



UBC-DIBS Working Paper 2021-CBI-06

Testing Two Nudges to Reduce Racial and Gender Bias in Student Evaluations of Teaching

Greg Lockwood, Andrea Vásquez Rodríguez, Rachel Yang, & Jiaying Zhao
University of British Columbia

Knowledge Summary: This project aimed to reduce racial and gender bias in Student Evaluations of Teaching at a large post-secondary institution in Western Canada. In a 2 x 2 factorial randomized controlled trial using midterm student evaluations from over 900 undergraduate students, a disclosure statement, pre-commitment, and disclosure statement plus pre-commitment were tested against a control condition. The results of the study were inconclusive, suggesting that other types of behavioral interventions, as well as systemic policy changes at the broader level, could be pursued to improve career outcomes for instructors from marginalized groups.

Keywords: *behavioural insights, nudge, disclosure, pre-commitment, racial bias, gender bias, bias reduction*


Suggested citation: Lockwood, G., Vásquez Rodríguez, A., Yang, R., & Zhao, J. (2021). *Testing two nudges to reduce racial and gender bias in student evaluations of teaching*. (UBC-DIBS Working Paper 2021-CBI-06).

UBC SAUDER
SCHOOL OF BUSINESS

DIBS

Decision Insights for Business & Society

UBC Decision Insights for Business & Society (UBC-DIBS) is a cluster of researchers and partners building a cross-sectoral centre of excellence to shape how Behavioural Insights (BI) are used in British Columbia and beyond. CBI Working Papers are capstone projects completed by BI Practitioners graduating from UBC's Advanced Professional Certificate in Behavioural Insights.



Acknowledgements

The project team would like to thank and acknowledge the following for their assistance with the completion of this project:

Project Sponsors: UBC Equity & Inclusion Office

Project Advisor: Jiaying Zhao

UBC Sauder Faculty: Kirstin Appelt, Dave Hardisty, Dale Griffin

BI Certificate TA: Rishad Habib

BI Certificate Pilot Cohort

Table of Contents

Executive Summary.....	4
Part A. Problem Background.....	5
Part B. Behaviour & Context.....	5
Part C. Exploratory Research.....	6
Part D. BI Solution.....	10
Part E. Research Design.....	13
Part F. Research Results.....	15
Part G. Recommendations.....	21
Part H. Discussion of BI & Research Ethics.....	22
Part I. Project Reflections.....	23
References.....	25
Appendices.....	28
Appendix I. Behaviour Map.....	28
Appendix II. Additional Secondary Research.....	29
Appendix III. Instructor Recruitment Questionnaire.....	30
Appendix IV. Student Evaluation of Teaching Survey.....	33
Appendix V. Data Cleaning Rules.....	37
Appendix VI. Additional Exploratory Analyses.....	38

Executive Summary

The teaching experiences and career opportunities, and lack thereof, of BIPOC (Black, Indigenous and People of Colour), women and gender-diverse instructors in higher education are under a microscope now more than ever. In academic institutions, the performance of post-secondary instructors is evaluated using a number of criteria, which varies across institutions. At the institution considered in this study, there is no single formula for performance evaluation; the process differs between faculties and even across departments, but it typically relies on data from student evaluations of teaching (SEoTs). As a result, student evaluations contribute in a significant way to the career advancement of instructors.

In recent years, there is growing evidence that BIPOC, and especially women, instructors are subjected to biased SEoTs (Bachen, McLoughlin, & Garcia, 1999; MacNell, Driscoll, & Hunt, 2014; Wagner, Rieger, & Voorvelt, 2016; Boring, 2017; Mengel, Sauermann, & Zölitz, 2019). However, the evidence is not conclusive and experimental research is under-represented in this area. Our study tackled this problem by developing and testing behavioural insights solutions that build on the existing evidence base and in-depth exploratory research.

The study tested two behavioural insights (BI) solutions: 1) Disclosure: an informative statement placed directly on the midterm student evaluations that described the impact of student evaluations on instructors' careers and the prevalence of racial and gender bias in instructor ratings, and 2) Pre-commitment: a request asking students to pre-commit to providing a fair and honest evaluation of their instructor. The study also included a condition that combined the two nudges, as well as a control condition. The solutions aimed to deactivate any implicit, unconscious biases students may hold that cause them to rate instructors based on their identity and/or appearance, rather than teaching ability.

919 undergraduate students across 41 course sections in the Winter 2021 Term 2 at the University of British Columbia (UBC) participated in the study in which a randomized controlled trial (RCT) methodology assessed whether the solutions causally influenced instructor ratings. The multi-treatment RCT allowed the project team to analyze the extent of the BI solutions' impact on students' ratings of BIPOC, women and gender-diverse instructors.

The results of the study were inconclusive, suggesting that other types of behavioral interventions, as well as systemic policy changes at the broader level, could be pursued to improve career outcomes for instructors from marginalized groups.

Part A. Problem Background

In our society's current moment of social awakening, organizations are expected more than ever to take action on the many processes where systemic racism and sexism permeate. While issues of inequality and inequity are not new, institutions are increasingly being called on to demonstrate the leadership needed to reduce and remove the sources of discrimination and bias that exist.

One way to address discrimination and bias at the University of British Columbia (UBC) is by improving the experience of BIPOC (Black, Indigenous and People of Colour), women and gender-diverse faculty at the university, including improving the rate of promotions for these equity-deserving groups. While it is well-known that BIPOC, women and gender-diverse employees at many organizations are subject to wage gaps, poor representation and discrimination, these issues are particularly important to UBC. As a demonstration of its commitments in this area, UBC has appointed Senior Advisors to the Provost on both Racialized Faculty, and Women and Gender-Diverse Faculty, specifically to "support institutional efforts to enhance the scholarly and leadership environment and opportunities" for BIPOC, women and gender-diverse faculty at UBC's Vancouver campus (Szeri, 2020).

The behaviour map in Appendix I illustrates that one potential way to address the overall goal of decreasing discrimination at UBC and improve the experience of equity-deserving instructors is to reduce bias in Student Evaluations of Teaching (SEoTs). The SEoT is an instrument that serves as an input into promotion and tenure decisions, which affect the careers of all UBC instructors.

Previous research provides evidence that there are differences in student evaluation ratings between men and women (Mengel, et al., 2019; Wagner, et al., 2016; MacNell et al., 2014). These biased differences in ratings could be driven by traditional gender stereotypes (Bachen, McLoughlin, Garcia, 1999; Boring, 2017) and could be described as unconscious and unintentional (Peterson et al., 2019). While there is less research around racial differences in student evaluations, some evidence has indicated that racialized instructors are evaluated more negatively than white faculty (Chávez & Mitchell, 2019; Reid, 2010). Because of the importance currently placed on SEoTs as an evaluation tool for instructors, poor SEoT ratings can negatively impact important decisions when it comes to promotions and tenure, and can disproportionately disadvantage equity-deserving groups, including BIPOC, women and gender-diverse instructors. This has detrimental effects on their career progression, especially those who receive lower ratings for reasons unrelated to teaching quality or subject matter knowledge.

Although it cannot be said that improving the SEoT alone will eliminate the systemic racism and sexism that exist in the academy, it is one potential avenue and a starting point for short-term improvements. A successful intervention involving the SEoT could nudge students to eliminate their racial and gender biases and submit a more fair and accurate evaluation of their instructors, thus improving the experience of equity-deserving faculty at UBC.

Part B. Behaviour & Context

The project team aimed to mitigate bias in students' ratings of BIPOC, women and gender-diverse instructors by reducing barriers to submitting fair evaluations. This target behaviour was analyzed using the MIST framework as follows:

- **Measurable:** The behaviour was observable through the evaluations. The SEoT is a quantitative

instrument in which students provide their perceptions of instructors' performance. Therefore, the completed questionnaire itself measured the relevant information for the project.

- **Important:** SEoTs are often used to inform tenure and performance decisions. Consequently, biased, SEoTs could negatively impact the career progression of BIPOC, women and gender-diverse instructors. Fair SEoTs would contribute to a more positive experience for equity-deserving faculty members, and may encourage more diverse faculty representation and an enriched learning experience for students.
- **Sizable:** The large undergraduate student population at UBC meant that a successful recruitment process with even just a very small fraction of participating students would enable the team to collect a sufficient number of valid survey responses for the execution of a rigorous trial.
- **Feasible:** The SEoT process was already in place and there were no costs associated with setting up the interventions. The timing for the project also aligned well with the midterm evaluation period and was viable administratively, much more so than the on-the-record end-of-term SEoTs. The team had also conducted several meetings with key stakeholders to ensure that recruitment was feasible.

The population of interest for this project is UBC's undergraduate student population of over 55,000. Our team's contacts in the university included the Equity & Inclusion Office which, along with the faculty from the BI certificate and other programs, facilitated and accelerated the recruitment process.

Influencing students' behavior was appropriate because of:

- **High impact:** Students' ratings of instructors in SEoTs highly impact instructors' careers and the institution's faculty diversity.
- **Availability:** Instructors and students demonstrated willingness to participate in this project.
- **Access:** Students were easily reached through the typical midterm evaluation process.

Three barriers to the target behaviour are:

- **Limited attention (distraction):** Students may have limited attention when completing SEoTs. In this context, unconscious biases can surface, especially while doing tasks that require little intellectual effort, like responding to simple questions.
- **Limited time:** Students have competing demands and therefore those who actually complete the evaluation might do so without proper care or thought. It is unclear if this subset of students is on average more or less biased than the group of students who choose not to complete SEoTs.
- **Limited knowledge and inertia:** Students might not be aware of the impact that the evaluations have on the instructors' careers, thus may invest little mental energy on completing the SEoTs.

The project team hypothesized that behavioural insights that increase awareness, request for commitment in advance, and extend completion time, could potentially help overcome these barriers.

Part C. Exploratory Research

To understand students' barriers to completing fair student evaluations of teaching, we conducted secondary and qualitative research. Our secondary research consisted of a literature review of previous studies focused on gender and racial bias in student evaluations of teaching (SEoTs.) For our qualitative research, the project team conducted six interviews (two with undergraduate students, four with faculty) to explore their opinions and feedback on the proposed interventions.

Secondary Research

The body of research involving SEoTs in a university setting has increased over the past 20 years, and a number of relevant studies provide a useful background for understanding implicit bias in SEoTs. Many focus on gender differences between men and women, while relatively fewer have focused on racialized or other diverse instructors. Most studies report small effect sizes and very few have been replicated.

Starting broadly, previous research provides evidence that there are indeed differences in SEoT ratings between groups such as men and women. Mengel, Sauermann and Zölitz (2019) concluded that women receive systematically lower teaching evaluation scores compared to their colleagues who are men, and that the effect is pronounced for junior women. Wagner, Rieger and Voorvelt (2016) also found that women faculty receive considerably lower evaluation scores when co-teaching a course with men, compared to teaching alone or with other women. Even when instructors' true identities were disguised in an online course, gender bias was still present, irrespective of the instructor's actual identity (MacNeill, Driscoll, & Hunt, 2014). These biased differences in ratings could be driven by traditional gender stereotypes (Bachen, McLoughlin, Garcia, 1999; Boring, 2017) and could be described as unconscious and unintentional (Peterson et al., 2019).

While there is less research around racial differences in SEoTs, one quasi-experimental design by Chávez and Mitchell (2019) found that women instructors and those who identify as people of colour received lower scores on SEoTs compared to white men, reinforcing the potential for bias in these ratings. Another study by Reid (2010) examined anonymous instructor ratings on the commonly used website *Ratemyprofessors.com* from high-ranking liberal arts colleges in the US. Racial minority faculty were evaluated more negatively than white faculty in terms of overall quality, helpfulness and clarity, but were rated higher on easiness. The author concluded that these findings can have implications for the tenure and promotion of racial minority faculty.

At UBC, where the planned intervention is taking place, an SEoT Working Group was formed in February 2019 to undertake a re-examination of its approach to student evaluations. The group also cited research conducted at UBC. An early study conducted at the UBC Vancouver campus with SEoTs (Hakstian, 2009) revealed that there are no practical differences in the aggregate ratings of women versus men instructors. However, course-specific effects, such as individual subject disciplines or course levels, showed slightly larger effects than the aggregate. A follow-up study years later showed that there was no significant gender bias in SEoT scores, despite there being some differences in scores for individual questions (UBC Centre for Teaching, Learning & Technology, 2015). Meanwhile, findings at the UBC Okanagan campus were mixed (UBC Okanagan Planning and Institutional Research, 2017 & 2020). Nonetheless, the SEoT Working Group concludes that "the lived experience of individual instructors may be quite different from this aggregate view." One of the working group's key recommendations is to retain student feedback as an input into instructor evaluation, along with other sources of data. It should be noted that understanding the experience of BIPOC faculty was not a primary focus in previous studies at UBC. Most were aimed at understanding the difference in scores between women and men instructors; in these studies, the binary sex variable was used as a proxy for gender, which is not accurate because it erases the experiences of trans and non-binary instructors.

Although evidence suggests that SEoTs appear to be biased against equity-deserving groups, SEoTs continue to be an important component (and sometimes the only component) of an instructor's performance review at many academic institutions across North America (UBC Student Evaluation of Teaching Working Group, 2020). It is anticipated that SEoTs at universities will remain a core component of an instructor's performance review for years to come. However, few studies have attempted to find a solution to mitigate the problem.

One relevant study from which the project team drew inspiration was conducted at Iowa State University (Peterson et al., 2019), which used a randomized controlled trial design to test a statement that brought awareness to students' implicit bias. With 249 participating students, the researchers found that as a result of adding a disclosure statement at the beginning of the SEoT alerting students to be aware of their own implicit bias (randomly shown to roughly half of the participants), students rated women instructors higher than in the SEoTs without the statement, while scores of men instructors did not change. This not only reinforced that implicit bias does seem to exist in SEoTs, but also that there are potential ways of reducing students' implicit bias.

Another recent study by Hoorens, Dekkers and Deschrijver (2020) in Belgium tested the effectiveness of self-affirmation in SEoTs. Non-self-affirmed participants showed a gender bias after being told they had received a bad grade, disadvantaging the women professors. However, self-affirmation (through either a value-affirming task or self-superiority priming) eradicated gender bias by lowering evaluations for the men professors, suggesting that students over-value men as professors. While not directly related to the planned intervention, these findings help to set the landscape of bias mitigation tactics that have already been tested.

The existing research around the topic of SEoTs, as well as the limited number of interventions tested in this area, provided a starting point for addressing the behavioural problem at hand. The project team felt that a disclosure statement was a potential route to explore and wanted to build on the success of Peterson et al. in their disclosure statement intervention. In addition, the team drew from social psychology, where Axt, Casola and Nosek (2019) conducted a study asking respondents to select honour society applicants based on academic credentials as well as irrelevant social categories such as attractiveness and ingroup status. The authors found that asking participants to avoid potential bias in one social category showed slight reductions in bias for that category. However, the authors concluded that the source of bias should be explicitly stated because mentioning multiple social biases or vaguely referring to bias did not have a consistent effect on reducing bias.

Another avenue the project team explored was the idea of a pre-commitment device. This tool draws on the behavioural insight that people generally want to take actions that are consistent with their self-image. Pre-commitment devices can be broadly described as "the arrangements people make to formalize and facilitate their goals" (Bryan et al., 2010, pp.672) and typically involve making a plan in advance with the aim of following through with the commitment in the future. Pre-commitments may include financial consequences (e.g., paying a monetary penalty if the goal is incomplete) or exclude them (e.g., making a free appointment). Our study excluded financial consequences due to impracticality, ethical reasons, and the mixed efficacy of financial pre-commitments (Derksen et al., 2021; John, 2020; Bryan et al., 2010). In terms of non-financial pre-commitments, results have also been somewhat mixed: interventions such as making and writing down appointment times have been shown to be very effective (IRS, 2017), while others such as using a signature line at the beginning of a form instead of at the end (Shu et al., 2012; Social and Behavioral Sciences Team, 2015) have been inconclusive or irreproducible (Kristal et al., 2020). Pre-commitment interventions to date have typically involved scenarios with tangible or direct outcomes (e.g., showing up to an appointment), or have been concerned with high-stakes legal or financial repercussions (e.g., in the domains of income tax, financial self-reporting, and the insurance and automotive industries). However, little exploration has been done with pre-commitments in more subjective areas such as unconscious bias.

One important consideration for this project and its potential impact is the prevalence of online learning. Due to the COVID-19 pandemic, UBC suspended all in-person classes for the 2020/2021 academic year, and there was an unprecedented number of students taking all of their courses online. There has been some, but relatively much less, research done in the area of online learning compared to traditional in-person

environments, particularly when it comes to student satisfaction with a course. Summers, Waigandt and Whittaker (2005) found that students enrolled in an online course were significantly less satisfied with the course than traditional classroom students. Shea et al. (2004) identified factors that impact student satisfaction, such as expressing clear expectations, timely feedback, low levels of technical difficulties and high-quality feedback on assignments. Roblyer and Ekhaml (2000) found that the quantity and quality of online interaction (whether with instructors or fellow peers) directly correlated with student satisfaction and learning. Awalt (2003) and Jackson, Jones and Rodriguez (2010) investigated the instructor's role in online learning, and concluded that faculty who email their students frequently, respond to emails promptly, hold regular office hours and introduce personal touches tend to be successful in the online learning environment. In designing BI solutions to reduce students' bias on SEoTs and assessing the impact and scalability of the project at hand, the differences between online learning and in-person classrooms are all factors to consider. For additional secondary research, refer to Appendix II.

Qualitative Research

The team conducted six interviews (two with undergraduate students, four with faculty) to explore their opinions and feedback on the proposed interventions. Recruitment was done through the Equity & Inclusion Office and no demographic information of students nor instructors were collected. The students interviewed did not take part in the trial.

With a small sample, it was difficult to establish patterns within the responses. However, some commonalities were found. For example, students and instructors in our interview sample viewed SEoTs as unfair assessments due to students' biases, social norms, sexism, racism, homophobia, and lack of knowledge on how to give constructive feedback. These perceptions reinforce the secondary research that SEoTs are unreliable measures of performance and that students tend to hold biases when evaluating their instructors.

When presented with the interventions, students and instructors saw them as meaningful, clear, and timely, but they all doubted their efficacy. That is, even though they thought the wording and the content were accurate, relevant, and easy to understand, they felt that the interventions would be insufficient to challenge students' biases. The project team addressed the interviewees' feedback by incorporating their suggestions into the interventions when possible and appropriate.

For instance, one interviewed student mentioned that even though the intervention statements might be well-worded, students would have trouble connecting them with their own behaviour. Some instructors noted that many students are convinced they are not racist or biased. Therefore, they would think the intervention is not applicable to them and would avoid further introspection. On a similar note, one instructor mentioned that asking students to recognize and manage their biases may not be enough for them to differentiate between their prejudices and their valid assessment of the instruction received.

After viewing draft BI solutions (see Part D for final versions), interviewees suggested using more directive language, more examples, and including a citation when mentioning that evidence shows that women, BIPOC and gender-diverse instructors receive systematically lower SEoT ratings. Considerable feedback revolved around use of the concept "fair/fairness" in the pre-commitment statement; participants stated that it is a broad term that might be interpreted in varying ways. One student pointed out how a "fair and honest" (evaluation) does not necessarily translate into an equitable one. However, no improvements to the term were suggested during the interviews.

Some interviewees mentioned that within the current digital realities, it was standard practice to accept certain terms of the survey before advancing to the next step or page without real acknowledgment of their

implications. Others thought that the pre-commitment would compel the students to be accountable, calling them to action on a specific matter and preventing them from skimming the text inattentively.

The qualitative research informed the interventions' final version. The most significant changes the team made included using more direct language throughout the statements and eliminating the option of "Prefer not to answer" in the pre-commitment, leaving only "Yes" or "No" as possible responses. The team also decided to add a citation of the evidence in the disclosure statement and eliminated redundant text.

The team also consulted with Dr. Toni Schmader, Canada Research Chair in Social Psychology at UBC, an expert in social identities and stereotypes. Dr. Schmader pushed the team to think about adding additional measures to ensure that students were actively thinking through the survey.

To ensure students actually read the statements they were shown, the team added timers to the online survey, preventing students from proceeding until 10 seconds (for the pre-commitment), 15 seconds (for the disclosure statement) and 20 seconds (for the combined intervention) had passed. No timer was added for the control condition. The team also added a self-reflection question, which asked "*How fair and honest do you feel your evaluation of this instructor was, based on the quality of the instruction you received?*" and was rated by students on a 5-point scale from "Extremely Fair" to "Not at All Fair". This was added to understand if there are any differences in how students perceive the fairness of their own evaluations compared to how they actually rated their instructors. This question was shown after all six SEoT questions, so it did not interfere with the students' ratings of their instructor.

After conducting secondary research and analyzing the exploratory research, we became more aware that asking students to recognize and combat their biases in SEoTs would be a tall, but not impossible, order. The team made the suggested improvements to the intervention and proceeded as planned.

Part D. BI Solution

The BI solutions for the project are a **disclosure statement** and a **pre-commitment device**.

Our secondary research indicated that bias in SEoTs is often implicit and unconscious, and that students may not be aware of their own biases when completing instructor evaluations (Peterson et al., 2019). During the process of completing the evaluation, students may not be thoughtfully responding to questions. A number of factors may be contributing to students' lack of thoughtfulness, including limited cognitive capacity due to feeling stressed from their workload, limited time and attention to complete SEoTs for all the courses they are taking, or being distracted (e.g., in an online learning environment, emails, notifications, etc. may interrupt students while they are completing SEoTs; in person, students may be distracted or influenced by their peers.)

The broad intention of an intervention in this project is to slow students down and encourage more thoughtfulness when completing instructor evaluations. This project tested two nudges: a disclosure statement and a pre-commitment device, as well as a combination of the two nudges, against a control condition (shown in Figure 1).

As shown in Figure, 2, the disclosure statement provided information about unconscious bias that may not have been known to students and a directive to resist stereotypes and focus on the quality of instruction, with the aim of encouraging students to think about their own biases when evaluating their instructor. This

information was included on the first page of the survey (after the consent form) as a prompt to provide more thoughtful ratings of their instructor.

As shown in Figure 3, the pre-commitment device asked students to commit to providing a fair and honest evaluation of their instructor *before* they started the survey. This question prompted students to reflect on the fairness and honesty of their impending evaluation before responding to the questions. This pre-commitment also leveraged the behavioural insight that people generally want to make decisions and take actions that are consistent with their self-image.

As shown in Figure 4, the combined condition included both the disclosure statement and the pre-commitment device. The complete SEoT viewed by students can be found in Appendix IV.

Figure 1. Control condition introductory text.

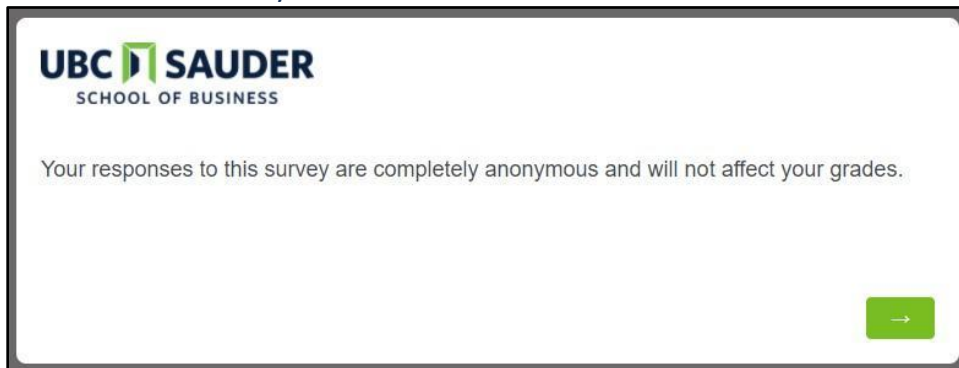


Figure 2. Disclosure statement condition introductory text.

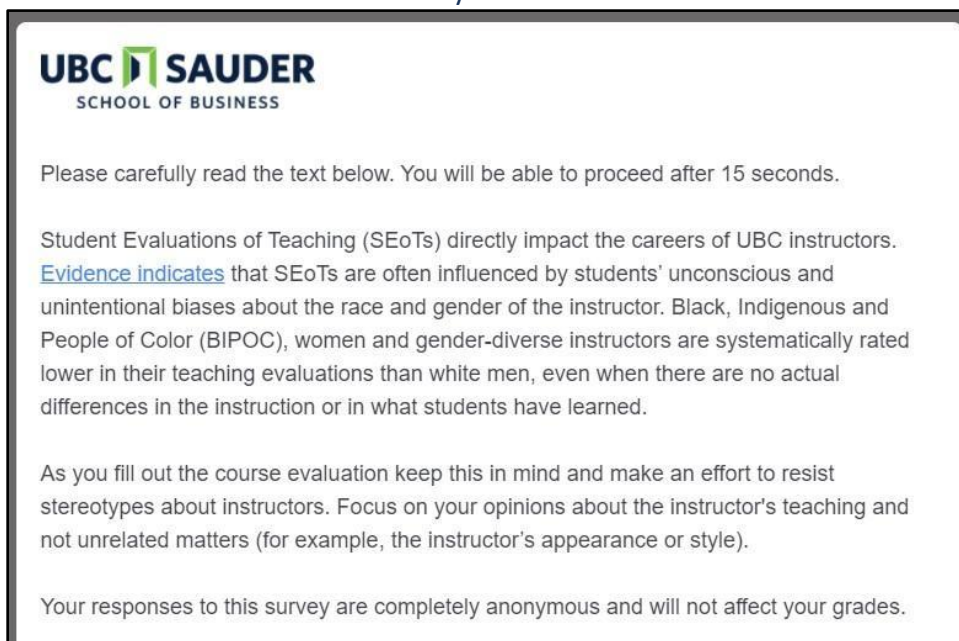


Figure 3. Pre-commitment condition introductory text.

UBC SAUDER
SCHOOL OF BUSINESS

Please carefully read the text below. You will be able to proceed after 10 seconds.

Your responses to this survey are completely anonymous and will not affect your grades.

Do you commit to completing a fair and honest evaluation of this instructor based on the quality of the instruction you received?

Yes, I commit

No, I do not commit

Figure 4. Combined disclosure statement plus pre-commitment condition introductory text.

UBC SAUDER
SCHOOL OF BUSINESS

Please carefully read the text below. You will be able to proceed after 20 seconds.

Student Evaluations of Teaching (SEoTs) directly impact the careers of UBC instructors. [Evidence indicates](#) that SEoTs are often influenced by students' unconscious and unintentional biases about the race and gender of the instructor. Black, Indigenous and People of Color (BIPOC), women and gender-diverse instructors are systematically rated lower in their teaching evaluations than white men, even when there are no actual differences in the instruction or in what students have learned.

As you fill out the course evaluation keep this in mind and make an effort to resist stereotypes about instructors. Focus on your opinions about the instructor's teaching and not unrelated matters (for example, the instructor's appearance or style).

Your responses to this survey are completely anonymous and will not affect your grades.

Do you commit to completing a fair and honest evaluation of this instructor based on the quality of the instruction you received?

Yes, I commit

No, I do not commit

In assuming that most students are well-intentioned and do not purposely submit biased instructor ratings, the project team anticipated that one or both of the nudges would encourage students to provide responses that are more thoughtful and fairer, thus helping to reduce racial and gender bias.

This intervention was deemed feasible because it was a no-cost solution and was straightforward to implement. Since UBC midterm SEoTs were already planned to be administered online during the 2020/2021

Winter Term 2, incorporating the intervention was simply a matter of adding additional text at the beginning of the survey before the core SEoT questions. The intervention was carried out in February and March 2021, which aligned with the typical schedule for midterm evaluations in the second winter term at UBC. This touchpoint was ideal, since midterm evaluations have no impact on instructors' career progression; they are for instructors' information purposes only. Institutional barriers were low; stakeholders at UBC were supportive of this project and generally agreed with the recommended approach. The publicity principle held well, with stakeholders indicating that they were willing to share the design and results with both instructors and students after the trial was completed.

Although there was evidence in one relevant study which showed that a disclosure statement had a positive effect on SEoT ratings for women instructors without lowering the scores of men instructors (Peterson et al., 2019), it was not guaranteed that the same result would be replicated in our trial. A pre-commitment statement had not been tested in previous experiments of this nature (to the knowledge of the project team), and the effectiveness of pre-commitments has been challenged in recent follow-up studies (Kristal, et al., 2020). In addition, the team also felt there was the possibility that a description of, and a call to resist, stereotypes, and provide "fair and honest" feedback, could result in a backfire effect if the statements paradoxically activated and normalized stereotypes. To summarize, it was unclear whether the intervention would have a significant impact, which justified the rationale for an experimental trial.

Part E. Research Design

The research design for this project is a **randomized post-test only design with three treatment arms and one control condition**.

We implemented a randomized controlled trial design to ensure an evaluation of the interventions' causal influence on students' SEoT behaviour. In randomly assigning students, the evaluation was able to determine any causal effects of either of the two nudges, or their combination, on students' evaluations of instructors.

Summary of the Treatment Conditions

The intervention was trialed on the optional midterm SEoTs distributed to students in February and March 2021. See Part D for a description of the treatments. A mandatory time delay was included in each of the treatment conditions to ensure that students read the information before proceeding. However, because there was no time delay in the control condition, the time delays in the treatment conditions served as an additional intervention and confounded the design.

Participants

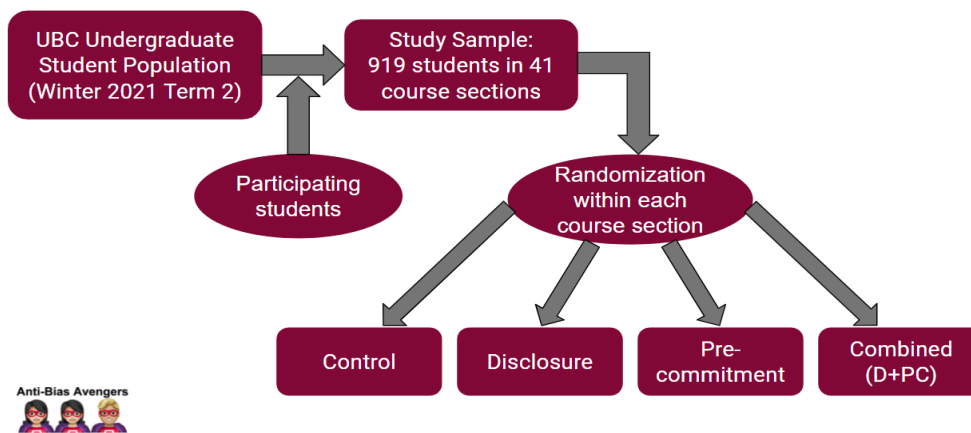
Participants are UBC undergraduate students: 1) in courses whose instructors consented to participate in this trial and 2) who completed the SEoT (the version to which they were randomly assigned). Instructors were eligible to participate if they taught an undergraduate course in the second Winter 2021 term. The project team developed a one-page project description and recruited instructors through informal and formal networks at UBC. Examples of these include the Centre for Teaching, Learning and Technology's Slack Channel, the Faculty Equity Leads Network, the EDI (Equity, Diversity & Inclusion) in Engineering committee and the Sauder Marketing & Behavioural Science Division. Overall, 30 instructors and 919 students participated in the study, which far surpassed expectations.

Method of Randomization

Participating instructors completed an enrollment survey in which they identified which course sections would be part of the study (see Appendix III). For each section, all enrolled students were randomized into one of the four conditions using the built-in randomization feature of the online survey software Qualtrics after they clicked the SEoT survey link and consented to participate. Students and instructors were not made aware of which condition each student was assigned to.

Figure 5. Visual illustration of the research design.

Research Design - Post-test only Design with 3 Treatment Arms



Method of Control

Upon clicking the SEoT link, students were asked for their consent to participate; the consent form included a broad description of the study. From the time they received the SEoT link, students had approximately ten days to complete the SEoT. (The survey was open from February 22 to March 7, 2021 and instructors were asked to send out their survey links during this time period.) Instructors received a reminder to resend the link approximately one week after they initially sent the SEoT links to students. Based on the study's implementation, demand characteristics, placebo effects and the Hawthorne effect (where individuals modify their behavior in response to being aware they are being observed) were not a concern. Contamination was only a possibility if students spoke with others in their class about the SEoT they received and chose to complete the survey multiple times. Based on the low number of repeat IP addresses (which were analyzed for possible exclusion according to a set of criteria; see Appendix V) and the even balance of participants across the four conditions, contamination was negligible or non-existent.

Measured Outcome/Dependent Variables

There are six outcome variables – students' ratings on each of the six midterm SEoT questions which are presented on a 5-point (Strongly disagree = 1 / Disagree = 2 / Neutral = 3 / Agree = 4 / Strongly agree = 5) scale. The six SEoT questions are listed in Appendix IV.

Hypotheses

The project team formulated the following three hypotheses - one for each treatment condition compared to the control condition. The hypotheses were part of the [pre-registration](#) the team submitted to Open Science Framework.

Hypothesis 1: We expect that the disclosure statement will increase Student Evaluations of Teaching (SEoT) more for instructors who identify as BIPOC, women, and/or gender-diverse (hereafter BIPOC/women/gender-

diverse instructors) than for instructors who identify as white men (hereafter white men instructors), compared to the control condition.

Hypothesis 2: We expect that the pre-commitment will increase Student Evaluations of Teaching (SEoT) more for BIPOC/women/gender-diverse instructors than for white men instructors, compared to the control condition.

Hypothesis 3: We expect that the combined condition (disclosure and pre-commitment) will increase Student Evaluations of Teaching (SEoT) more for BIPOC/women/gender-diverse than for white men instructors, compared to the control condition.

Data Collection

There were two primary data collection tools: 1) the instructor enrollment survey (distributed prior to the SEoTs, see Appendix III), and 2) the midterm SEoT randomized by the four conditions (see Appendix IV).

The instructor survey collected the two key independent variables from participating instructors: a) self-reported BIPOC identity and b) self-reported gender identity. The primary independent variable of interest, a new binary variable created by the project team which captured the instructors' BIPOC/women/gender-diverse or white man identity, was constructed from these two variables.

The midterm SEoTs for this project were designed in Qualtrics and included UBC's standard six SEoT questions. Of these six dependent variables, the primary dependent variable of interest is the final question, UMI-6 (University Module Item), which the project team hypothesizes links most directly to the intervention and is typically viewed as the question that carries the most weight in instructors' performance evaluations: "*Overall, this instructor was effective in helping me learn.*" The SEoT also included similar demographic questions that were posed to instructors in the instructor survey. Instructors were also given the option of including additional questions in their own SEoTs; these were not included in the analysis. No personally identifying information (names, email addresses, student numbers, etc.) was collected from participating students.

Instructors were assigned unique survey links which allowed the project team to track SEoT data for each individual course section. The project team sent instructors the survey links for distribution to their students in late February. Within two weeks of the survey's closing, the project team created and sent instructors an aggregate report of their course section's evaluations, removing all RCT-related data.

The project team applied a thorough data cleaning protocol (see Appendix V) and carefully reviewed the data sets to create a master data file for analyzing the research results, which are described in the next section.

Part F. Research Results

Study Sample

Of the 30 participating instructors, 21 self-identified as BIPOC/women/gender-diverse, while the remaining nine identified as white men. A total of 919 completed midterm Student Evaluations of Teaching were included in the final sample. Further information about the study sample is shared in Table 1.

Table 1. Study sample with sample sizes for instructors and students.

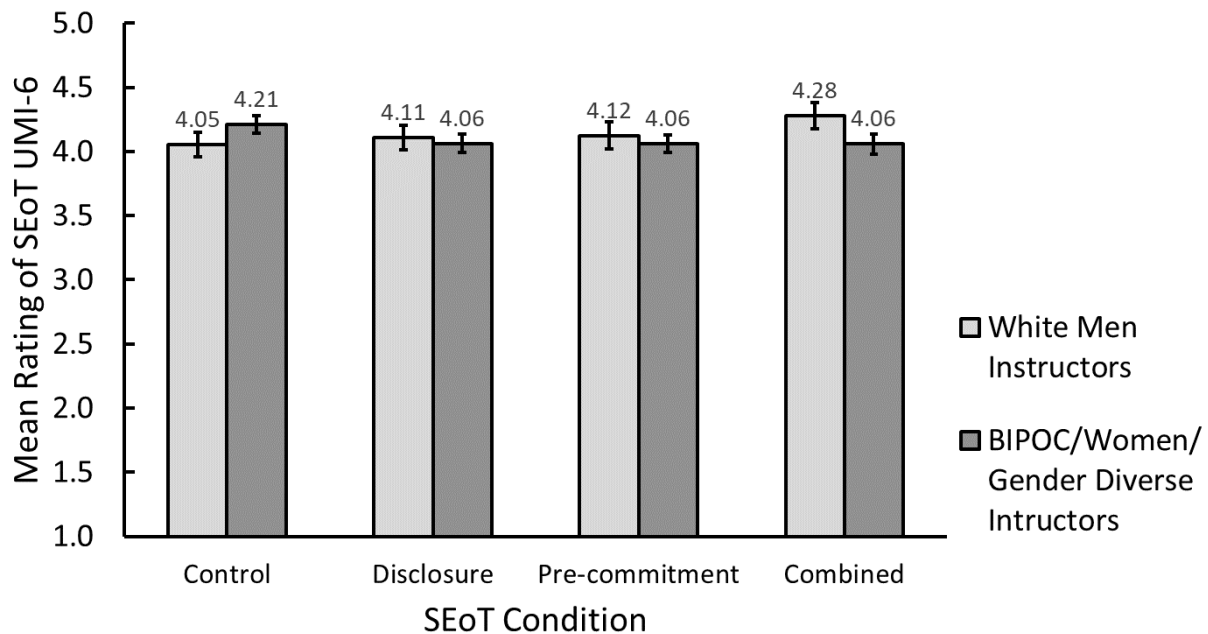
Independent Variable	Identity / Condition Assignment	Number of Instructors	Number of Students (Note: Students who selected 'Prefer not to answer' are not included)
BIPOC/Women/Gender Diverse vs. White Men	BIPOC/Women/Gender Diverse	21	832
	White Men	9	68
Gender	Women	17 (includes Gender-Diverse)	576
	Gender Diverse	N/A	14
	Men	13	294
Race	BIPOC	9	639
	White	21	262
Condition	Control	N/A	238
	Disclosure		230
	Pre-commitment		245
	Combined		206

It is important to note that in the study sample, none of the nine BIPOC instructors identified as Black or Indigenous and seven of nine identified as East Asian (e.g., Chinese, including Hong Kong and Macau, Japanese, Korean, etc., and including Asian-Canadian, Asian-American, etc.).

Analysis of the Hypotheses

Figure 6 illustrates students' average UMI-6 rating and the associated standard error bars in each of the four conditions, grouped by the BIPOC/women/gender-diverse and white men instructor identities. Notably, in the control condition, the mean score for BIPOC/women/gender-diverse instructors is higher than that of white men instructors, but this difference is reversed in the treatment conditions. The difference is most pronounced in the combined condition. This is an indication that each of the three treatment conditions did not have the intended effect of mitigating students' bias and increasing the ratings of BIPOC/women/gender-diverse instructors more than white men instructors.

Figure 6. Mean Instructor Rating of UMI-6 among BIPOC/women/gender-diverse and white men instructors in the four SEoT conditions. Means are displayed on the outside end of each bar. Error bars indicate +/- one standard error.



To assess the extent to which the BI solution mitigated students' implicit bias and changed their behaviour, we tested each of the hypotheses by comparing the ratings of the SEoT UMI-6 of one of the three treatment conditions to the control condition. We used standard two-tailed tests and adjusted our criteria for multiple comparisons using a Bonferroni correction.

In Hypothesis 1, we expect that the disclosure statement will increase SEoT more for BIPOC/women/gender-diverse instructors than for white men instructors, compared to the control condition. To test this, we conducted a 2 (instructor identity: BIPOC/women/gender-diverse vs. white men) x 2 (condition: disclosure statement vs. control) between-subjects ANOVA on the dependent measure (mean rating of SEoT UMI-6). There were no main effects of instructor identity or condition, $F(1, 464) = 0.45, p = .502, corrected p = 1, partial \eta^2 = .001$ and $F(1, 464) = 0.28, p = .596, corrected p = 1, partial \eta^2 = .001$, respectively. Contrary to Hypothesis 1, there was also no significant interaction, $F(1, 464) = 1.40, p = .237, corrected p = .711, partial \eta^2 = .003$. In fact, the scores for BIPOC/women/gender-diverse were lower in the disclosure condition ($M = 4.06, SD = 0.93$) than in the control condition ($M = 4.21, SD = 0.84$), whereas scores for white men in the disclosure condition ($M = 4.11, SD = 0.80$) were higher than in the control condition ($M = 4.05, SD = 0.84$).

In Hypothesis 2, we expect that the pre-commitment device will increase SEoT more for BIPOC/women/gender-diverse instructors than for white men instructors, compared to the control condition. To test this, we conducted a 2 (instructor identity: BIPOC/women/gender-diverse vs. white men) x 2 (condition: pre-commitment device vs. control) between-subjects ANOVA on the dependent measure (mean rating of SEoT UMI-6). There were no main effects of instructor identity or condition, $F(1, 479) = 0.32, p = .570, corrected p = 1, partial \eta^2 = .001$ and $F(1, 479) = 0.21, p = .644, corrected p = 1, partial \eta^2 = .000$, respectively. Contrary to Hypothesis 2, there was also no significant interaction, $F(1, 479) = 1.70, p = .193, corrected p = .579, partial \eta^2 = .004$. In fact, the scores for BIPOC/women/gender-diverse were lower in the pre-commitment condition ($M = 4.06, SD = 0.89$) than in the control condition ($M = 4.21, SD = 0.84$), whereas scores for white men in the pre-commitment condition ($M = 4.12, SD = 0.95$) were higher than in the control condition ($M = 4.05, SD = 0.84$).

In Hypothesis 3, we expect that the combination of disclosure and pre-commitment will increase SEoT more for BIPOC/women/gender-diverse instructors than for white men instructors, compared to the control condition. To test this, we conducted a 2 (instructor identity: BIPOC/women/gender-diverse vs. white men) x 2 (condition: combined disclosure and pre-commitment vs. control) between-subjects ANOVA on the dependent measure (mean rating of SEoT UMI-6). There were no main effects of instructor identity or condition, $F(1, 440) = 0.11, p = .737, \text{corrected } p = 1, \text{partial } \eta^2 = .000$ and $F(1, 440) = 0.18, p = .671, \text{corrected } p = 1, \text{partial } \eta^2 = .000$, respectively. Contrary to our hypothesis, there was also no significant interaction, $F(1, 440) = 4.645, p = .032, \text{corrected } p = .096, \text{partial } \eta^2 = .010$. In fact, if anything, the scores for BIPOC/women/gender-diverse were lower in the combined condition ($M = 4.06, SD = 0.92$) than in the control condition ($M = 4.21, SD = 0.84$), whereas scores for white men in the combined condition ($M = 4.28, SD = 0.88$) were higher than in the control condition ($M = 4.05, SD = 0.84$).

In summary, none of our BI solutions boosted UMI-6 scores of BIPOC/women/gender-diverse instructors relative to white men instructors. These null results applied to the first five UMI questions as well (see Appendix VI). That is, when the same ANOVA was conducted using each of the other five questions as the dependent measure, the BI solutions did not significantly impact students' ratings of BIPOC/women/gender-diverse instructors relative to white men instructors.

Exploratory Analyses

In addition to testing the primary hypotheses discussed above, we conducted exploratory analyses of the data set to investigate how the BI solution impacted evaluation ratings in other ways.

Analyzing Instructor Identity by Gender. As a first set of exploratory analyses, it was important to parse instructor identity by both gender and race separately and examine any effects the interventions might have had on students' evaluations of instructors. The gender analysis is especially relevant, given that the bulk of the literature on bias in SEoTs has demonstrated evidence of bias against women instructors.

For the analysis by gender, we conducted a 2 (instructor identity: women vs. men) x 2 (condition: disclosure statement vs. control) between-subjects ANOVA on the dependent measure (mean rating of SEoT UMI-6.) We repeated this analysis with the other two treatment conditions – pre-commitment device vs. control and combined disclosure and pre-commitment vs. control. Once again, the individual and combined interventions did not significantly impact students' ratings when comparing men to women instructors (Figure 7). Notably however, there was a main effect of gender. Women instructors were rated higher than men instructors in both the disclosure and pre-commitment conditions, but not the combined condition, to the control, $F(1, 458) = 8.90, p = .003, \text{corrected } p = .009$ and $F(1, 473) = 6.97, \text{corrected } p = .0027$, respectively.

Analyzing Instructor Identity by Race. The analysis by race followed the same structure as the above analysis by gender. We conducted a 2 (instructor identity: BIPOC vs. white) x 2 (condition: disclosure statement vs. control) between-subjects ANOVA on the dependent measure (mean rating of SEoT UMI-6.) We repeated this analysis with the other two treatment conditions – pre-commitment device vs. control and combined disclosure and pre-commitment vs. control. Once again, the individual and combined interventions did not significantly impact students' ratings when comparing white to BIPOC instructors (Figure 8). Similar to the analysis by gender, there was a main effect of race. White instructors were rated higher than BIPOC instructors when comparing all three treatments, disclosure, pre-commitment and combined, $F(1, 464) = 13.24, p < .001, \text{corrected } p < .001, \text{partial } \eta^2 = .028, F(1, 464) = 8.74, p = .003, \text{corrected } p = .009, \text{partial } \eta^2 = .018, F(1, 464) = 7.40, p = .007, \text{corrected } p = .021, \text{partial } \eta^2 = .017$, respectively.

Figure 7. Mean Instructor Rating of UMI-6 among women vs. men instructors in the four SEoT conditions. Means are displayed on the outside end of each bar. Error bars indicate +/- one standard error.

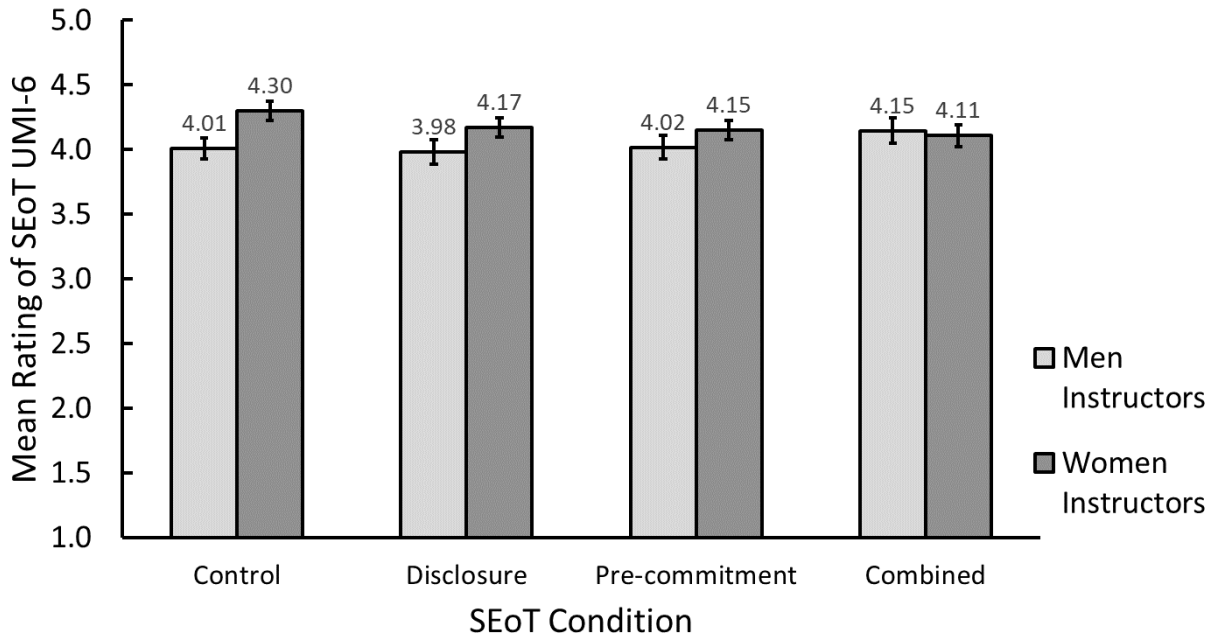
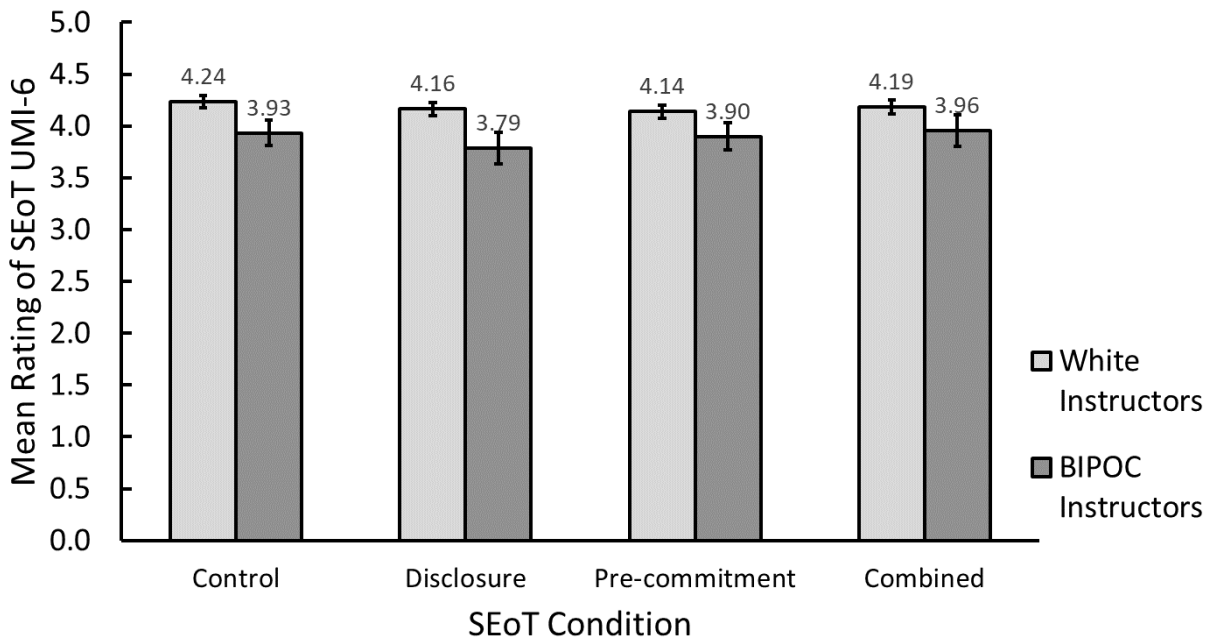


Figure 8. Mean Instructor Rating of UMI-6 among BIPOC vs. white instructors in the four SEoT conditions. Means are displayed on the outside end of each bar. Error bars indicate +/- one standard error.



Analyzing Instructor Identity and Student Identity. To explore any relationships between instructor identity, student identity and the four SEoT conditions on evaluation ratings, we conducted a 2 (instructor identity: BIPOC/women/gender-diverse vs. white/men) x 2 (student identity: BIPOC/women/gender-diverse vs. white/men) x 4 (condition: disclosure vs. pre-commitment vs. combination vs. control) ANOVA on UMI-6 scores (see Table 2).

As shown in Table 5, there was a main effect of student identity ($p = .023$), indicating that white men students ($M = 4.26$, $SD = 0.82$) rated instructors higher than BIPOC/women/gender-diverse students ($M = 4.05$, $SD =$

.91) on UMI-6. There was a marginal two-way interaction between instructor identity and condition; this interaction was analyzed in the study’s three hypotheses above. There were no other significant main effects or interactions, indicating that students were not differentially impacted by the BI solutions.

Table 2. Results of a 2X2X4 ANOVA comparing instructor identity, student identity and the four SEoT conditions on UMI-6.

Predictor	df	F	p-value	η_p^2
Instructor Identity	1, 885	0.637	.425	.0007
Student Identity	1, 885	5.167	.023	.006
Condition	3, 885	0.168	.918	.0005
Instructor Identity X Student Identity	1, 885	0.294	.588	.0003
Instructor Identity X Condition	3, 885	2.681	.046	.009
Student Identity X Condition	3, 885	0.178	.911	.0006
Instructor Identity X Student Identity X Condition	3, 885	2.007	.111	.007

Additional exploratory analyses can be found in Appendix VI.

Understanding the Findings

The project team reflected on the research results and offers the following insights to help understand the findings.

The study was tested in a different context from that of the secondary research. Our study tested a similar version of the disclosure statement that improved ratings of women instructors at Iowa State University (Peterson et. al, 2019). However, there are a number of differences between the two contexts that could have contributed to the differences in the results. UBC is located in a large metropolitan centre and is home to a demographically diverse student body, whereas Iowa State is located in a rural setting with a demographic that identifies as predominantly white (Data USA, 2020). These factors could seemingly limit the scalability of the intervention.

Students were taking courses in an online environment. As mentioned in the secondary research, the online learning environment is an important consideration for this project. Findings from Summers, Waigandt and Whittaker (2005) that students enrolled in an online course were significantly less satisfied with the course than traditional classroom students, as well as other differences between the in-person and online learning environments, may have impacted the ways in which students responded to the interventions.

Calling out implicit biases may have contributed to reactance in students. The racial justice movement of 2020 has made racial inequities more visible than ever to the broader population and might have impacted how an anti-bias disclosure statement is viewed by students. A study that examined public support of policies with racial framing (English & Kalla, 2021) demonstrates that despite more progressive public attitudes towards issues of racial equality, racial framing decreases support for race-neutral progressive policies. These

same mechanisms could be at play in the interventions that included the anti-bias disclosure statement. More generally, students could have experienced psychological reactance to the statement and behaved in the opposite way of what the interventions intended.

The use of timers may have skewed the results. In an effort to ensure that students took time to read the disclosure statement and pre-commitment in full, we integrated a timer system into the surveys so that students were not able to proceed until the minimum time had passed. Although the timers were relatively short - 10 seconds for the pre-commitment, 15 seconds for the disclosure and 20 seconds for the combined intervention - students are likely not accustomed to waiting for a timer before advancing to the next screen, and so this effect could have also contributed to a backlash in which they rated instructors lower than they otherwise would have, or once again in opposition of what is suggested in the disclosure statement or pre-commitment. In retrospect, we could have included a timer in the control condition to mitigate any such impact, or not included timers in any conditions.

Part G. Recommendations

The inconclusive results of this study suggest that other types of behavioral interventions, – for instance, implicit, experience-based interventions that demonstrate the excellence of BIPOC, women and gender-diverse faculty – could be pursued as a means to reducing students’ racial and gender bias of instructors. Because it is difficult to determine what influences the unconscious mind, especially regarding highly charged topics such as racial and gender inequities, influencing processes and evaluating interventions that advance inclusion from other fronts, such as upstream systemic changes to tenure processes, might present equally promising opportunities to mitigate racial and gender biases that negatively impact the careers of post-secondary instructors.

Although Student Evaluations of Teaching will continue to impact UBC instructors’ prospects for career advancement, it is widely established that the SEoT is an imperfect instrument with far-reaching impacts on career outcomes of UBC faculty. In recent months, UBC has taken steps to modify the SEoT, which includes changing the wording of the SEoT questions. Beginning in Fall 2021, it will take a new name - the Student Experience of Instruction (SEoI.)

The redesign to the new SEoI follows UBC’s Student Evaluation of Teaching Working Group’s 16 recommendations to the Senate which include two recommendations on dealing with bias. Most notably, Recommendation #13 states:

“UBC needs additional and regularized analysis of our own data to answer questions related to potential bias, starting with instructor ethnicity, as it is frequently highlighted as a potential source of bias in the literature on student evaluation of teaching.”

This recommendation aligns with our study’s finding that BIPOC instructors are rated significantly lower than white instructors and reiterates the need for further research to understand and mitigate student bias toward BIPOC instructors.

Kreitzer et. al (2021) reviewed a novel dataset of over 100 articles on bias in student evaluations of teaching and offer recommendations for better evaluations. UBC’s move to the new SEoI incorporates some of these recommendations, which include contextualizing evaluations as perceptions of student learning and experience, recommending minimum response rates and reporting metrics that are better measures of central

tendency and variability, such as interpolated median and dispersion index. They also recommend that administrators not rely on student evaluations as the sole method of assessing teaching. Peer evaluations, observations of teaching portfolios and reviews of course materials represent alternative methods for evaluating teaching beyond student evaluations.

More broadly, the problem of racial and gender bias in the academy runs much deeper than SEoTs. BIPOC and women instructors' opportunities for career advancement are negatively impacted by a number of factors: these include experiences of racism, sexism and other forms of discrimination. Continued and further research on behavioural insights, behavioural interventions, education and training, as well as institutional systems change, should be tested and evaluated to improve the evidence base for what works in improving career outcomes of instructors from marginalized groups.

Part H. Discussion of BI & Research Ethics

The project's research complied with the ethics protocol established by the program and ran only the pre-approved activities by the UBC Ethics Review Board Approval for Capstone Projects. In addition, the team engaged the project advisor in ongoing discussions of the ethical aspects related to qualitative and quantitative research, as well as the use of the BI solution.

The project team's considerations were as follows:

Conducting Qualitative Research (Instructor and Student Interviews)

- *Ensuring a transparent process.* The students and instructors were asked if they wished to participate in the study. The interviewer reviewed the consent forms with the interviewees, which the interviewees then signed.
- *Participation could be discontinued at any time.* The project team avoided questions that risked the emotional safety of participants since the goal of the research was merely informative. The team guaranteed that participants could abandon the interviews at any point. The interviewer reinforced the option of disconnecting from the Zoom session at any point.
- *Respectful treatment of participants.* Interviews with women and BIPOC instructors were conducted by a woman, BIPOC researcher to ensure a safer space for discussion and greater trust in the interview process.

Conducting Quantitative Research (The Intervention)

- *Ease of opt-out.* Before the intervention, students were asked if they wanted to participate in the study and if so, digitally signed the designated consent form. Some students decided not to participate and exited the study.
- *Confidentiality.* In compliance with ethical guidelines, the research team did not collect information that could have resulted in participant identification, such as names, email addresses or student ID numbers. IP addresses were only used to monitor duplicate SEoT completions. The IP addresses were deleted after this process was completed.
- *No harm principle.* The research prevented any potential negative effects resulting from the instructors' or students' participation in the research. For instructors, the midterm evaluations are not used for career development processes as per UBC policy; therefore, the results of the experiment did not directly impact career progression. Moreover, the implementation and results of this trial might

inform other researchers who are considering implementing similar experiments using end-of-term evaluations, which are typically used in instructors' performance evaluations.

- *Data security.* Survey completion was anonymous and the data was stored on a secure and private UBC server.

The BI Solution

- *Nudges for good.* According to Soman's and others' (2019) definitions of nudges, the proposed intervention made it easy to make a "good" decision. Even if the trial produced null results, the intervention aimed to improve diversity and inclusion at UBC by promoting unbiased evaluations of BIPOC, women, and gender-diverse instructors.
- *Freedom of choice was maintained.* Participating students had the option of responding to the pre-commitment device with "No". They were informed that they would not face any consequences for selecting this response or any other responses to SEoT questions. Similarly, students could ignore the disclosure statement without any consequences. The decision to participate required no more effort than the decision to not participate.
- *Transparency.* Participants knew they were participating in a study and might have identified the disclosure statement and the pre-commitment as the treatments. To preserve the integrity of the study, the consent form did not specifically mention that the trial tested a BI solution to mitigate bias in SEoTs.
- *Complied with the publicity principle.* The project team summarized the interventions for the instructors. The team will disclose the results of the study to any and all interested parties.
- *Evidence base.* The strongest available evidence indicated that the intervention would not backfire; it showed that women have benefited from a similar intervention (Peterson et al., 2019). No negative impact on other populations was reported.
- *Feasibility of scaling.* Although the team ran a rigorous evaluation through a randomized controlled trial which ensured internal validity, the null and somewhat counter-intuitive results do not support scaling the intervention. However, if the study had produced significantly positive results, the scalability of the intervention would have been feasible with the support of UBC leadership.

Part I. Project Reflections

Some of the limitations of the project that might have influenced its outcomes are:

A very unusual context. The project was executed amid the COVID-19 pandemic which impacted students and instructors in numerous ways; classes were conducted virtually, populations were socially isolated, screen fatigued, and struggling with mental health challenges. In 2020, the murder of George Floyd and many other Black Americans reignited a worldwide conversation on systemic racism, which included the urgency to improve diversity and inclusion within workplaces. The very polarized US election of November 2020 also raised the volume levels on racial and gender justice issues.

Translating identity and lived experience into quantitative data. To understand how students rated instructors, the team grouped instructors according to their race and gender - white men instructors and instructors that were not white men (BIPOC, women and gender-diverse) The lived experiences and levels of discrimination for each person in the latter group can vary greatly.

Self-selection bias. Instructors participated voluntarily in the study. Instructors that have experienced historically lower ratings (including those that are BIPOC, women and gender-diverse) may have decided not to participate to avoid exposing themselves and their negative evaluation experiences under a larger microscope.

The project faced challenges throughout implementation that were addressed with different strategies:

Establishing a variable on racial identity. As race and ethnicity are social constructs, the way people define their racial identity is not straightforward. To create a consistent variable, the team added a question on ancestry, and included non-European/white identifying participants in the BIPOC group (even if they did not self-identify as such directly.) This practice was established by Ninan Abraham, Associate Dean, Equity and Diversity of the Faculty of Science at UBC, and has become a standard practice in the university to collect more accurate data on race.

Recruiting instructors from marginalized groups. BIPOC instructors continue to be underrepresented in the academy, and some raised concerns about the possible repercussions of their participation on their careers. The project team reassured instructors that their midterm evaluations would have no direct impact on their performance evaluations while at the same time applying absolutely no pressure on them to participate.

Some lessons learned from this project might inform new interventions on unconscious bias prevention:

Consider students' perceived identity of their instructors. If it is viable from an ethical perspective, consider including a variable on how students perceive the instructors' identity to capture more accurate instructor data. For example, if an instructor self-identifies as non-binary but students perceive them to be a man, this could provide some additional insight on students' ratings of that particular instructor.

Shift focus to influencing systemic processes. It is difficult to determine what influences the unconscious mind, especially regarding highly charged topics such as racial and gender inequities. Evaluate interventions that advance inclusion from other fronts, such as systemic changes to tenure processes.

Include timers in the control condition or do not use timers at all. Students might have felt a loss of autonomy when they could not move forward as soon as the interventions were presented. To ensure this is not a factor that influences their ratings of instructors, include a timer in the control condition as well. Alternatively, do not use timers in any of the conditions.

Keep text succinct. Lengthy text might have taken a toll on students, causing confusion or overload, especially in the combined intervention. In written manipulations, it might be beneficial to keep the text shorter. Consistent text lengths also reduce the possibility of confounding effects that could result from varying lengths.

Execute trials when stakes are low. Applying the interventions during the midterm evaluations proved to be the only viable option from an ethical perspective since the manipulation might have negatively influenced ratings. Final SEoTs, unlike midterm SEoTs, impact instructors' career progression. Execute interventions in settings where potential backfire effects have minimal or no impacts.

References

- Awalt, C. J. (2003). Moving from the classroom to online teaching: a study of change in faculty attitudes. Unpublished dissertation, University of Texas, Austin. <https://www.learntechlib.org/p/125217/>.
- Axt, J.R., Casola, G., & Nosek, B.A. (2019). Reducing social Judgment biases may require identifying the potential source of bias. *Personality and Social Psychology Bulletin*, 45(8), 1232-1251. DOI: 10.1177/0146167218814003.
- Bachen, C., McLoughlin, M., & Garcia, S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education*, 48, 193-210. DOI: 10.1080/03634529909379169
- Beg, S., Fitzpatrick, A., & Lucas, A.M. (2019). Gender bias in assessments of teacher performance in Ghana. Innovations for Poverty Action. <https://www.poverty-action.org/printpdf/45546>
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41. DOI: 10.1016/j.jpubeco.2016.11.006
- Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*. DOI: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1
- Chávez, K., & Mitchell, K. M. (2019). Exploring bias in student evaluations: Gender, race, and ethnicity. *PS: Political Science & Politics*, 53(2), 270 - 274. DOI: 10.1017/S1049096519001744
- Data USA (2020). Iowa State University 2019, Enrollment by Race and Ethnicity. <https://datausa.io/profile/university/iowa-state-university#:~:text=The%20enrolled%20student%20population%20at%20Iowa%20State%20University%20is%2071.4,Hawaiian%20or%20Other%20Pacific%20Islanders.>
- English, M., & Kalla, J. (2021). Racial equality frames and public policy support: Survey experimental evidence. DOI: 10.31219/osf.io/tdkf3
- Hakstian, A. R. (2009). An investigation into the effects of instructor gender, field of study, and student-responder gender on UMI scores in the 2008-09 SEoT administration. *UBC Vancouver Senate Report*. May 27, 2020, (pp. 234-249).
- Hoorens, V., Dekkers, G. & Deschrijver, E. (2021). Gender bias in student evaluations of teaching: Students' self-affirmation reduces the bias by lowering evaluations of male professors. *Sex Roles*, 84, 34-48. DOI: 10.1007/s11199-020-01148-8
- Jackson, L.C., Jones, S.J., Rodriguez, R.C. (2010). Faculty actions that result in student satisfaction in online courses. *Journal of Asynchronous Learning Networks*, 14(4), 78-96.
- Khazan, E., Borden, J., Johnson, S., & Greenhaw, L. (2019). Examining gender bias in student evaluations of teaching for graduate teaching assistants. *NACTA Journal*, 430-435. https://www.researchgate.net/publication/345178456_Examining_Gender_Bias_in_Student_Evaluations_of_Teaching_for_Graduate_Teaching_Assistants

- Kogan, L.R., Schoenfeld-Tacher, R. & Hellyer, P.W. (2010). Student evaluations of teaching: Perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, 15(6), 623-636. DOI:10.1080/13562517.2010.491911
- Kreitzer, R., & Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*. DOI: 10.1007/s10805-021-09400-w
- Kristal, A., Whillans, A., Bazerman, M., Gino, F., Shu, L., Mazar, N., & Ariely, D. (2020). Signing at the beginning versus at the end does not decrease dishonesty. *Proceedings of the National Academy of Sciences the United States of America*, 117(13), 7103-7107. DOI: <https://doi.org/10.1073/pnas.1911695117>
- La Touche, R., Kowalchuk, L., & Wijesingha, R. (2020). *(Re)prioritizing pedagogic feedback: Faculty experiences with qualitative comments from SETs*. Canadian Sociological Association.
- London, B., Downey, G., Romero-Canyas, R., Rattan, A., & Tyson, D. (2012). Gender-based rejection sensitivity and academic self-silencing in women. *Journal of Personality and Social Psychology*, 102(5), 961–979. DOI: 10.1037/a0026615
- MacNell, L., Driscoll, A. & Hunt, A.N. (2015). What’s in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40, 291-303. DOI: 10.1007/s10755-014-9313-4
- Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535-566. DOI: 10.1093/jeea/jvx057
- Mitchell, K. M., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(03), 648-652. DOI: 10.1017/s104909651800001x
- Peterson D. A. M., Biederman L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS ONE* 14(5): e0216241. DOI: 10.1371/journal.pone.0216241
- Reid, L. D. (2010). The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.Com. *Journal of Diversity in Higher Education*, 3(3), 137–152. <https://doi.org/10.1037/a0019865>
- Reinsch, R., Goltz, S., & Hietapelto, A. (2020). Student evaluations and the problem of implicit bias. *Journal of College and University Law*, 45(1). (pp. 114-140)
- Roblyer, M. D., & Ekhaml, L. (2020) How interactive are your distance courses? A rubric for assessing interaction in distance learning. *Online Journal of Distance Learning Administration*, 3. <http://www.westga.edu/%7Edistance/roblyer32.html>
- Shea, P. J., Fredericksen, E. E., Pickett, A. M., & Pelz, W. E. (2004). Faculty development, student satisfaction, and reported learning in the SUNY learning network. In T. M. Duffy & J. R. Kirkley (Eds.), *Learner-Centered Theory and Practice in Distance Education*: (pp.343-377). DOI: DOI:10.4324/9781410609489-27

Shu, L., Mazar, N., Gino, F., Bazerman, M., & Ariely, D. (2012). *Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. Proceedings of the National Academy of Sciences*. 109(38), 15197–15200.

Social and Behavioral Sciences Team. (2015). Annual Report.

https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/sbst_2015_annual_report_final_9_14_15.pdf.

Soman, D., Cowen, D., Kannan, N., & Feng, B. (2019). Seeing sludge: Towards a dashboard to help organizations recognize impedance to end-user decisions and action [White paper].

<https://www.rotman.utoronto.ca/-/media/Files/Programs-and-Areas/BEAR/White-Papers/BEARxBIOrg-Seeing-Sludge-1.pdf?la=en&hash=5CAB338A32025E08D366F4297AF4F59EABC8781D>

Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. DOI: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Summers, J.J., Waigandt, A. & Whittaker, T.A. (2005). A comparison of student achievement and satisfaction in an online versus a traditional face-to-face statistics class. *Innovative Higher Education*, 29, 233–250. DOI: 10.1007/s10755-005-1938-x

Szeri, A. (2020). Extension of appointment of Dr. Minelle Mahtani, Senior Advisor to the Provost on Racialized Faculty. <https://academic.ubc.ca/academic-community/appointments/extension-appointment-dr-minelle-mahtani-senior-advisor-provost>

Szeri, A. (2020). Naznin Virji-Babul appointed Senior Advisor to the Provost on Women and Gender-Diverse Faculty. <https://academic.ubc.ca/academic-community/appointments/naznin-virji-babul-appointed-senior-advisor-provost-women-and-gender>

UBC Centre for Teaching, Learning & Technology (2015). Examining the effect of field of study and gender on Students' Evaluation of Teaching (SEoT): A case study of the University Module Items (UMI) scores in the 2014-2015 academic year. *UBC Vancouver Senate Report*. May 27, 2020, (pp. 250-256).

UBC Equity and Inclusion Office (2020). About Us. <https://equity.ubc.ca/about/>

UBC Okanagan Planning and Institutional Research (2017). SET gender analysis: 2015/16 lecture evaluations. *UBC Vancouver Senate Report*. May 27, 2020, (pp. 228-230).

UBC Okanagan Planning and Institutional Research (2020). SEoT gender bias analysis: 2018W1 and 2018W2 lecture evaluations. *UBC Vancouver Senate Report*. May 27, 2020, (pp. 231-233).

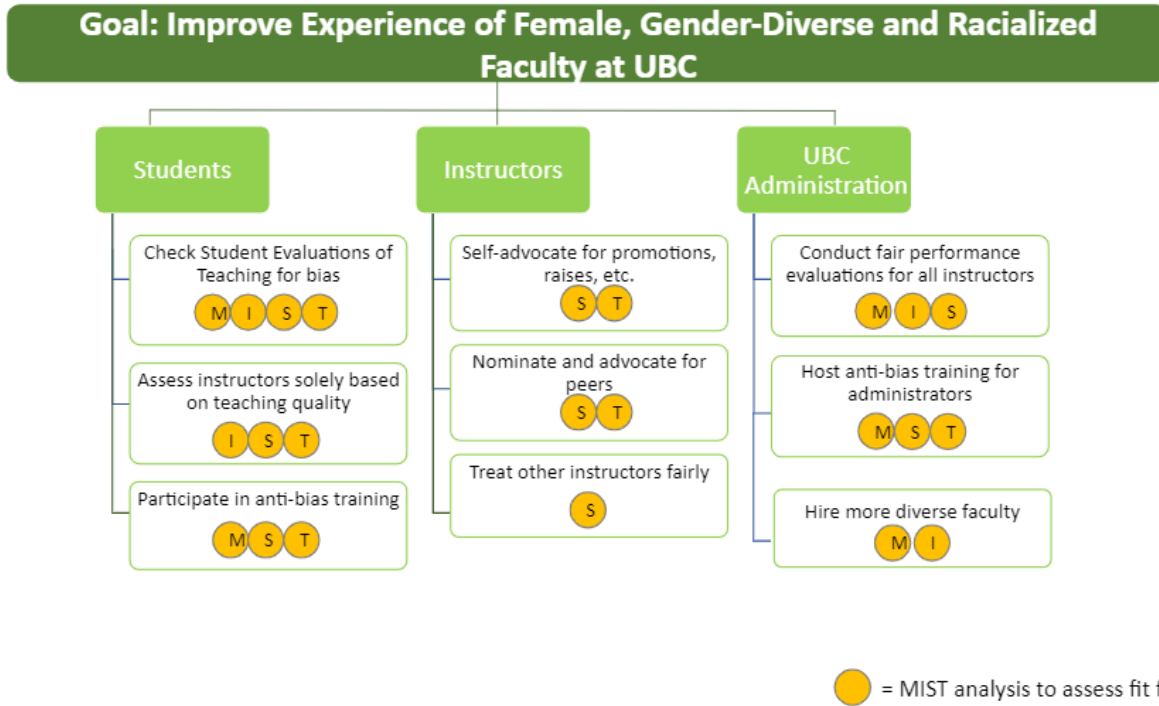
UBC Student Evaluation of Teaching Working Group (2020). Report to Vancouver and Okanagan Senates, *Vancouver Senate Report*. May 27, 2020, (pp. 180-265).

Wagner, N., Rieger, M., & Voorvelt, K. (2016). Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams. *Economics of Education Review*, 54, 79–94. DOI: 10.1016/j.econedurev.2016.06.004

Appendices

Appendix I. Behaviour Map

Behaviour Map - Part A (Problem Background)



Appendix II. Additional Secondary Research

The effect of implicit bias against women is problematic not only at the university instructor level, but even among teaching assistants (TAs). Using a similar “disguised identity” approach as MacNell, Driscoll and Hunt (2014), Khazan et al. (2019) found that unconscious bias exists even at the teaching assistant (TA) level, with women TAs receiving lower scores than men TAs, even though all TA work was actually completed by a woman. This discouragement early on in one’s career may create a barrier that prevents women TAs from pursuing a career in academia or as a university instructor in the future. This is relevant to existing work on rejection sensitivity (London et al., 2012), where perceptions of gender-based threats are addressed and coped with using self-silencing, which reinforces feelings of alienation and diminished motivation.

In addition to students’ implicit bias potentially having an effect on the careers of BIPOC, women and gender-diverse university instructors, it could also be causing adverse effects on their mental health. Mitchell and Martin (2018) studied the language students use in evaluations and found that the language used to evaluate men faculty is considerably different than the language used to evaluate women faculty. Rea (2018) uncovered that some students use SEoTs to make disrespectful or abusive comments. While the most common theme was around instructors’ competency to teach a subject, another theme was around gender, background and spoken English. Kogan, Schoenfeld-Tacher and Hellyer (2010) examined qualitative data to show that student comments affected women instructors more negatively than men instructors. La Touche, Kowalchuk and Wijesingha (2020) concluded that equity-deserving instructors who received negative comments which were irrelevant to teaching quality experienced negative effects on their well-being and professional self-esteem.

There can even be bias in evaluating instructors amongst peers. At the elementary school level, Beg, Fitzpatrick and Lucas (2019) found that school principals in Ghana were less likely to rate women teachers as effective, even though they were in fact more effective on an objective measure (student test scores.)

Several studies even conclude the SEoT is an inappropriate method of evaluating instructors, especially when it is the only instrument used during an instructor’s performance review. Some studies indicate that SEoTs do not predict actual student learning outcomes, and are therefore a poor indicator of teaching quality (Boring, Ottoboni, Stark, 2016; Stark & Freishtat, 2014). Reinsch, Goltz and Hietapelto (2020) also suggest that SEoTs should not be the only measure of an instructor’s evaluation, and that multiple evaluation methods should be adopted for a more holistic approach. Although SEoTs are here to stay in the short-term, these are findings that could be considered for long-term planning.

Appendix III. Instructor Recruitment Questionnaire

[CONSENT FORM]

QConsent.

Behavioural Insights Certificate Course Project Study Overview and Consent

Welcome! This is a survey about you and your courses for the January to April 2021 term. We are conducting this survey as part of a Behavioural Insights class project at UBC Sauder.

The survey should take about 3 to 5 minutes. You will be asked questions about your courses, as well as demographic information. These questions are important for designing the study logistics and analyzing the data. Your answers will only be used to facilitate your participation in the study, and any personal information collected will be reported in aggregate. You may withdraw from the survey at any time by closing your browser window. If you decide that you wish to withdraw from the study after you complete this survey, you can do so by sending an email to the student project leader (contact information below).

Data will be stored on the advising professor's encrypted, password-protected computer for a period of at least six months.

If you have any questions or complaints, you may contact any of the following:

Student Project Leader: Greg Lockwood, greg.lockwood@ubc.ca

Advising Professor and Principal Investigator (PI): Dr. Jiaying Zhao, jiayingz@psych.ubc.ca

If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

Clicking the button below indicates that you consent to participate in this study.

Q1. Your first and last name:

[OPEN-END TEXT BOX]

Q2. Your email address:

[OPEN-END TEXT BOX] [EMAIL FORMAT VALIDATION]

Q3. Please list the **courses and sections** you would like to include in the study. (e.g., APSC 100, Section 103)

If you are teaching the same course for several sections, please enter the course and section numbers on different lines. For example, if you are teaching COMM123 sections 101 and 102, please enter them on separate lines:

-Course A: COMM123

-Section for Course A: 101

-Course B: COMM123

-Section for Course B: 102

Course: [OPEN-END TEXT BOX]

2021-CBI-06

Section: [OPEN-END TEXT BOX, NUMERICAL VALIDATION]

[INSERT TEXT BOXES FOR 5 COURSES/SECTIONS. AT LEAST ONE COURSE/SECTION MUST BE ENTERED]

Q4. Please indicate the approximate number of students in each course and section. *Select one for each.*

[ROWS - INSERT COURSES AND SECTIONS FROM Q3]

[COLUMNS]

1-24
25-49
50-99
100-149
150-199
200+

[NEW SCREEN - D1 TO D5 ON THE SAME PAGE]

D1. How long have you been a post-secondary instructor at UBC or any other institutions? Enter the number of years and months below.

If less than one year, enter '0' for 'Years' and enter the number of months in 'Months'.

Years: [NUMERIC RANGE 0-70]

Months: [NUMERIC RANGE 0-11]

D2. Do you identify as a woman, man or non-binary person?

Woman

Man

Non-Binary Person

Prefer not to answer

D3. Do you identify as someone who is Black, Indigenous or a Person of Colour (BIPOC)? *Select one.*

Yes

No

Prefer not to answer

D4. How do you identify your ancestry? *Select all that apply.*

- African/Black (e.g., African, African-American, African-Canadian, Afro-Caribbean, etc.)
- Arab (e.g., Algerian, Lebanese, Tunisian, etc.)
- East Asian (e.g., Chinese, including Hong Kong and Macau, Japanese, Korean, etc., and including Asian-Canadian, Asian-American, etc.)
- European/Non-white (e.g., Roma, etc.)
- European/White (e.g., Belgian, Croatian, English, Spanish, etc.)
- Filipina/Filipino
- Indigenous (outside of North America)
- Indigenous (within North America)
- Latin, South or Central American (e.g., Brazilian, Chilean, Colombian, Mexican, etc.)
- South Asian (e.g., Indian, Pakistani, Sri Lankan, etc., and including Indo-Caribbean, Indo-African, Indo-

Fijian, West Indian, etc.)

- Southeast Asian (e.g., Cambodian, Indonesian, Laotian, Vietnamese, etc.)
- West Asian (e.g., Afghan, Iranian, etc.)
- If none of the above, please specify: _____
- Prefer not to answer [**Please note:** If you choose this response, all of your other responses to this question will not be considered in the data analysis.] [**EXCLUSIVE**]

D5. Do you currently have children or dependents at home? *Select one.*

Yes

No

[NEW SCREEN]

D6. Below are the 6 standard questions that will be asked on the midterm Student Evaluation of Teaching (SEoT):

1. The instructor made it clear what I was expected to learn.
2. I think that the instructor communicated the subject matter effectively.
3. The instructor engaged me in the subject matter.
4. I have received feedback that supported my learning.
5. I think that the instructor showed concern for student learning.
6. Overall, this instructor was effective in helping me learn.

Rating scale: Strongly Agree (5), Agree (4), Neutral (3), Disagree (2), Strongly Disagree (1)

Would you like to include any additional questions in the student evaluation form specific to you or your course?

If so, please provide specific instructions below, including the question text, answer options, single-choice or multi-choice response, etc. We will contact you if anything requires clarification.

If you have no additional questions to include, please leave the field blank.

[OPEN-END PARAGRAPH, RESPONSE OPTIONAL]

[NEW SCREEN]

D7. Do you have any additional comments about the study?

[OPEN-END PARAGRAPH, RESPONSE OPTIONAL]

[EXIT SCREEN]

Thank you for your participation! We will follow up in early-mid February with instructions for distributing the midterm SEoT to your students.

If you have any questions about the study, please contact Greg Lockwood at greg.lockwood@ubc.ca.

Appendix IV. Student Evaluation of Teaching Survey

[CONSENT FORM]

QConsent.

Behavioural Insights Certificate Course Project Study Overview and Consent

Welcome! This is a survey about one of your specific courses for the January to April 2021 term. We are conducting this survey as part of a Behavioural Insights class project at UBC Sauder.

The survey should take about 3 minutes. You will be asked questions about your course, as well as demographic information. These questions are important for designing the study logistics and analyzing the data. Your answers will only be used to facilitate your participation in the study, and any personal information collected will be reported in aggregate. You may withdraw from the survey at any time by closing your browser window. If you decide that you wish to withdraw from the study after you complete this survey, you can do so by sending an email to the student project leader (contact information below).

Data will be stored on the advising professor's encrypted, password-protected computer for a period of at least six months.

If you have any questions or complaints, you may contact any of the following:

Student Project Leader: Greg Lockwood, greg.lockwood@ubc.ca

Advising Professor and Principal Investigator (PI): Dr. Jiaying Zhao, jiayingz@psych.ubc.ca

If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

If you would like to participate in the study, click 'Yes, I consent to participating in the study'.

If you do not wish to participate in the study, click 'No, I do not consent to participating in the study'.

Yes, I consent to participating in the study

No, I do not consent to participating in the study

[ONLY SHOW RCT1, RCT2, RCT3, CTRL IF SELECTED 'YES' AT QCONSENT. RANDOMLY ASSIGN RESPONDENT TO ONE OF THE FOLLOWING: RCT1, RCT2, RCT3, CTRL]

RCT1. Please read the text below carefully. You will be able to proceed after 15 seconds.

Student Evaluations of Teaching (SEoTs) directly impact the careers of UBC instructors. [Evidence indicates](#) that SEoTs are often influenced by students' unconscious and unintentional biases about the race and gender of the instructor. Black, Indigenous and People of Color (BIPOC), women and gender-diverse instructors are systematically rated lower in their teaching evaluations than white men, even when there are no actual differences in the instruction or in what students have learned.

As you fill out the course evaluation keep this in mind and make an effort to resist stereotypes about instructors. Focus on your opinions about the instructor's teaching and not unrelated matters (for example, the instructor's appearance or style).

Your responses to this survey are completely anonymous and will not affect your grades.

RCT2. Please read the text below carefully. You will be able to proceed after 10 seconds.

Your responses to this survey are completely anonymous and will not affect your grades.

Do you commit to completing a fair and honest evaluation of this instructor based on the quality of the instruction you received?

Yes, I commit

No, I do not commit

RCT3. Please read the text below carefully. You will be able to proceed after 20 seconds.

Student Evaluations of Teaching (SEoTs) directly impact the careers of UBC instructors. [Evidence indicates](#) that SEoTs are often influenced by students' unconscious and unintentional biases about the race and gender of the instructor. Black, Indigenous and People of Color (BIPOC), women and gender-diverse instructors are systematically rated lower in their teaching evaluations than white men, even when there are no actual differences in the instruction or in what students have learned.

As you fill out the course evaluation keep this in mind and make an effort to resist stereotypes about instructors. Focus on your opinions about the instructor's teaching and not unrelated matters (for example, the instructor's appearance or style).

Your responses to this survey are completely anonymous and will not affect your grades.

Do you commit to completing a fair and honest evaluation of this instructor based on the quality of the instruction you received?

Yes, I commit

No, I do not commit

CTRL. Your responses to this survey are completely anonymous and will not affect your grades.

[ASK ALL]

QSEoT. [GRID]

To what extent do you agree or disagree with the following statements about your instructor **[INSTRUCTOR NAME]** for **[COURSE & SECTION]**? *Select one for each.*

[ROWS - DO NOT RANDOMIZE]

The instructor made it clear what I was expected to learn.

I think that the instructor communicated the subject matter effectively.

The instructor engaged me in the subject matter.

I have received feedback that supported my learning.

I think that the instructor showed concern for student learning.
Overall, this instructor was effective in helping me learn.

[COLUMNS]

- Strongly Disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly Agree (5)

[NEW SCREEN - ONLY SHOW SELF_REFL IF SELECTED 'YES' AT QCONSENT.]

SELF_REFL. How fair and honest do you feel your evaluation of this instructor was, based on the quality of the instruction you received? *Select one.*

- Extremely fair
- Very fair
- Somewhat fair
- Not very fair
- Not at all fair

[INSERT CUSTOM INSTRUCTOR QUESTIONS (IF ANY)]

[NEW SCREEN - INTRODEMO AND D1 TO D3 ON THE SAME PAGE. ONLY SHOW INTRODEMO AND D1 TO D3 IF SELECTED 'YES' AT QCONSENT.]

INTRODEMO. Finally, please respond to the following demographic questions.

D1. Do you identify as a woman, man or non-binary person? *Select one.*

- Woman
- Man
- Non-Binary Person
- Prefer not to answer

D2. Do you identify as someone who is Black, Indigenous or a Person of Colour (BIPOC)? *Select one.*

- Yes
- No
- Prefer not to answer

D3. How do you identify your ancestry? *Select all that apply.*

- African/Black (e.g., African, African-American, African-Canadian, Afro-Caribbean, etc.)
- Arab (e.g., Algerian, Lebanese, Tunisian, etc.)
- East Asian (e.g., Chinese, including Hong Kong and Macau, Japanese, Korean, etc., and including Asian-Canadian, Asian-American, etc.)
- European/Non-white (e.g., Roma, etc.)
- European/White (e.g., Belgian, Croatian, English, Spanish, etc.)
- Filipina/Filipino
- Indigenous (outside of North America)
- Indigenous (within North America)

- Latin, South or Central American (e.g., Brazilian, Chilean, Colombian, Mexican, etc.)
- South Asian (e.g., Indian, Pakistani, Sri Lankan, etc., and including Indo-Caribbean, Indo-African, Indo-Fijian, West Indian, etc.)
- Southeast Asian (e.g., Cambodian, Indonesian, Laotian, Vietnamese, etc.)
- West Asian (e.g., Afghan, Iranian, etc.)
- If none of the above, please specify: _____
- Prefer not to answer **[EXCLUSIVE]**

[END OF SURVEY MESSAGE]

Thank you for your participation!

Appendix V. Data Cleaning Rules

If duplicate surveys are suspected, remove if:

1. Surveys have duplicate IP addresses
2. Surveys have the exact same answers for all six SEoT questions
3. Surveys have the exact same demographic answers (e.g., age, gender, BIPOC identification, ethnic ancestry)

All three conditions must be met to delete any survey responses from the data set. If surveys meet only one or two of the conditions, the response is kept.

The project team removed 1 duplicate survey from the data set based on the conditions above.

Pre-commitment: If students responded 'No' to the pre-commitment question in the pre-commitment condition or combined condition, the response was deleted.

The project team did not remove any surveys from the data set based on this criterion.

Appendix VI. Additional Exploratory Analyses

Analyzing Student Fairness Ratings. Students were asked how fair and honest they felt their evaluation of their instructor was, based on the quality of the instruction received. To explore any relationships between instructor identity, student identity and the four SEoT conditions on students' perceived levels of fairness and honesty, a 2 (instructor identity: BIPOC/women/gender-diverse vs. white/men) x 2 (student identity: BIPOC/women/gender-diverse vs. white/men) x 4 (condition: disclosure vs. pre-commitment vs. combination vs. control) ANOVA on this self-reflection question was analyzed.

As shown in Table A1, there was again a main effect of student identity ($p = .015$), indicating that white men students ($M = 4.25$, $SD = 0.58$) rated their evaluations as fairer and more honest than BIPOC/women/gender-diverse students ($M = 4.02$, $SD = 0.71$). There were no other significant main effects or interactions from this analysis.

Taken together, the results shown in Tables 2 (from Part F above) and A1 indicate that white men students rate instructors higher than BIPOC/women/gender-diverse students and that white men students associate their higher ratings with increased honesty and fairness.

Table A1. Results of a 2X2X4 ANOVA comparing instructor identity, student identity and the four SEoT conditions on the self-reflection question.

Predictor	<i>df</i>	<i>F</i>	<i>p</i> -value	η_p^2
Instructor Identity	1, 885	0.856	.355	.001
Student Identity	1, 885	5.889	.015	.007
Condition	3, 885	0.685	.561	.002
Instructor Identity X Student Identity	1, 885	0.203	.652	.0002
Instructor Identity X Condition	3, 885	0.480	.696	.002
Student Identity X Condition	3, 885	0.978	.402	.003
Instructor Identity X Student Identity X Condition	3, 885	0.459	.711	.002

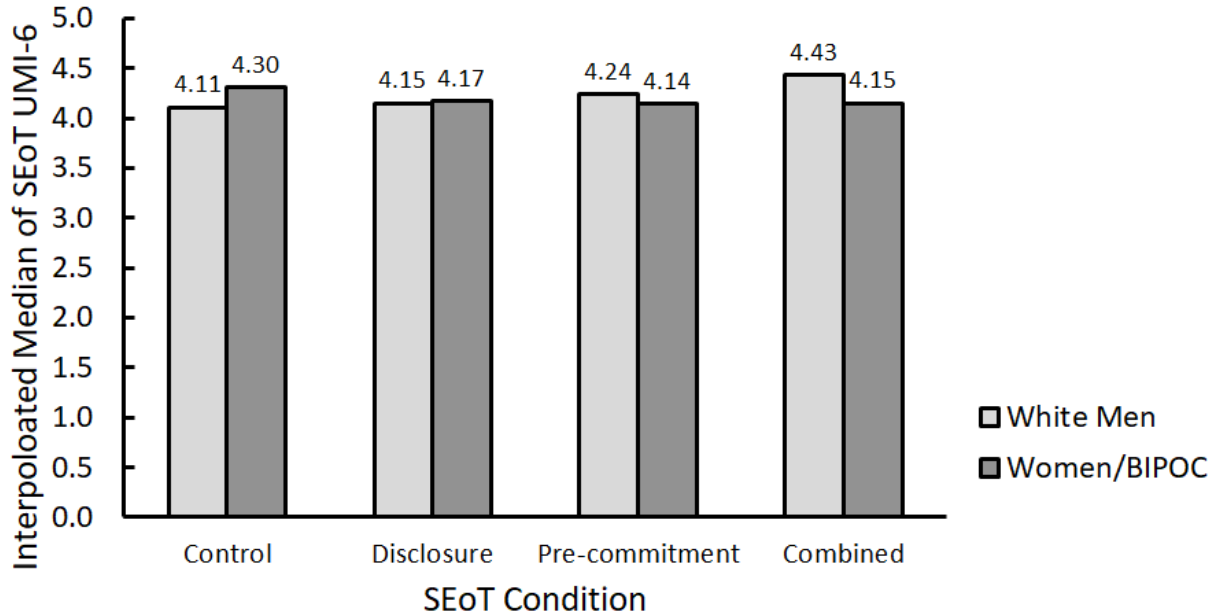
Interpolated Medians of BIPOC/Women/Gender-Diverse and White Men Instructors

To describe the UMI-6 data, we calculated the mean and the interpolated median for both the BIPOC/women/gender-diverse and white men instructors in each of the four conditions. UBC will be replacing the mean with the interpolated median in reporting SEoT ratings because the interpolated median better takes into account all respondents' feedback when estimating the central tendency of the response data for Likert-scale response items.

Figure A1 illustrates the interpolated median (IM) of the UMI-6 scores by BIPOC/women/gender-diverse and white man identity in the four conditions. The IM is calculated as the median M , plus the difference between n -plus (the number of data points greater than the median) and n -minus (the number of data points less than the median) divided by $2n$ (twice the number of data points equal to the median):

$$IM = M + (n\text{-plus} - n\text{-minus}) / 2n.$$

Figure A1. Interpolated Median of UMI-6 ratings for women/BIPOC and white men instructors in the four SEoT conditions



Notably, and mirroring the results for the analyses on the mean score, the IM of UMI-6 for BIPOC/women is higher (IM = 4.30) than that of white men (IM = 4.10) in the control condition but lower than that for white men in each of the other conditions. The difference is most pronounced in the combined condition where the IM for BIPOC/women equals 4.15 and the IM for white men equals 4.43. This is an indication that each of the three treatment conditions did not have the intended effect of mitigating students’ bias and increasing the ratings of BIPOC/women/gender-diverse instructors more than white men instructors.

UMI-1 to UMI-5

The analysis that tested the three hypotheses on UMI-6 were also tested on UMI-1 to UMI-5. As shown in Tables A2a to A2e, each interaction produced a null result. There was one significant main effect (after taking into account a Bonferroni correction): white men were rated significantly higher than BIPOC, women and gender-diverse instructors when students were presented with the pre-commitment condition in UMI-4, $F(1, 479) = 6.496, p = .011, corrected p = .033, partial \eta^2 = .013$.

Table A2a. Results of a 2X2 ANOVA comparing ratings on UMI-1 of BIPOC/women/gender-diverse instructors with white men instructors in each of the treatment conditions to the control condition.

Predictor		df	F	p-value	η_p^2
Disclosure Statement	Instructor Identity	1, 464	0.083	.773	.0002
	Condition	1, 464	0.665	.415	.001
	Instructor Identity X Condition	1, 464	0.097	.756	.0002
Pre-commitment	Instructor Identity	1, 479	0.788	.375	.002

	Condition	1, 479	0.073	.787	.0002
	Instructor Identity X Condition	1, 479	0.826	.364	.002
Combined (Disclosure + Pre-commitment)	Instructor Identity	1, 440	0.436	.509	.001
	Condition	1, 440	0.127	.722	.0003
	Instructor Identity X Condition	1, 440	0.409	.523	.0009

Table A2b. Results of a 2X2 ANOVA comparing ratings on UMI-2 of BIPOC/women/gender-diverse instructors with white men instructors in each of the treatment conditions to the control condition.

Predictor		<i>df</i>	<i>F</i>	<i>p</i> -value	η_p^2
Disclosure Statement	Instructor Identity	1, 464	0.370	.543	.0008
	Condition	1, 464	0.376	.540	.0008
	Instructor Identity X Condition	1, 464	0.002	.967	.000004
Pre-commitment	Instructor Identity	1, 479	0.002	.962	.000005
	Condition	1, 479	0.308	.579	.0006
	Instructor Identity X Condition	1, 479	0.354	.552	.0007
Combined (Disclosure + Pre-commitment)	Instructor Identity	1, 440	1.608	.206	.004
	Condition	1, 440	0.005	.944	.00001
	Instructor Identity X Condition	1, 440	0.526	.469	.001

Table A2c. Results of a 2X2 ANOVA comparing ratings on UMI-3 of BIPOC/women/gender-diverse instructors with white men instructors in each of the treatment conditions to the control condition.

Predictor		<i>df</i>	<i>F</i>	<i>p</i> -value	η_p^2
Disclosure Statement	Instructor Identity	1, 464	0.058	.810	.0001
	Condition	1, 464	0.597	.440	.001
	Instructor Identity X Condition	1, 464	0.045	.833	.0001
Pre-commitment	Instructor Identity	1, 479	0.016	.900	.00003
	Condition	1, 479	0.038	.846	.00008
	Instructor Identity X Condition	1, 479	0.339	.339	.0007
Combined (Disclosure	Instructor Identity	1, 440	0.977	.323	.002

+ Pre-commitment)	Condition	1, 440	1.141	.286	.003
	Instructor Identity X Condition	1, 440	2.111	.147	.005

Table A2d. Results of a 2X2 ANOVA comparing ratings on UMI-4 of BIPOC/women/gender-diverse instructors with white men instructors in each of the treatment conditions to the control condition.

Predictor		<i>df</i>	<i>F</i>	<i>p</i> -value	η_p^2
Disclosure Statement	Instructor Identity	1, 464	3.846	.050	.008
	Condition	1, 464	0.654	.419	.001
	Instructor Identity X Condition	1, 464	0.009	.926	.00002
Pre-commitment	Instructor Identity	1, 479	6.496	.011	.013
	Condition	1, 479	0.004	.948	.000009
	Instructor Identity X Condition	1, 479	0.203	.653	.0004
Combined (Disclosure + Pre-commitment)	Instructor Identity	1, 440	2.127	.145	.005
	Condition	1, 440	0.936	.334	.002
	Instructor Identity X Condition	1, 440	0.269	.604	.0006

Table A2e. Results of a 2X2 ANOVA comparing ratings on UMI-5 of BIPOC/women/gender-diverse instructors with white men instructors in each of the treatment conditions to the control condition.

Predictor		<i>df</i>	<i>F</i>	<i>p</i> -value	η_p^2
Disclosure Statement	Instructor Identity	1, 464	0.940	.333	.002
	Condition	1, 464	1.986	.159	.004
	Instructor Identity X Condition	1, 464	0.543	.462	.001
Pre-commitment	Instructor Identity	1, 479	0.028	.868	.00006
	Condition	1, 479	0.294	.588	.0006
	Instructor Identity X Condition	1, 479	0.163	.687	.0003
Combined (Disclosure + Pre-commitment)	Instructor Identity	1, 440	0.365	.546	.0008
	Condition	1, 440	0.015	.904	.00003
	Instructor Identity X Condition	1, 440	0.138	.710	.0003

Teaching Experience

We also controlled for the amount of teaching experience of instructors (in years) into our analysis to construct an ANCOVA model. With the exception of teaching time in the pre-commitment vs. control comparison, the model again concluded that there were no effects of the interventions on student ratings of UMI-6 by instructor BIPOC/women/gender-diverse or white man identity. These results are displayed in Table A3.

Table A3. Results of a 2X2 ANCOVA comparing ratings on UMI-6 of BIPOC/women/gender-diverse instructors with white men instructors in each of the treatment conditions to the control condition

Comparison with the Control Condition	Case	df (Degrees of Freedom)	F	p-value	Corrected p-value	η_p^2 (Partial eta squared)
Disclosure Statement	Identity Interaction	1	1.462	.227	0.681	.003
	Total Time Teaching	1	1.566	.211	0.633	.003
Pre-commitment	Identity Interaction	1	1.763	.185	0.555	.004
	Total Time Teaching	1	6.951	.009	0.027	.014
Combined Condition	Identity Interaction	1	4.649	.032	0.096	.010
	Total Time Teaching	1	0.708	.401	1.203	.002

Courses by Faculty

To explore any differences in SEoT ratings across the conditions as they related to the faculty within which the courses are situated, we classified the courses into three main categories: 1) Arts (mostly Psychology courses), 2) Sauder and 3) Science & Applied Science. Again, the analysis indicates that the intervention did not impact ratings on the basis of the BIPOC/women/gender-diverse or white man identity. The results are illustrated in Table A4 below.

Table A4. Results of a 2X2 ANOVA comparing ratings on UMI-6 of BIPOC/women/gender-diverse instructors with white men instructors in each of the treatment conditions to the control condition across three faculty categories.

Faculty	Comparison with the Control Condition	df (Degrees of Freedom)	F	p-value	Corrected p-value	η_p^2 (Partial eta squared)
Arts (n=273)	Disclosure Statement	1	0.584	.446	1.338	.004

	Pre-commitment	1	0.224	.637	1.911	.002
	Combined Condition	1	0.370	.544	1.632	.003
Sauder (n=151)	Disclosure Statement	1	0.214	.645	1.935	.003
	Pre-commitment	1	0.026	.873	2.619	.0003
	Combined Condition	1	0.434	.512	1.536	.006
Science & Applied Science (n = 495)	Disclosure Statement	1	0.603	.438	1.308	.002
	Pre-commitment	1	1.530	.163	0.489	.007
	Combined Condition	1	2.421	.121	0.363	.010

Self-Reflection Question

As a manipulation check on the self-reflection question, we conducted a one-way ANOVA across the four conditions to determine if students in the pre-commitment conditions rated themselves with a different level of fairness and honesty than students that were not in the pre-commitment conditions. With $F(3) = 0.571$ and $p = .634$, this was not the case.

We verified this result by conducting independent samples t-tests between each of the two pre-commitment conditions and the control condition. For the pre-commitment only condition, $t(479) = 1.108$, $p = .269$, while for the combined condition, $t(438) = -0.066$, $p = .948$. These results verified the finding that students did not rate themselves differently on this question across the four conditions. In other words, contrary to our predictions, students who pre-committed to being fair and honest did not rate their SEoTs as being fairer and more honest than other students. This provides further evidence that the pre-commitment condition did not impact how students rated their instructors and helps to explain the null results.