

(Please cite the published version)

“The Many Harms of SETs in Higher Education”

Symposion: Theoretical and Applied Inquiries in Philosophy and Social Sciences. Special issue: Diversity in Philosophy. Guest Editors, Helen Beebee and Anne-Marie McCallion

Cecilea Mun

Independent Scholar, cecileamun@icloud.com

There are many reasons to be angry currently in the United States: On September 20, 2020, the U.S. surpassed 200,000 deaths due to COVID-19 (Chappell 2020), and a disproportionate number of those who have died are Black and Latinx members of the society (Ford, Reber, and Reeves 2020). The U.S. Bureau of Labor Statistics reported that in July the unemployment rate was at 10.2%,¹ with Black and Latinx community members constituting the highest unemployed racial groups. In May 2020, during the last month in which lockdowns were near universally imposed across the country, the unemployment rate for women was 14.3% compared to 11.9% for men, with Hispanic women (19.5%) having the highest rate of unemployment in the nation compared to the rate for women and men of any other major racial or ethnic group; and the unemployment rate among those aged 16-24 also rose to 25.3%, which was more than double the rate of unemployment for those who were 35 and older (Kochhar 2020). The main contributing causes were the global COVID-19 pandemic and the Trump administration’s failure to take appropriate actions.² Another consequence of these two causes is that approximately 30-40 million people, with 80% being people of color and especially Black and Latinx people, are currently at risk of eviction (Benfer et. al 2020). Finally, the abuse, violence, and injustice that Black, Latinx, and Native Americans—both men and women—have suffered throughout the years at the hands of local, state, and federal police officers have inspired Black Lives Matter protests at approximately 4,395 locations around the world, since May 25, 2020 and as of August 1, 2020, but mostly in the United States.³

That such harms are ultimately a consequence of racial and economic injustice is obvious, and it is right for one to be angry about such injustices. What is not obvious, however, are the pernicious harms to marginalized groups—such as women, people of color, genderqueer people, and transgender people—and ultimately a society as a whole, of the continued endorsement and use of student evaluation of teaching (SET) instruments for hiring, promotion, reappointment, and recognition practices in higher education. Yet these harms are as or even more detrimental than those gender and racial injustices that give rise to systematic inequalities in health, employment, housing, and the law. One reason why is that these harms work to undercut the distribution of knowledge, skills, and epistemic authority within a society—those

¹ U.S. Bureau of Labor Statistics. 2020. "Employment Situation Summary." Economic News Release. Last updated August 2020. Accessed August 23, 2020. <https://www.bls.gov/news.release/empsit.nr0.htm>.

² In contrast, the unemployment rate in July of the previous year was 3.7%. Read U.S. Bureau of Labor Statistics. 2019. “The Employment Situation — July 2019.” News Release. Accessed August 2, 2019. https://www.bls.gov/news.release/archives/empsit_08022019.pdf.

³ Black Lives Matter Protests 2020. Online Map. <https://www.creosotemaps.com/blm2020/>.

(Please cite the published version)

goods that ultimately mediate access to healthcare, employment, housing, and equality before the law—while at the same time making at least some of the victims of such harms not only willing, but sometimes eager and enthusiastic accomplices in reinforcing the systemic injustices that cause their oppression. In this paper I call attention to the problem of continuing to rely on SETs for hiring, reappointment, promotion, and award decisions in higher education, including the problem of continuing to permit the use of SETs despite the clear and explicit acknowledgement of their problems. I argue that to do so manifests a failure to acknowledge the weight of the actual and potential harms of doing so. I then provide an outline of such harms in order to clearly convey not only the weight but the extent of such harms, especially on marginalized job candidates, administrators, pre-existing faculty, and non-privileged students. I conclude by arguing that, given all of the foregoing, the use of SETs should be abolished for hiring, reappointment, promotion, and award decisions in higher education.

I. Illusions/Delusions of Just Concern with SET Use in Higher Education

The use of SETs in U.S. higher education goes as far back as the early 1920s (Stroebe 2020; Uttl, White, and Gonzalez 2017; Adams and Umbach 2012; Carpenter, Witherby, and Tauber 2020). In 1973 only 29% of colleges relied on the use of SETs; the numbers jumped to 68% in 1983, to 86% in 1993, and then to 94% in 2010 (Stroebe 2020). Its near ubiquitous use today in higher education, both nationally and internationally, is typically attributed to the ease of collecting, presenting, and analyzing data for the purposes of improving teaching quality, as well as its usefulness in appraisal processes (e.g., tenure, promotion, and award decisions), providing institutional assurances and accountability with respect to an institution's quality of education, and as a mechanism for creating a more "democratic" and "inclusive" educational environment for students (Spooren et al. 2017). In business terms, it is supposed that SETs make the business of education more "efficient" and "effective," as well as provide students a sense of "consumer satisfaction" (Katopes 2009). Yet from both a business perspective and the perspective of egalitarian values of diversity and inclusiveness, the continued reliance on SETs by administrators and pre-existing faculty, especially for the purposes of making hiring, reappointment, promotion, and award decisions is self-defeating since they are invalid measures of teaching effectiveness, their adoption is motivationally questionable, and they pose a legal threat to institutions.⁴

⁴ One might note that there has been an increase throughout the years in diversity among the faculty in higher education, along with an increase use in SETs in higher education, and this observation might be used to suggest that the use of SETs are, therefore, not barriers to diversity and inclusiveness in higher education. This argument, however, fails to attend to the fact that advancements toward diversity and inclusiveness in higher education may have been made despite the use of SETs. As I argue in section two, the use of SETs for hiring, reappointment, promotion, and award decisions equip administrators and pre-existing faculty with a more effective tool for gating since it provides a ready excuse to deny a marginalized candidate a well-earned job, promotion, or award while also providing committee members with a valid excuse for discounting SETs for any particular decision-making process. Furthermore, although gender there have been some improvements toward equity in higher education, there are still many problems (e.g., vertically stratified inequities among the higher ranks of academia, inequities in

(Please cite the published version)

Although scholars in the past touted the validity of SETs in measuring what their architects propose they measure, evidence against the validity of SETs in measuring teaching effectiveness has been available as early as the 1990s (Hornstein 2017). There is now even more evidence against their validity (e.g., Esarey and Valdes 2020; Fan et. al 2019; Rivera and Tilcsik 2019; Mitchell and Martin 2018; Arguete, Slater, Mwaikinda 2017; Hornstein 2017; Wagner, Reiger, and Voorvelt 2016; MacNell, Driscoll, and Hunt 2015; and Smith and Hawkins 2011). Given such evidence, the American Sociological Association (ASA) released a “Statement on Student Evaluations of Teaching” (2019), which was endorsed by 23 additional societies or associations. This statement cited evidence of 1) a weak relation between student evaluations and student learning, 2) methodological problems with studies reporting a positive correlation between student evaluations and teaching effectiveness, and 3) gender and racial bias as reasons for doubting the validity of student evaluations in measuring teaching effectiveness. It further observed that there was an emerging consensus among scholars that “using SETs as the primary measure of teaching effectiveness in faculty review processes can systematically disadvantage faculty from marginalized groups” (ASA 2019, 1).

Furthermore, there is now more evidence for questioning the initial motivation for the flawed research and arguments that have been previously published in support of the validity of SETs, especially during the 1980s when the use of SETs jumped to 68% and continued to significantly increase thereafter. According to a study by Spooren et al. (2017), the following three scholars are credited for authoring at least one of the top 10 articles, by total citations according to Google Scholar, in SET research: H. W. Marsh, W. J. McKeachie, and P. A. Cohen. These three authors, along with J. A. Centra, were also found to be among the top 11 authors in the field of SET research, with Marsh identified as the top author and the rest taking a place among the upper ranks of this list (Spooren et al. 2017). Given the significant influence these authors have had in the field, the fact that their research has been associated with at least 1-3 conflicts of interests (corporate, evaluative, administrative, or SET author), which were not originally disclosed to their readers (Uttl, Cnudde, and White 2019), challenges the integrity of the underlying motivation, arguments, and movement to adopt SETs on the basis of their research. In other words, the research that motivated the adoption of SETs as effective tools for measuring teaching effectiveness, especially during the 1980s, were not only methodologically flawed (Uttl, White, Gonzalez 2017), but also motivationally flawed.

Finally, student evaluations have failed to pass legal validation as effective measures of teaching effectiveness. In the recent arbitration case in Ontario, Canada, between Ryerson University and the Ryerson Faculty Association,⁵ arbitrator William Kaplan reviewed the evidence provided by both Ryerson University and the Ryerson Faculty Association, and determined that the expert evidence “establishes, with little ambiguity, that a key tool in teaching effectiveness [SETs] is flawed” (Kaplan 2018, 10). Not only this, but while Kaplan’s decision upheld the importance and right of Ryerson University to assess faculty teaching effectiveness, it also established that “the ubiquity of the SET tool is not a justification, in light of the evidence about its potential impact, for its continuation, or for mere tinkering” (Kaplan 2018, 10).

equal pay for equal work, and the continued persistence of gender norms and stereotypes; read Fan 2020) that need to be address, which the disuse of SETs may contribute to doing.

⁵ William Kaplan (Arbitrator), 2018, “Ryerson University vs. Ryerson Faculty Association,” CanLII 58446 (ON LA). Accessed August 30, 2020. <http://canlii.ca/t/hsqkz>.

(Please cite the published version)

Arbitrator Kaplan continued to note that “the evidence is dispositive that some of the [Ryerson University SET] questions do not elicit any useful information about teaching effectiveness and are subject to bias, while the use of averages—individual, Departmental, Faculty and University—provides no relevant information about teaching effectiveness” (Kaplan 2018, 10).

It should, therefore, be surprising that most universities across the country, and around the world, still rely on SETs for hiring, reappointment, promotion, and award decisions. According to Carpenter, Witherby, and Tauber (2020), some studies have shown that over 94% of deans and administrators have reported that they always use SETs as a basis for assessing faculty teaching, and more than 80% of administrators have reported using SETs for promotion and tenure decisions (Carpenter, Witherby, and Tauber 2020, 143). Even the ASA and Ryerson University are guilty of a lacking in their integrity. Rather than denouncing the use of SETs in such decisions, the ASA instead recommends 1) that SET questions should focus on student experiences, as an opportunity for students to give their feedback rather than an opportunity to formally assess their instructor’s teaching; 2) that SETs should be used as a part of a holistic approach; 3) that SETs can be used to appropriately document patterns in an instructor’s feedback; 4) that interpretive context should be provided with quantitative SET scores through the reporting of distributions, sample sizes, and response rates; and 5) that evaluators (e.g., chairs, deans, hiring committees, and tenure and promotion committees) should be trained to appropriately interpret and use SETs as a part of a holistic assessment of teaching effectiveness (ASA 2019, 2). Ryerson University has also failed to equally extend arbitrator Kaplan’s decision, which was limited to the use of Ryerson’s SETs (“Faculty Course Surveys) for the purpose of promotion and tenure decisions, to protect job candidates during their hiring processes from being harmed by similar problems with SETs conducted by other institutions. In short, the ASA, the 23 other associations or societies that have endorsed their statement, and Ryerson University remain complicit in the harms of the continued use of SETs in hiring, promotion, reappointment, or award decisions, despite explicitly acknowledging all the problems associated with SETs. This demonstrates a deficiency in their recognition of the actual and potential harms that the continued reliance on SETs authorize and enact, especially against marginalized members of a society. It does, however, put associations like the ASA ahead of associations like the American Philosophical Association, which has yet to endorse any kind of statement regarding the use or disuse of SETs, especially with respect to its aim towards diversity and inclusiveness in the discipline of philosophy.

From the perspective of a professional educator, what the ASA, the 23 additional associations or societies who also endorsed the ASA’s statement, and Ryerson University (including the Ryerson Faculty Association) failed to realize is that, in general, to continue to permit the use of SETs for hiring, reappointment, promotion, and award purposes, even under the five conditions outlined by the ASA, is as ethically problematic as permitting the use of SETs as a sole measure of teaching effectiveness. A moral wrong is a moral wrong regardless of whether or not it is committed within a wider practice that is thought to be more ethical, and embedding such wrongs in what is perceived to be a more ethical “holistic” practice makes such wrongs insidious. The most significant reason why is that these wrongs—the injustices against underrepresented and minority people that result from the continued use of SETs for hiring, reappointment, promotion, and award decisions—are consequences of the continued use, and therefore legitimization, of what is explicitly acknowledged to be an invalid measure of teaching effectiveness. In other words, you can paint a donkey pink and teach it to squeal, but it ain’t gonna ever give you bacon. The ASA’s primary justification for their recommendation is the

(Please cite the published version)

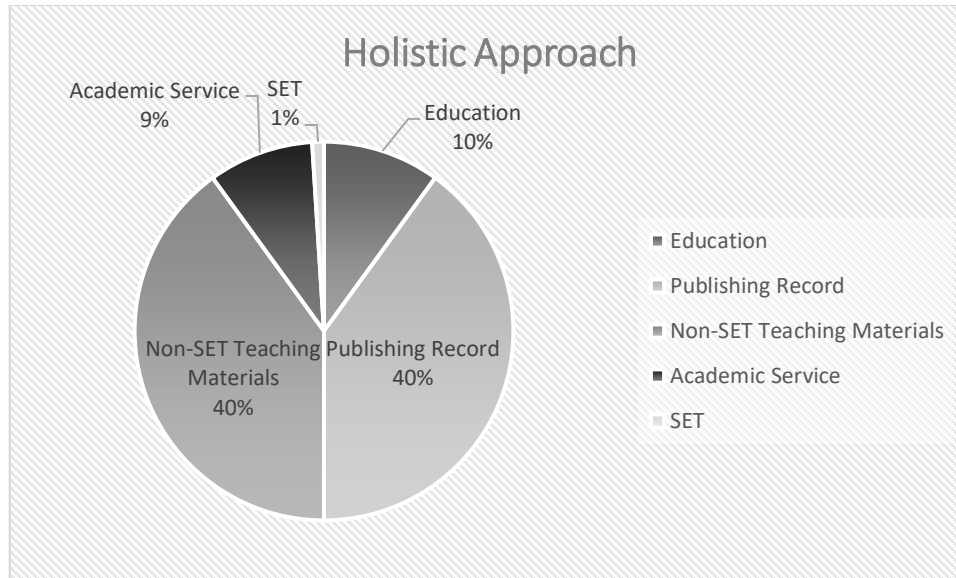
“wide use” of such holistic approaches (ASA 2019, 2). Yet as with arbitrator Kaplan’s reason for awarding a judgment on behalf of the Ryerson Faculty Association, the ubiquity of such holistic approaches is also not a justification for the continued reliance on such faulty measures for teaching effectiveness. And if this is the case for promotion and tenure decisions, then it is equally true for the purpose of hiring, reappointment, and award decisions. If necessarily flawed measures, such as SETs (Esarey and Valdes 2020), are allowed to be given any weight in a decision-making process, not only is the entire decision-making process dubious, but the explicit acknowledgment of the inadequacies of the defective measures is at best lip service and at worst a calculated sleight of hand. In either case, no real change is implemented in order to ensure that SETs do not continue to detrimentally impact those who are the most harmed. I now turn to enumerating these harms.

II. The Harms Against Marginalized Educators and Educators in General

Consider how, for example, how the ASA’s recommendation to take a “holistic approach” in making hiring decisions might actually play out, especially when one is considering a diversity of candidates. Even if a search committee considers multiple items (e.g., education, publishing record, non-SET teaching materials, and academic service), including SETs as an additional item with any weight in the decision-making process would still unfairly bias search committees against hiring an underrepresented or minority candidate. The decision to hire one candidate rather than another is often based on a very narrow margin, which can simply amount to a difference in SET scores, especially since many candidates can be equally matched in all the other factors. Given that SETs have been shown to be biased against women (e.g., Rivera and Tilcsik 2019; MacNell, Driscoll, Hunt 2015; Mitchell and Martin 2018; Wagner, Reiger, Voorvelt 2016; Holroyd and Saul 2016), as well as against women and non-English speaking instructors (e.g., Fan et. al 2019), and minority and especially Black faculty (e.g., Smith and Hawkins 2011; Aurguete, Slater, Mwaikinda 2017), to include these measures at all as a part of any decision-making process is to introduce a systematically disadvantaging mechanism into that process.

Although the consequences of including SETs as a factor within a holistic approach to decision-making may be at times negligible, especially when a candidate for a job, promotion, or award faces little to no competition or is a shoe-in (which should be rare), it is also possible for SETs to play perhaps the single most important decision-making factor in a hiring decision despite the attempt to give SETs very little weight, as a part of a holistic approach. To illustrate this second point, consider a case in which a job search committee decides to take the following holistic approach, which aims to minimize the weight of SETs to what they believe to be a trivial degree: education = 10%, publishing record = 40%, non-SET teaching materials 40%, academic service 9%, and SETs = 1%. One can visualize this holistic approach in the following way:

(Please cite the published version)



One might conclude that such an approach could not possibly disadvantage any underrepresented or minority candidate in the job market since SETs hold so little weight (1%), and is a part of a holistic approach. Yet in practice, such an approach will have a systematically disadvantaging consequence, especially when an underrepresented or minority candidate is competing against a non-marginalized candidate who is equally matched on all other factors except for SETs.⁶ When we implement the above holistic approach with two hypothetical candidates—a non-marginalized (Candidate 1) and a marginalized candidate (Candidate 2)—in which the only difference between the two candidates is that the marginalized candidate has a total SET score that is only 0.01% lower than the non-marginalized candidate’s total SET score, regardless of how the search committee calculates a “total” SET score for its candidates, the actual, practical effect of the weight of the SETs, even when theoretically held at 1% and with only a 0.01% difference between the two candidates, is stark: it becomes the deciding factor.

⁶ One may wonder how often it would be the case that a search committee would ever have to decide between two candidates with different marginalization statuses who are equally matched in every other way but their SETs. If one considers, however, the fact that professional academics must all have the appropriate qualification, publication, teaching experience, and experience garnered through academic service in order to fulfill many of the requirements for even submitting an application, and that competition is a constitutive aspect of life for professional academics, then there should almost always be a case in which at least two candidates are equally matched in every way except in SETs, especially with a candidate pool of over 300 applicants as it currently is in the discipline of philosophy. The main reason why is that all the other factors except for SETs are within the sphere, to a certain extent, of a professional academic’s control; these are the things that we have been trained to do and we remain in the job market because we continue to do them. In other words, within a pool of over 300 candidates, one should find at least two candidates (which is all one needs) of differing marginalization status who are equally competitive, able, and effective, simply as a matter of statistical probability. That one does not may be an indication that various biases are at work in the mind of the perceiver.

(Please cite the published version)

	Candidate 1	Candidate 2
Education (10%)	10	10
Publishing Record (40%)	40	40
Non-SET Teaching Materials (40%)	40	40
Academic Service (9%)	9	9
Student Evaluation (1%)	1	0.99
Total	100	99.99

SETs can, therefore, actually carry a significant amount of weight even when well-intended search committees believe that they have protected their hiring process from possible biases by reducing the weight of SETs to a negligible amount.⁷ In other words, every time a search committee is deciding between two equally matched candidates, in which one is a non-underrepresented, non-minority and the other is a marginalized candidate, then the marginalized candidate will systematically lose out given that the student biases reflected in SETs systematically disadvantage marginalized instructors. This is how biases, and therefore injustices, become systemic. It is in these kinds of “little things” or “minor details”—the significant effects of which go unnoticed by most—which introduce and sustain many of the systemic injustices in a society. The late Justice Ruth Bader Ginsberg once said that “Real change, enduring change, happens one step at a time”; but this is true for changes toward justice as well as injustice. Furthermore, although some research has shown that some biases can be mitigated (Peterson et. al 2019; Rivera and Tilcsik 2019), no study to date has shown that harmful student biases can be entirely eliminated.

SETs, therefore, do not merely serve, under the guise of quality control, as an illusory measure of teaching effectiveness, but more significantly they serve as a mechanism for gating the diversity and inclusiveness of academic departments, institutions, and academia as a whole by providing administrators and pre-existing faculty members with a ready-made justification—which administrators and pre-existing faculty members can practically use at will—for denying a marginalized job candidate or faculty a position, promotion, or award. Allowing SETs to have any weight in hiring, reappointment, promotion, and award decision-making processes is to give the harmful biases captured by SETs a default legitimacy, and since SETs systematically disadvantage underrepresented and minority faculty, no issue would be raised if a marginalized candidate or faculty is given what may be a well-deserved position, promotion, or award despite their low SET scores. Administrators or pre-existing faculty who are responsible for making hiring, reappointment, promotion, or award decisions can, therefore, simply make decisions in accordance with their preferences (much of which can be illegitimately based).

⁷ One might suggest that such a strict quantitative approach to holistic decision-making is unrealistic, that in practice strict quantitative approaches like this are very rarely used. The point, however, is to provide a quantitative illustration to highlight how much of an impact fine-grained differences can make. My assumption is that qualitative approaches ultimately underlie these quantitative judgments. Furthermore, such effects are not restricted to quantitative approaches. They can also result from taking a qualitative approach that uses no numerical value in the decision-making process. Purely qualitative approaches must still ultimately include the weighing of various factors, and all one needs are conditions in which one has two final candidates with only a difference between the weighting of their SETs for SETs to have this kind of effect.

(Please cite the published version)

SETs, therefore, serve the purpose of alleviating the need of administrators and pre-existing faculty to provide any kind of legitimate justification for their decision, except when deciding to discount SETs on behalf of a candidate, which is always easy to do given their established invalidity. This reliance on SETs for such gating purposes is especially salient when one considers Ryerson University's current policy, which bars the use of SETs for tenure or promotion decisions but does not do so for hiring decisions. Apparently, according to Ryerson University and the Ryerson Faculty Association, SETs are fine to use in order to discriminate against external job candidates, but when applied to internal candidates it has been judged to be irredeemably problematic. Yet one must ask, why the difference? Thus an institution's decision to include SETs in hiring, reappointment, promotion, and award decisions is a decision not only to maintain SETs as instruments for reinforcing irrelevant and harmful student biases, but also as a mechanism for reinforcing irrelevant and harmful administrative or pre-existing faculty biases, which have not gone undocumented (e.g., Turner, Myers, and Creswell 1999; Bernal and Villalpando 2010; O'Meara, Culpepper, and Templeton 2020). It is to continue the systematization of the direct discrimination committed by students against marginalized faculty, and thereby transforming them into wrongs of indirect discrimination with disparate impact.⁸ Once again, this is how such marginalizing biases become systemic.

Some have argued for the continued use of SETs for reasons other than hiring, reappointment, promotion, and award decisions, such as for providing "useful" information about how a teacher affected a student's learning experience or engagement with a subject matter, and whether a student received adequate feedback in a timely manner.⁹ The ASA's statement suggests as much in its recommendations. Yet such reasons fail to appropriately acknowledge the inherently discriminatory nature of SETs, and the harms they commit by exacerbating already hostile conditions in which marginalized faculty work (e.g., read Misra and Lundquist 2015; Alexander and Moore 2008; Jackson and Crawley 2003).

First, given the evidence that SETs reflect a variety of harmful student biases, including gender and racial/ethnic biases, SETs are literally records of students' discrimination against underrepresented and minority faculty. To judge an underrepresented faculty as being less qualified due to one's implicit bias against women is an act of what Rasmussen (2020) refers to as "non-intentional disparate treatment discrimination." In most cases, acts of discrimination are not carried out as acts that are justified by explicit discriminatory beliefs. A student with a gender bias against women and a racial bias against black people does not usually think to themselves, "This teacher is incompetent and her authority is illegitimate because she is a black woman." They simply think, "This teacher is incompetent," and refuse to follow her instructions. They also do so, unbeknownst to them, because the teacher is a black woman. Students may be ignorant of perpetrating their unintentional acts of discrimination while completing their SETs,

⁸ Read Rasmussen 2020 regarding forms of direct and indirect discrimination.

⁹ Some might also suggest that SETs help marginalized students have a voice against possibly biased instructors. The problem with this rationale is that such voices are too easily disregarded given all the problems with SETs. In other words, an administrator or department chair can simply choose not to give such an SET any validity, since the general invalidity of SETs is now well established. So the best forum for having such voices heard is not SETs, but instead forums such as those established by Title IX policies. For concerns regarding racial/ethnic discrimination, I would recommend that students demand the establishment of analogous reporting policies and forums at their institution.

(Please cite the published version)

but this does not mean they are not guilty of committing any acts of discrimination. Students are also not always naive actors. According to Carpenter, Witherby, and Tauber (2020), in some studies students have reported that SET ratings are based on the grades that students have been receiving throughout a course, and that students intentionally provide inaccurate information in order to “get back” at their instructors for the grades they have been receiving; students (36.5%) also reported knowing other students who have intentionally submitted false information about an instructor because they did not like the instructor. Second, given that SETs are records of student biases, the kind of criticism collected through SETs are not equally distributed among the faculty even when the circumstances on which these criticisms are based are. For example, evidence suggests that students rate female instructors as taking significantly more time to return feedback compared to male professors even when both female and male instructors returned feedback at the same time (MacNell et al. 2015). That this occurs should not be a surprise since such unequal distributions are a feature of discriminatory acts. This is an aspect of what makes them *discriminatory*. One should also take care to note that it is these kinds of biases—gender and racial/ethnic biases—that are also at the root of disparate effects due to the COVID-19 crisis and the Trump administration’s failure to appropriately respond, the injustices against which Black Lives Matter activist are fighting, and the wrongs committed by what are now known as “Internet Karen’s,” in reference to the now infamous woman who made a false police report against a black man who was birdwatching in a park simply because he asked her to leash her dog, in accordance with posted park regulations.

Given the foregoing, the reliance on SETs for any reason also places instructors under conditions that hold them responsible for factors that are beyond their control: students’ implicit biases. This places undue burdens, especially on marginalized educators, and can incentivize many instructors, perhaps out of the sheer fear of professional survival, to implement alternative strategies for raising their SETs, such as grade inflation and the use of passive learning strategies. The continued legitimization of the use of SETs for hiring, reappointment, promotion, and award decisions, including as a method for appropriately “documenting patterns in instructor feedback,” can therefore work to motivate marginalized instructors, as well as non-marginalized instructors, to—perhaps under a coercive threat of “documenting patterns in instructor feedback”—lower the quality of their teaching. It can also punish both marginalized and non-marginalized instructors who have fought to resist the implementation of less effective teaching strategies due to student preferences for such strategies. I discuss these effects in more detail in the subsequent section on the harms of SETs against non-privileged students and a democratic society, in general.

The use of SETs in hiring, reappointment, promotion, and award decisions also commit harms against administrators and pre-existing faculty, as members of job search committees, tenure review committees, and award committees by both ethically and practically overburdening them with the responsibility to consider *necessarily invalid* data as a legitimate source of information in order for them to make a *just* decision about a job candidate or faculty member.¹⁰ Not only is this impossible, but it also requires them to commit ethically questionable,

¹⁰ I grant that letters of recommendations can also be as or even more biased than SETs, but that this is the case does not discount the bias in SETs nor does it discount any argument to exclude SETs in hiring, reappointment, promotion, and award decisions. The more reasonable inference to draw would be that one ought to also discount letters of recommendation for hiring,

(Please cite the published version)

and perhaps even illegal acts of discrimination in order to make even an attempt at doing so. To base a decision about a job, a promotion, or an award on factors that have been proven to systematically introduce a variety of harmful biases, including gender and racial/ethnic biases, is to amplify these discriminations and thereby compound the harms of these biases. As mentioned earlier, this is how biases become systematized, and is a form of indirect discrimination with disparate impact. Furthermore, it is unreasonable to presume that any committee would be able to make an appropriate judgment about a candidate for a job, promotion, or award with the addition of necessarily invalid SET and harmfully biased measures if they cannot already make such a decision based on all the other less controversial information that they are already required to review, such as non-SET materials in a teaching portfolio, publication quality and record, grant procurement record, history of academic service, and recommendation letters. As noted earlier, what search committees might regard to be “helpful” information in SETs might in fact be completely misleading. Furthermore, up-to-date institutional teaching observations conducted by experts that are external to a department, review of syllabi and other teaching materials that were actually used in an instructor’s course, a review of an instructor’s feedback on students’ assignments, a review of student reflections on what they learned in a course, and a review of any formal complaints submitted by students through appropriate forums should provide enough legitimate and adequately unbiased information for making a more fair and accurate assessment of an instructor’s quality of teaching effectiveness. As such, not only is the reliance on SETs for hiring, reappointment, promotion, and award decisions both ethically and practically burdensome for job search committees, tenure review committees, and award committees, but they are unnecessarily so. They, therefore, also commit harms against the administrators and pre-existing faculty who serve on these committees. SETs should, therefore, never be a factor in any hiring, reappointment, promotion, or award decision-making process.

III. The Harms Against Non-Privileged Students and Society

Rather than helping to ensure a student’s quality of education, the use of SETs not only harm students in general, but they also non-privileged students (students of both the general and underserved populations), accomplices to their own oppression. Although in a society like the U.S., in which gender and racial injustices are systemic, it is not always the case that hard work brings just rewards, it is still the case that those who succeed have worked hard to get to where they are. The lesson here is that hard work is, in general, a necessary condition for success even if it is not a sufficient condition. This is also true in education. It is only when students are asked to work through and overcome challenges that one can say that they have truly learned something. Learning is a kind of growth, and simply demonstrating that one is able to do what one has been able to do is not learning. It is also true that such learning not only prepares students with the skills to fulfill their future workplace responsibilities, but it also helps students build the endurance and resilience they will need in order to meet workplace demands and life challenges, which they will eventually have to face after they graduate if they want to achieve

reappointment, promotion, and award decisions. But to make such an argument is beyond the scope of this paper.

(Please cite the published version)

even just the possibility of what many believe to be indicative of a successful, flourishing life: a professional career, in whatever form that it might take, in some industry.¹¹

Providing students with a high quality of education, which is the kind of education that achieves the above aims in teaching, is not only beneficial to students but also to a society as a whole. Armed with such knowledge and habits of success, students can enter the workforce with the confidence to achieve bigger and better things for themselves, and also the skills and abilities to do so. Such achievements include access to better health, employment, housing, and recognition before the law. In a litigious, capitalistic, “buyer beware” society like the U.S., it is often the case that one must demand what one is owed through being appropriately informed and knowing how to argue for one’s claims. This includes access to good, affordable healthcare and housing, rewarding employment, and equal recognition before the law. It is also through having a well-informed voting public that a democracy like the U.S. can ensure its flourishing—with appropriate checks and balances—but this cannot happen when the public is not appropriately educated. For example, being able to read and think at the level of higher education is crucial for ensuring that one practices their right to vote in non-self-defeating ways. To deny students the opportunity to develop these skills and habits, therefore, denies such students the opportunity to compete both in the job market and in life. It is to fail in providing students with the kind of adequate education they were promised as good faith consumers of education and as members of a democratic society that touts the value of a higher education.

Given the foregoing, one should be able to agree that to simply give students the impression that they have developed a certain level of knowledge and skills without them actually doing so, and to regard this as an acceptable outcome not only fails to furnish students with their just deserts as good faith consumers and members of our society, but it is to swindle them. Institutions who do so are committing acts of fraud against their students, and to use student feedback as a mechanism to achieve these ends is to make students accomplices to the harms committed against them. Yet this is exactly what the use of SETs in hiring, reappointment, promotion, and award decisions do. The causal chain is indirect, but easy to follow, especially under the presumption of a business model of education, which operates with the primary purpose of generating a profit and has now become the predominant model for U.S. higher education (Katopes 2009). A growing number of studies are finding that the correlation between higher SET scores and student grades may be more so an indicator of grade inflation than teaching effectiveness, and arguments are mounting for the conclusion that the reliance on SETs for hiring, reappointment, promotion, and award decisions are leading to more ineffective rather than effective teaching (e.g., Stroebe 2020; Carpenter, Witherby, and Tauber 2020; Deslauriers et al. 2019; Lee et. al 2018; Kornell and Hausman 2016; Yunker and Yunker 2003).

For example, Stroebe (2020) notes that GPAs in the U.S. have been increasing for decades; yet rather than university students demonstrating that they have become more hardworking and better qualified for college, students report spending less time on academic

¹¹ As to whether or not such a life is one that ought to be valued by all is another question. My assumption throughout this paper, however, is that at least one reason why someone would choose to pursue an undergraduate education is because they hope that their education will contribute to their eventual success in the job market. Yet even if this does not apply, as long as a student hopes to learn, as I have defined it, through their education, then my arguments about the harms of using SETs for hiring, reappointment, promotion, and award decisions are still applicable.

(Please cite the published version)

pursuits, SAT scores show a downward trend, and students have developed less critical thinking, complex reasoning, and writing skills compared to students from a number of decades ago. What explains this inconsistency, according to Stroebe, between student GPAs and other indicators of high quality learning is grade inflation, which can be traced back to the motivation that teachers have to achieve higher SET scores due to their use in hiring, reappointment, promotion, and award decisions. Stroebe cites an abundance of empirical evidence suggesting that students reward lenient grading, as well as the fact that instructors resort to lenient grading in order to achieve higher SET scores. This fact should also not be a surprise for anyone with a good amount of actual teaching experience: learning is difficult and students demonstrate a consistent preference for passive teaching methods that leave them with the impression of learning compared to more effective active learning methods (Kornell and Hausman 2016; Deslauriers et al. 2019; Carpenter et al. 2020).

To illustrate, consider the findings of Deslauriers et al. (2019). They were concerned with the continued widespread use of passive learning practices in STEM programs (e.g., listening to lectures) despite the extensive research on student learning which has established that active learning practices (e.g., deliberative practice of concepts in problem solving) are more effective teaching strategies. The fear of lower SETs was cited as one of the reasons why instructors chose more passive teaching strategies compared to more active teaching strategies. As Deslauriers et al. noted:

Indeed, one-third of instructors who try active teaching eventually revert to passive lectures, many citing student complaints as the reason (23). Instructors report that students dislike being forced to interact with one another (15, 17, 24), they resent the increase in responsibility for their own learning (21, 22), and they complain that “the blind can’t lead the blind” (19). (Deslauriers et al. 2019, 19251)

To test whether or not students have a bias toward passive learning strategies, Deslauriers et al. (2019) conducted a study on students in physics courses at a major U.S. university. They randomly assigned students into two groups, and they toggled the experimental condition of using active learning strategies between the two groups. In other words, when one group was using active learning strategies during a class period the other group was using a passive learning strategy. They also gave both groups a survey at the end of each class in order to record student perceptions about their class as well as a multiple choice test to record actual learning measurements. They also conducted a follow-up one-on-one with student participants to learn more about their perceptions of learning. According to their results, they found that although students significantly demonstrated more actual learning using active learning strategies, they reported that the class in which they practiced the passive learning strategy was more enjoyable and that they felt that they learned more in the passive learning strategy class. They also judged the instructor when using a passive learning strategy to be more effective and reported that they wished all their physics courses were taught using passive learning strategies. As Deslauriers et al. summarized, “In this report, we identify an inherent student bias against active learning that can limit its effectiveness and may hinder the wide adoption of these methods” (Deslauriers et al. 2019, 19251), which is problematic since active learning strategies have been proven to be more effective learning strategies than passive learning strategies. Yet the main reason why student biases would hinder the adoption of these methods is because of the reliance on SETs for hiring, reappointment, promotion, and award decisions, and not simply because students have an

(Please cite the published version)

inherent bias against these kinds of methods. If you give kids a choice between a bowl of fruit and a bowl of candy, they will most likely choose the bowl of candy. People have an inherent preference for sugar. But maybe the problem is the fact that they are given such a choice. Are not educators, like parents, supposed to know better than students? Why are students being given the opportunity to choose between these two kinds of teaching methods? Students can also be taught to appreciate active learning methods. In other words, it is only in contrast with passive learning strategies and the focus on student preferences through SETs that such inherent student biases problematically hinder the adoption of more effective learning strategies. Although teachers should never implement abusive teaching strategies or any other kind of ineffective strategies, there is a difference between such strategies and proven, effective, learning strategies, which students prefer less compared to passive learning strategies that have been proven to be less effective.

What is worse is that students are not only being given a worse education compared to the quality of education that was given to students a few decades ago, but the reliance on SETs as feedback for educators to “improve” on their teaching effectiveness has led teachers to focus on employing strategies that give students the impression of learning. Carpenter, Witherby, and Tauber (2020) argue that qualities such as better organization, fluency of style, and enthusiasm, which are common factors that are measured in SETs, leave students with the impression of learning, but do not have matching effects on a student’s learning. Furthermore, Carpenter, Witherby, and Tauber also argue that because students have inaccurate perceptions of what effective learning feels like, they misjudge effective teaching strategies as being ineffective and report that they learn more effectively from passive approaches rather than those concrete approaches—such as retrieval practice, distributed practice, and active learning—that have been proven to be effective. Thus student responses in SETs encourage instructors to improve on those factors that leave students with the belief that they learned while jettisoning those strategies that lead to improved learning but lower SET scores. In other words, students—in virtue of their naive SET responses about their learning experiences—bring about the implementation of bargain-basement learning conditions for themselves and their fellow students, while also feeling like they have made a significant contribution to increasing the quality of their education, institution, and gaining a sense of “consumer satisfaction.” Furthermore, students may also intentionally do so by knowingly making false reports in their SETs because they resented the grades that their teacher gave them, or simply because they disliked their teacher. Again, as Carpenter, Witherby, and Tauber (2020) noted, some students reported intentionally submitting inaccurate information in order to “get back” at their instructors for the grades they received, and some students (36.5%) also reported knowing other students who had intentionally submitted false reports because they did not like the instructor. Such are the consequences of an entitled consumerism, which is fostered by a business model of education that relies on the use of SETs for hiring, reappointment, promotion, and award decisions and which can turn some students into accomplices to their own oppression.

When such consequences also necessarily have differential effects on marginalized students in a community, the implementation of the mechanism that brings about these consequences is the implementation of a structural injustice. The result is an undereducated general population—which are the population of non-privileged people—with an inflated impression of having gained a certain level of knowledge and skills, which they will be unable to actually demonstrate in the workplace and in everyday life. They will move through the highly competitive job market without understanding why they are failing or to eventually realize that

(Please cite the published version)

they were sold a bill of goods, especially after finding themselves under a mountain of financial debt from the educational loans they took out as an investment in their future. This is what the business model of education, which reinforces the use of SETs in hiring, reappointment, promotion, and award decisions do. Yet not every student will be subject to such effects because unlike students from the general populations—which constitute the majority of U.S. students in higher education—students from privileged backgrounds have the resources (e.g., highly educated parents; tutors, academic advisors, and academic consultants external to their institutions; as well as nepotistic professional networks) to help them realize that the training they are receiving is inadequate, and to make-up for these deficiencies in one way or another. The use of SETs for hiring, reappointment, promotion, and award decisions, therefore, leave non-privileged students at a considerable disadvantage. It unwittingly encourages non-privileged students—who typically have little to no recourse for the poorer quality of education that SETs incentivize—to structure their educational environment through their unconscious, biased SET responses in such a way that leave them with educators who lower their quality of teaching to match students' preferences while giving them the impression of learning, so that such educators can continue to survive within a systematically oppressive system that has been reinforced by the use of SETs.

From the business perspective of education, one might argue that it is a student's responsibility to inform themselves about what constitutes effective teaching. As the saying goes in business: *Buyer beware!* But the inapplicability of this rule to educational goods ought to be a clear indication that a business model, along with its reliance on SETs for hiring, reappointment, promotion, and award decisions, is an inappropriate model for education. One cannot hold students responsible for being knowledgeable consumers of a product that has the purpose of helping them develop the skills and habits that would allow them to be such consumers. To do so betrays not only the unreasonableness of institutions who continue to rely on SETs despite the evidence that clearly establishes their invalidity and the harms they confer on unsuspecting students, as well as job candidates and pre-existing faculty, but it also betrays that institution's lack of ethical concern for the students they serve and their employees. The reliance on SETs for hiring, reappointment, promotion, and award decisions also enacts procedures that hinder the implementation of at least one corrective to the kind of student biases that make SETs especially problematic. Fan et. al (2019) conducted a longitudinal study at a leading Australian university of 523,703 individual student surveys, across five different faculties with diverse cultural backgrounds (38% of the university faculty had non-English speaking backgrounds), over a period of 7 years (2010-2016). According to their findings, not only is there a significant interaction between gender and culture, such that students seem to unconsciously prefer instructors who are more like them, but they also found evidence for the conclusion that students' gender and cultural biases may be significantly reduced when they are exposed to conditions of diversity, such as having a more diverse faculty. Accordingly, the reliance on SETs for hiring, reappointment, promotion, and award decisions also harm students and society as a whole by acting as a mechanism to deny students in the general population (i.e., non-privileged students) the conditions that would allow them to overcome their biases in order to be able to make judgements that are the most beneficial for their well-being. SETs should, therefore, never be a factor in any hiring, reappointment, promotion, or award decision-making process.

IV. Conclusion

(Please cite the published version)

SETs have been proven to be invalid measures of teaching effectiveness, the motivation for their use is questionable, and they have failed legal validation. The use of SETs in hiring, reappointment, promotion, and award decisions, therefore, enact numerous harms against professional and students educators in higher education, and a society as a whole. As a part of a holistic approach in a hiring process, they can keep administrators and pre-existing faculty from having those difficult discussions and arguments about why a candidate should or should not receive a job, promotion, or award—discussions and arguments that can reveal, and sometimes dismantle or re-establish, the same kind of harmful biases that are documented in SETs. But these are the kinds of conversations that any committee ought to have in order to ensure that the best possible candidate (rather than what some might believe would be the best possible candidate) for their particular department is hired, despite the kinds of explicit or implicit biases that various administrators or pre-existing faculty may hold. These are the kinds of conversations that need to be had in order to make academia more diverse and inclusive for current and future generations of students. The use of SETs in hiring, reappointment, promotion, and award decision-making processes also legitimizes the direct discrimination committed by students against marginalized educators, and thereby work to sustain and reinforce a hostile environment for marginalized educators. Their use systematize the discriminatory acts in SETs into acts of indirect discrimination with disparate impact. They are mechanisms for systemically sustaining and reinforcing gender, racial/ethnic, and other harmful biases. Their use place administrators and pre-existing faculty in ethically impossible and practically overburdensome situations, while also allowing such administrators and pre-existing faculty to offload the moral weight of making such decisions onto student preferences (i.e., student biases). Finally, they have led to the implementation of low-quality teaching methods as a response to fulfilling such preferences, and leave many students with having the impression that they have a good quality education rather than actually having a good quality education. This not only harms students, and especially non-privileged students, but also the aims of a democratic society as a whole. A democracy requires an educated general populace so that they can appropriately exercise their rights to truly protect their life, liberty, and pursuit of happiness. A populace with delusional impressions about their level of knowledge and skills cannot do so. SETs should, therefore, never be a factor in any hiring, reappointment, promotion, or award decision-making process.

(Please cite the published version)

References

- Adams, Meredith J. D. and Paul D. Umbach. 2012. "Nonresponse and Online Student Evaluations of Teaching: Understanding the Influence of Salience, Fatigue, and Academic Environments." *Research in Higher Education* 53, no. 5: 576-591. <http://www.jstor.org/stable/23257596>.
- Alexander, Rudolph, Jr., and Sharon E. Moore. 2008. "The Benefits, Challenges, and Strategies of African American Faculty Teaching at Predominantly White Institutions." *Journal of African American Studies* 12: 4-18. <https://doi.org/10.1007/s12111-007-9028-z>.
- American Sociological Association (ASA). 2019. "Statement on Student Evaluations of Teaching." https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_feb132020.pdf.
- Aurquete, Mara S., Joshua Slater, and Sekela R. Mwaikinda. 2017. "The Effects of Professors' Race and Clothing Style on Student Evaluations." Special Issue: Ethnic Differences, Experiences, and Connections for All Students in Education. *The Journal of Negro Education* 86, no. 4: 494-502.
- Badal, S., and Harter, J. K. 2014. "Gender Diversity, Business-Unit Engagement, and Performance." *Journal of Leadership & Organizational Studies*, 21(4), 354-365. <https://doi.org/10.1177/1548051813504460>
- Benfer, Emily, David Bloom Robinson, Stacy Butler, Lavar Edmonds, Sam Gilman, Katherine Lucas Mckay, Zach Neumann, Lisa Owens, Neil Steinkamp, And Diane Yentel. 2020. "The Covid-19 Eviction Crisis: An Estimated 30-40 Million People In America Are At Risk." The Aspen Institute. Accessed August 23, 2020. <https://www.aspeninstitute.org/blog-posts/the-covid-19-eviction-crisis-an-estimated-30-40-million-people-in-america-are-at-risk/>.
- Bernal, Dolores Delgado and Octavio Villalpando. 2010. "An Apartheid of Knowledge in Academia: The Struggle Over the 'Legitimate' Knowledge of Faculty of Color." *Equity and Excellence in Education* 35, no. 2: 169-180. <https://doi.org/10.1080/713845282>.
- Carpenter, Shana K., Amber E. Witherby, and Sarah K. Tauber. 2020. "On Students' (Mis)judgments of Learning and Teaching Effectiveness." *Journal of Applied Research in Memory and Cognition* 9: 137-151. <https://doi.org/10.1016/j.jarmac.2019.12.009>.
- Chappell, Bill. 2020. "'Enormous And Tragic': U.S. Has Lost More Than 200,000 People To COVID-19." NPR. Online. Accessed October 6, 2020. <https://www.npr.org/sections/coronavirus-live-updates/2020/09/22/911934489/enormous-and-tragic-u-s-has-lost-more-than-200-000-people-to-covid-19>.
- Deslauriersa, Louis, Logan S. McCartya, Kelly Millerc, Kristina Callaghana, and Greg Kestina. 2019. Measuring Actual Learning versus Feeling of Learning in Response to Being Actively Engaged in the Classroom. *Proceedings of the National Academy of Sciences* 116, no. 39: 19251-19257. <https://doi.org/10.1073/pnas.1821936116>.
- Esarey, Justin and Natalie Valdes. 2020. "Unbiased, Reliable, and Valid Student Evaluations Can Still Be Unfair." *Assessment and Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2020.1724875>.
- Fan, C. Cindy. 2020, "Gender Inequity in American Higher Education." In *Scholarly Engagement and Decolonisation: Views from South Africa, the Netherlands, and the United States*, edited by Maurice Crul, Liezl Dick, Halleh Ghorashi, and Abel Valenzuela

(Please cite the published version)

- Jr., 55-80. *On Higher Education Transformation*, vol. 1. Stellenbosch, S. Africa: African Sun Media/Sun Media.
- Fan, Y., L. J. Shepard, E. Slavich, D. Waters, M. Stone, R. Abel, E. L. Johnston. 2019. "Gender and Cultural Bias in Student Evaluations: Why Representation Matters." *PLoS ONE* 14, no. 2: e0209749. <https://doi.org/10.1371/journal.pone.0209749>.
- Ford, Tiffany, Sarah Reber, and Richard V. Reeves. 2020. "Race Gaps in COVID-19 Deaths are Even Bigger than They Appear." *Middle Class Memos*. Brookings Institute. Accessed August 25, 2020. <https://www.brookings.edu/blog/up-front/2020/06/16/race-gaps-in-covid-19-deaths-are-even-bigger-than-they-appear/>.
- Gutiérrez-Fernández, M. and Y. Fernández-Torres. 2020. "Does Gender Diversity Influence Business Efficiency? An Analysis from the Social Perspective of CSR." *Sustainability* 12, no. 9: 3865. <https://doi.org/10.3390/su12093865>.
- Holroyd, Jules and Jennifer Saul. 2016. "Will the Teaching Excellence Framework Be Sexist." *The Guardian*. Accessed August 31, 2020. <https://www.theguardian.com/higher-education-network/2016/apr/04/will-the-teaching-excellence-framework-be-sexist>.
- Hornstein, Henry A. 2017. "Student Evaluations of Teaching are an Inadequate Assessment Tool for Evaluating Faculty Performance." *Cogent Education* 4: 1304016.
- Jackson, Ronald L., II, and Rex L. Crawley. 2003. "White Student Confessions about a Black Male Professor: A Cultural Contracts Theory Approach to Intimate Conversations about Race and Worldview." *The Journal of Men's Studies* 12, no. 1: 25-41.
- Katopes, Peter. 2009. "The Business Model is the Wrong Model." *Inside Higher Education*. Accessed October 5, 2020. <https://www.insidehighered.com/views/2009/02/16/business-model-wrong-model>.
- Kochhar, Rakesh. 2020. "Unemployment rose higher in three months of COVID-19 than it did in two years of the Great Recession." *Fact Tank*. Pew Research Center. Accessed August 27, 2020. <https://www.pewresearch.org/fact-tank/2020/06/11/unemployment-rose-higher-in-three-months-of-covid-19-than-it-did-in-two-years-of-the-great-recession/>.
- Kornell, Nate, and Hannah Hausman. 2016. "Do the Best Teachers Get the Best Ratings?" *Frontiers in Psychology* 7: 570. <https://doi.org/10.3389/fpsyg.2016.00570>.
- Lee, LaDoris J., Connolly, Manju E., Dancy, Melissa H., Henderson, Charles R., Christensen, Warren M. 2018. "A Comparison of Student Evaluations of Instruction vs. Students' Conceptual Learning Gains." *American Journal of Physics* 86, no. 7: 531-535. <https://doi.org/10.1119/1.5039330>.
- MacNell, Lillian, Adam Driscoll, and Andrea N. Hunt. 2015. "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education* 40: 291-303.
- Misra, Joya, and Jennifer Lundquist. 2015. "Diversity and the Ivory Ceiling." *Inside Higher Ed*. Accessed October 8, 2020. <https://www.insidehighered.com/advice/2015/06/26/essay-diversity-issues-and-midcareer-faculty-members>.
- Mitchell, Kristina M. W. and Jonathan Martin. 2018. "Gender Bias in Student Evaluations." *PS: Political Science & Politics* 53, no. 2: 270-274. doi:10.1017/S1049096519001744.
- O'Meara, KerryAnn, Dawn Culpepper, Lindsey L. Templeton. 2020. "Nudging Toward Diversity: Applying Behavioral Design to Faculty Hiring." *Review of Educational Research* 90, no. 3, 311-348: <https://doi.org/10.3102/0034654320914742>.
- Peterson, David A., Lori A. Biederman, David Anderson, Tessa M. Ditonto, and Kevin Roe. 2019. "Mitigating Gender Bias in Student Evaluations of Teaching." *PLoS ONE* 14, no. 5: e0216241. <https://doi.org/10.1371/journal.pone.0216241>.

(Please cite the published version)

- Rasmussen, Katharina B. 2020. "Implicit Bias and Discrimination." *Theoria*.
<https://doi.org/10.1111/theo.12227>.
- Rivera, Lauren A. and András Tilcsik. 2019. "Scaling Down Inequality: Ratings Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84, no. 2: 248-274. <https://doi.org/10.1177/0003122419833601>.
- Settles, Isis H., Martinque K. Jones, NiCole T. Buchanan, and Kristie Dotson. 2020. "Epistemic Exclusion: Scholar(ly) Devaluation that Marginalizes Faculty of Color." *Journal of Diversity in Higher Education*. Advance online publication.
<https://doi.org/10.1037/dhe0000174>
- Smith, Betty P. and Billy Hawkins. 2011. "Examining Student Evaluations of Black College Faculty: Does Race Matter?" *The Journal of Negro Education* 80, no 2: 149-162.
- Spooren, Pieter, Frederic Vandermoeren, Raf Vanderstraeten, and Koen Pepermans. 2017. "Exploring High Impact Scholarship in Research on Student's Evaluation of Teaching (SET)." *Educational Research Review* 22: 129-141.
- Stroebe, Wolfgang. 2020. "Student Evaluation of Teaching Encourages Poor Teaching and Contributes the Grade Inflation: A Theoretical and Empirical Analysis." *Basic and Applied Social Psychology* 42, no. 4: 276-294.
<https://doi.org/10.1080/01973533.2020.1756817>.
- _____. 2016. "Why Good Teaching Evaluations May Reward Bad Teaching: On Grade Inflation and Other Unintended Consequences of Student Evaluations." *Perspectives on Psychological Science* 11, no. 6: 800-816.
- Turner, C. S. V., S. Myers, and J. Creswell. 1999. "Exploring Underrepresentation: The Case of Faculty of Color in the Midwest." *The Journal of Higher Education* 70, no. 1: 27-59.
doi:10.2307/2649117. <http://www.jstor.org/stable/2649117>.
- Uttl, Bob, Kelsey Cnudde, and Carmela A. White. 2019. "Conflict of Interest Explains the Size of Student Evaluation of Teaching and Learning Correlations in Multisection Studies: A Meta-Analysis." *PeerJ* 7: e7225. <http://doi.org/10.7717/peerj.7225>.
- Uttl, Bob, Carmela A. White, Daniela Wong Gonzalez. 2017. "Meta-Analysis of Faculty's Teaching Effectiveness: Student Evaluation of Teaching Ratings and Student Learning are Not Related." *Studies in Educational Evaluation* 54: 22-42.
<https://doi.org/10.1016/j.stueduc.2016.08.007>.
- Wagner, Natascha, Matthias Rieger, Katherine Voorvelt. 2016. "Gender, Ethnicity and Teaching Evaluations: Evidence from Mixed Teaching Teams." *Economics of Education Review* 54: 79-94. <https://doi.org/10.1016/j.econedurev.2016.06.004>.
- Wallisch, Pascal and Julie Cachia. 2018. "Are Student Evaluations Really Biased by Gender? Nope, They're Biased by 'Hotness.'"
- Weinberg, Bruce A., Masanori Hashimoto, and Belton M. Fleisher. 2009. "Evaluating Teaching in Higher Education." *The Journal of Economic Education* 40, no. 3: 227-261.
<http://dx.doi.org/10.3200/JECE.40.3.227-261>.
- Yunker, Penelope J. and James A. Yunker. 2003. "Are Student Evaluations of Teaching Valid? Evidence from an Analytical Business Core Course." *Journal of Education for Business* 78, no. 6: 313-317. <https://doi.org/10.1080/08832320309598619>.