

A report prepared for the Canadian Sociological Association (CSA)

(Re)Prioritizing Pedagogic Feedback: Faculty Experiences with Qualitative Comments from Student Evaluations of Teaching (SETs)

Rachel La Touche, Lisa Kowalchuk, Rochelle Wijesingha

JUNE 2020

CSA Canadian Sociological Association
Société canadienne de sociologie



Table of Contents

EXECUTIVE SUMMARY	5
Background and Justification	5
Methodology	5
Hypotheses and Key Quantitative Findings of the Survey	6
Key Qualitative Findings of the Survey	8
Recommendations	9
INTRODUCTION	10
LITERATURE REVIEW	12
Contextualizing SETs Historically	12
Research on Biases in SETs Ratings	12
The Growing Concerns with SETs Comments	15
SETs Comments and Emotional Well-being	17
SETs' Impact on Teaching and Learning	18
SETs and Contingent Faculty	19
Institutional Responses	20
HYPOTHESES	21
METHODOLOGY	22
Study Participants	22
Data Collection, Design and Procedure	22
Limitations of the Study	23
Data Analysis	24
QUANTITATIVE DATA ANALYSIS	25
Experience of the Administration and Use of SETs	25
Impact of SETs Comments on Sense of Well-being	25
Criteria that Students Apply to SETs Comments	30
Impact of SETs Comments on Pedagogy	35
Intersectional Effects of SETs Comments	41
Contingent Faculty and SETs Comments	42

QUALITATIVE DATA ANALYSIS	43
Pedagogy	43
Personality, Personhood and Perceived Identity	44
Personality	47
Personhood	47
Perceived Identity: Gender	51
Perceived Identity: Race and Ethnicity	53
Perceived Identity: Age	55
Perceived Identity: Language/Accent/Voice	55
Perceived Identity: Sexuality	56
DISCUSSION	57
RECOMMENDATIONS	61
REFERENCES	66
APPENDIX A: Survey Instrument	72
APPENDIX B: Sample Characteristics	81

Acknowledgments

The authors thank the Canadian Sociological Association (CSA) for supporting this project by providing the lion's share of funding for it, and for the CSA Executive's patience as we carefully curated our results. We also wish to acknowledge funding generously provided by the Ontario Confederation of University Faculty Associations (OCUFA).

We are extremely grateful to Sherry Fox, the CSA's Executive Director, for her invaluable logistical advice and moral support provided from the very inception of the project until its completion. We also thank Carson Balestreri for his invaluable research assistance in collecting and annotating literature for us during the summer of 2018, work which in turn was funded by the University of Guelph's Undergraduate Research Assistantship program. Special thanks go to Dr. Valerie Damasco for lending her expertise to preliminary coding for the qualitative analysis.

We also owe a debt of gratitude to several of our colleagues, all Canadian sociologists, whose invaluable suggestions helped shape and motivate the project in its critical early stages, especially: Fiona Nelson, Bruce Ravelli, Kenneth Cain, Rochelle Coté, and Neil McLaughlin. Finally, we thank all three members of the CSA's sub-committee on Policy, Ethics, and Professional Concerns (PEPC), Johanne Jean-Pierre, Qiang Fu, and Rochelle Coté, as well as anonymous members of the CSA Executive Committee, for their careful and insightful feedback on the report as it was nearly finalized.

Report Funded By

Canadian Sociological Association (CSA)

Ontario Confederation of University Faculty Associations (OCUFA)

Executive Summary

Background and Justification

In recent years, mounting evidence of the dubious validity of Student Evaluations of Teaching (SETs) has generated a series of challenges to their use in decision-making on faculty careers across North America. The most noteworthy of these culminated in June 2018 with an Ontario provincial arbitrator ruling in favour of the Ryerson University Faculty Association, whereby SETs scores should no longer be used for tenure and promotion decisions. This established an empowering precedent for faculty elsewhere. However, largely absent from the scholarly literature on SETs, and from the Ryerson decision itself, is what to do with the open-ended feedback students provide on SETs, which faculty are expected to utilize to improve their teaching. There is sparse research into how these qualitative comments affect faculty mental health, well-being, professional self-esteem, and pedagogical practice. Yet, anecdotal accounts of unfair, hostile, even harassing comments are increasing with the transition to online delivery of the survey instruments. The potential for unevenness in psychological burden that this imposes raises issues of equity, given that identity-based biases in SETs scores are well documented.

While chairing the PEPC in 2017, one of the three authors of this report, Lisa Kowalchuk, obtained financial support from the CSA for a survey-based study to address these questions. A sharper picture of how SETs comments affect Sociology

faculty's mental health and well-being, as well as their pedagogy, would provide a basis for recommendations to Sociology departments across the country regarding improved approaches to obtaining student feedback.

Methodology

The target population for our study were faculty employed as professors or instructors in Canadian Sociology departments, or departments which combine Sociology with another discipline. The study received ethics clearance through the Research Ethics Boards (REBs) of the University of Guelph and the University of Toronto. Our online survey consisting of 26 questions, including three that were open-ended, was emailed to 1,179 faculty in August 2019, attaining a response rate of 26.4% (n=311).

“

Underpinning our study is the assumption that university students and professors both gain when the former attain skills for providing respectful and constructive feedback. Further, we believe that professors' improvement and growth as educators are maximized when teaching conditions foster their professional morale and holistic well-being.

”

Hypotheses and Key Quantitative Findings of the Survey

H1. A majority of sociology instructors engage with their qualitative SETs comments and use them to improve their teaching practice.

Confirmed: over 90% reported that they read their comments.

H2. A majority of sociology instructors make pedagogical choices to preempt hostile feedback.

Not confirmed for the sample overall: the percentage of those who curb innovation and reduce difficulty levels was only about 40%, less than the majority but substantial nonetheless. There were noteworthy differences by gender, race and age:

- ◆ The percentages of women who avoided innovation (46%), reduced the amount of material (51%), and reduced difficulty level (46%), were consistently higher than men; 27%, 34%, and 34% respectively.

- ◆ Forty-nine percent of racialized respondents reduced the amount of material, and 46% reduced the difficulty level, compared to 40% and 39% of their non-racialized peers.

- ◆ Among those aged 40 and under, 45% avoided innovation (compared to 33% for those 60 and older), 51% reduced amount of material (compared to 35%), and 49% reduced difficulty level (vs. 37%).

- ◆ Only small differences were found on these items by sexual orientation.

- ◆ *A related finding concerns how faculty use the feedback to modify pedagogy. Only one-third of the*

overall sample regarded SETs comments as an effective means of determining teaching effectiveness, and 38% believe that the comments have contributed to eroding academic standards. Yet, almost two-thirds reported that SETs comments are helpful in improving course design. One reason for this apparent contradiction may be that faculty define improvement in terms of achieving better SETs scores and comments in future courses, by pre-empting abusive feedback. Further research is needed to probe this possibility.

H3. A majority of sociology instructors will have experienced more negative than positive effects of SETs comments on their self-esteem; self-confidence in their ability to teach; morale; motivation for teaching; and emotional well-being.

Confirmed. Only a minority of respondents – ranging from 20% to 29% – indicated that SETs comments positively affect these key aspects of professional self-esteem. Conversely, when asked if anticipation of SETs comments provokes anxiety, 58% reported affirmatively.

H4. A majority of sociology instructors will receive comments that reflect pedagogically irrelevant, personalized criteria.

Confirmed for our sample overall, with percentages ranging from 71% to 90% agreeing that the comments reflect ascribed characteristics of the instructors themselves rather than pedagogical criteria.

H5. Higher percentages of faculty from equity-seeking groups will receive comments that do not reflect pedagogical criteria, and are hostile and abusive.

Largely confirmed:

◆ Women, more than men, reported receiving SETs comments with abusive, hostile or bullying content (64% vs 55%). They more often experienced SETs comments as biased by factors unrelated to pedagogy (with differences ranging from 9% to 26%), found SETs comments less helpful to improving their pedagogy, and made more modifications to their teaching practice (reducing the difficulty level), to avoid unconstructive feedback. They were also praised more often for adhering to gender stereotypes of nurturance, and less commended for intellectual abilities.

◆ Racialized respondents more often received abusive and hostile content in their comments than non-racialized faculty.

◆ A higher proportion of LGBTQ respondents considered SETs comments to be reflective of non-pedagogical criteria.

H6. Higher percentages of faculty from equity-seeking groups will report negative effects of SETs comments on well-being.

Confirmed for women and racialized faculty:

◆ Women more often reported that SETs comments negatively affected their well-being (39% vs. 25%), and that SETs provoke anxiety due to fear of negative feedback (68% vs. 45%).

◆ Racialized respondents also experienced more negative effects on well-being from SETs comments than their non-racialized peers: 49% vs. 32%.

The findings were less clear cut for sexual minorities, demonstrating only small differences in effect on overall well-being and anxiety, and a difference in the opposite direction for receiving abusive or bullying comments compared to their heterosexual peers.

H7. We anticipate intersectional effects: a higher proportion of sociology instructors who belong to more than one equity-seeking group will experience these negative effects on well-being.

Confirmed. SETs comments had a more negative impact on sense of well-being, and were more anxiety-provoking, for respondents who belonged to two or more equity-seeking groups, than those who belonged to only one or none. These respondents also more commonly saw SETs comments as reflective of non-pedagogical factors; especially gender, race, age and language proficiency, and more commonly received abusive comments.

H8. Contingent instructors will report more negative impacts of SETs comments on well-being and professional self-esteem.

Confirmed for only one of our indicators of well-being and professional morale: SETs comments have prompted twice the proportion of contingent faculty, who make up 18.6% of our total sample, to consider leaving academia, compared to tenured/tenure track faculty (33% versus 13%).

◆ Contingent faculty more often received abusive comments than tenured/tenure-track counterparts (74% vs. 58%), and the abusive comments they received were more often experienced as a threat to safety.

Key Qualitative Findings of the Survey

Of the three open-ended questions in the survey summarizing faculty reflections on SETs qualitative feedback, we distilled four major themes in the responses: pedagogy, personality, personhood, and perceived identity.

Pedagogy

Respondents most frequently recalled a combination of polarizing favourable and harshly disapproving feedback. The least favourable comments recount classes (or instructors themselves) as boring, incomprehensible, repetitious, worthless, and biased. Alongside these highs and lows are the emotional rollercoaster respondents described as they attempt to make pedagogic sense of the qualitative feedback.

Personality, Personhood and Perceived Identity

Equity seeking groups – including women, people of colour, and those whose first language is not English – more frequently recalled student feedback about non-pedagogic factors. Students' comments about pedagogy were often hard to disentangle from comments regarding personality, personhood, and perceived identity. Nevertheless, SETs comments on the latter take a greater emotional toll.

Personality

Many respondents experienced SETs as a popularity contest, reflecting socially biased perceptions and expectations of teaching personality, at times in lieu of critical feedback. Faculty's affective responses to assessments of personality included anxiety, sadness, hurt, confusion, annoyance, disappointment, and resentment. While not all student comments are inherently negative, few respondents detailed long-lasting, positive emotional impact.

Personhood

These student comments were directed at personal or professional character, and were exclusively negative. We also found that they were more prevalent amongst SETs received by instructors from marginalized backgrounds – namely with respect to gender and race.

Perceived identity

Gendered stereotypes are reflected in expectations of emotional labour and flexibility in exercising course policy. Gender is also activated in students' assessments of physical appearance/attraction and presentation of self. For respondents who identified as women and genderqueer/genderfluid, students' evaluations of gender represented norms or expectations of physical appearance/attraction and presentation of self, and were routinely experienced as demeaning. Additionally, students often bring instructors' race and ethnicity into judgements of course material and how they present it, referring to instructors as biased. Respondents' comments about how age figures in student feedback were exclusive to women and intertwined with sexist remarks. They reflected discrimination toward both younger and older faculty depending on the category of assessment. Regarding language, perceived accent, and voice, respondents' experiences demonstrated that some accents are interpreted as signals of intelligence, knowledgeability and even condescension, while others are not.

Recommendations

- 1.** The overarching goal of open-ended commentary from students on the quality of teaching must be solely formative, not summative or comparative. That is, it should be one source of pedagogic insight for course development, growth, and improvement, rather than used to appraise and rank overall teaching performance relative to peers.
- 2.** Relatedly, comments on teaching should be solicited as courses are underway, and not at the end of the semester as has long been the convention.
- 3.** Faculty must be allowed to decide on the content of the open-ended questions for student feedback on their courses.
- 4.** Before completing SETs surveys, students should be trained regarding their use, both as an instrument for pedagogic improvement and as a measure of teaching effectiveness in faculty's tenure and promotion files.
- 5.** University administrations should invest in alternative or supplementary methods of evaluating teaching to replace end-of-term surveys of students' opinions.
- 6.** End of term SETs, both ratings and comments, should be suspended immediately for contingent faculty.

Introduction

A vast literature on the appropriateness of student evaluations of teaching (SETs) as a means to gauge and reward university instructor's competence has been accumulating over the past three decades (Singh Kang and Sidhu 2015). Indeed, "the subject of student rating of teaching and courses is one of the most highly researched in education" (Ogier 2005: 477). More recently, scrutiny has translated into action. Mounting evidence of the dubious validity and identity-based biases of SETs, and the exacerbation of these problems with the transition to online delivery of the survey instruments, has equipped organized challenges of their use in decision-making on faculty careers. To date, the most significant of these challenges was in June 2018, in Toronto, Ontario, where a provincial arbitrator awarded Ryerson University's Faculty Association a victory in a grievance filed in 2009, over the use of SETs for evaluating the performance of its faculty for tenure and promotion. This decision appears to have set an empowering precedent for faculty unions across North America (Bothwell 2018; CAUT 2018; Farr 2018).

A striking feature of most of the scholarship on SETs, and of the literature and expert testimony that informed the Ryerson decision, is a focus on the numerical scores generated by the closed-ended questions. Fewer studies examine the content and impact of the comments that students provide on open-ended questions that are included in most

SETs instruments. A probable reason for this imbalance in scholarly analysis of SETs is that the numerical ratings are seen to play a greater role in decisions around hiring, contract renewal, and performance evaluation. This in turn owes to their heuristic convenience based on the assumption, increasingly interrogated, that the scores can be directly compared with departments and across instructors of the same faculty (which many instruments are in fact set up to do). As well, problems of validity and reliability in the scores, including biases against particular groups, can be compellingly demonstrated or falsified through quantitative analysis.

But SETs comments, too, can affect faculty careers. When students sign their SETs surveys, their free-form comments are typically available to evaluative committees. More importantly, instructors are expected to engage with students' qualitative feedback and account to peers and administrators for how it informs their teaching practice. SETs comments have the potential not only for nuancing numerical scores, but also for personalized judgements not relevant to pedagogy. Though this has always been inherent to SETs comments, it can only be amplified by the nearly ubiquitous transition to the online mode of survey completion. The greater leeway for hostile and even harassing comments in SETs, which are reported to be on an upswing (Eidenger 2017), carries implications for emotional well-being, sense of safety, self-concept, morale, and even the

teaching abilities of instructors. Given the dearth of research on mental health among professors as an occupational group in general (CAUT 2018; Flaherty 2015), these are burdens insufficiently explored, and may be unevenly borne by particular groups.

Our report seeks to understand the impact of SETs comments on these dimensions of emotional and mental well-being through a survey of instructors (n=311) in Sociology departments across Canada in the summer of 2019. We also looked at respondents' views of the usefulness of SETs comments in their own teaching practice and in general. Ironically, there has been little research into the influence that SETs have on teaching practice, with the noteworthy recent exception of Omer et al (2020 forthcoming). For our study, we confined the sampling frame to Sociology and kindred disciplines and/or interdisciplinary programs, allowing us to control for factors that may vary somewhat between fields, such as the types of topics being taught.

To briefly summarize our key findings, almost two thirds of the respondents report that they receive abusive, hostile, or bullying content in their SETs comments, a proportion that is considerably higher for women and racialized faculty. These results echo a recent national survey of Australian faculty (NTEU 2018). Nonetheless, our respondents overwhelmingly engage with their SETs comments as opposed to ignoring them. A third report a negative impact on their overall emotional well-being, but this is much higher for women and racialized groups. Asked specifically about anxiety over SETs comments, well over half answered affirmatively, a proportion that is, again, higher for women, racialized people, LGBTQ

and those 40 and under. For the majority of the sample, SETs comments reflect judgements that are not relevant to pedagogy, most often pertaining to personal traits of the professor, class size, and expected grades of students. Again, there are disturbing differences by gender, sexual orientation, and age.

We anticipate that SETs will be reconfigured but maintained at many institutions, perhaps with changes in the parameters around their use, instructions to students, etc. If diminishing weight is placed on SETs quantitative ratings in light of the Ryerson decision, instructors may be expected to engage more deeply with, and demonstrate more fully their use of, the qualitative comments in hiring, contract renewal, and performance evaluations (Omer et al 2020, forthcoming). Our overarching goal, then, is to establish grounds for rethinking the qualitative component of SETs, through a better understanding of how instructors in social sciences experience and use them. In particular, attention is needed to the goals of soliciting comments in SETs, as well as control over their content, the timing of questionnaire administration, and institutions' investment in maximizing the constructiveness of student commentary and in innovations in additional, supplementary methods of evaluating teaching. Underpinning our study is the assumption that university students and professors both gain when the former attain skills for providing respectful and constructive feedback. Further, we believe that professors' improvement and growth as educators are maximized when teaching conditions foster their professional morale and holistic well-being.

Literature Review

Contextualizing SETs Historically

Since they were first introduced in U.S. universities in the 1920s, SETs have been broadly similar in their basic structure and mechanics: administered near the end of the academic semester, they are typically composed of scaled questions on a set of indicators deemed to operationalize the quality of students' experiences in courses. As well, there is usually an option for students to provide open-ended commentary for each item (Algozzine et al 2004). For decades, SETs were a means by which professors could opt to solicit feedback directly from students in the 1970s, pressures from a more empowered student body, and from governments demanding greater quality and accountability of universities (Calkins and Micari 2010; Langen 2011), led administrators to increasingly usurp the delivery of SETs and impose them across the board. A steady rise in their use, from 29% to 86% of institutions from the early 1970s to early 1990s, has been charted by several analysts (Stroebe 2016). This evolution in the control and purposing of term-end evaluations as a source of information for managerial objectives (Algozzine et al 2004; NTEU 2018), and the spread of this model throughout the US, Canada, and much of the Western world, coincides with reduced public spending on higher education, and reduced university investment in teaching staff. ¹ For those in the professoriate who are most affected by budgetary austerity, namely the

ballooning numbers of those employed on a contingent, part-time basis, SETs have been an even more important determinant of their prospects for hiring and contract renewal (Heller 2012; Langen 2011).

Research on Biases in SETs Ratings

Not all of the scholarship on the merits of the scores generated by SETs concludes definitively that the surveys are invalid, unreliable, or biased. But the bulk of research done in the past decade, including a ground-shifting meta-study that addresses long-overlooked flaws in many prior studies (Uttil et al 2017), shows SETs to be at best a very problematic quantitative gauge of teaching effectiveness, and a source of occupational discrimination against already disadvantaged groups. Arguments in defense of the predominant way that SETs are designed and used become even more untenable with the digitization of the survey instrument. While the transition to online SETs was bound to alter results of both the scores and the qualitative comments, for the former, the dramatic decline in response rates (Vasey and Carroll 2014) makes their outcomes little more than “pseudo-statistics” (NTEU 2018). As well, experimental research supports the existence of validity problems that had long been suspected. It has been shown that far from measuring pedagogical performance, SETs results are better indicators of, for example, the pleasantness of the weather on the date of the SETs survey (Braga et al 2014), the provision of tasty

¹ In Ontario, an additional impetus for deepening the reliance on SETs may have been the 2012 Drummond Commission which ramped up the use of metrics to evaluate institutional performance across the civil service (OCUFA 2019).

treats prior to survey completion (Hessler et al 2018), and grade expectations (Boring et al 2016). The study by Braga et al (2014) suggests that SETs scores are inversely related to teaching effectiveness, lending credence to the argument that SETs have a paradoxically perverse effect on teaching and learning.

Studies of identity-based bias illuminate deeper, equity-related problems with taking SETs as a valid measure of teaching quality. There is simply too much evidence that the scores reflect discrimination toward equity-seeking groups to warrant their use for summative and comparative purposes, if universities claim to respect and uphold human rights of all their members. With regard to gender, three recent experimental studies compellingly support the existence of bias against female instructors (Boring et al 2016; MacNeill et al 2015; Mitchell and Martin 2018). In the most recent of these, “women are rated more poorly than men even in identical courses and when all personality, appearance, and other factors are held constant,” for every one of 23 items on the survey (Mitchell and Martin 2018: 1).

In an earlier experiment that exposed students to a lecture by a gender- and age-neutral stick figure, but in which gender and age identities were assigned on the pseudo-evaluation forms, younger male professors received the highest ratings and were more positively valued for their expressiveness. This interaction of instructor gender and age suggests that the criteria students apply to evaluate teaching vary by these and other ascribed traits of the instructor: “[S]tudents may expect their professors to express themselves in the classroom according to sociocultural scripts that

limit full human development and influence people to appraise men and women differently even when they are displaying identical behaviors” (Arbuckle and Williams 2003: 514). Research has also found that ratings are affected by the sexual orientation of instructors, disadvantaging LGBT instructors albeit in subtle and contradictory ways (Ewing et al 2003).

Student ratings of professors have also been found to correlate with the race and ethnicity of instructors. On specific dimensions of teaching quality, and in questions on overall effectiveness, US-based studies have found lowest scores given to African Americans, and highest scores to whites (See for example, Bavishi et al 2010; Smith and Hawkins 2011). This pattern is mirrored in a study of ratings on the RateMyProfessors.com (RMP) website (Reid 2010). A subtype of ethnic and racial bias in SETs concerns language and accent. Student ratings have been found to penalize instructors whose first language is presumed to differ from the predominant language of the institutional setting. Analysis of official SETs responses in one New Zealand university finds that in science- and math-based courses especially, ratings on communication skills have a disproportionate effect on the overall effectiveness rating (Ogier 2005). Similarly, a study of RMP ratings and comments for math and statistics courses found those with Korean and Chinese last names score appreciably lower than their counterparts, especially on the clarity and helpfulness, 0.6 to 0.8 points lower on a 5-point scale. Ogier (2005) implicitly accepts that students in his study are responding to an objective effect of language fluency on communication ability. But an alternative and more disturbing explanation is that

there is at play an “ideological construction of the [non-native English Speaker] as incomprehensible Other” (Subtirelu 2015: 35). Experimental research has yielded damning evidence of the impact of job applicants’ names on race-based labour market discrimination in a number of fields (Bertrand and Mullainathan 2004). Arguably, similarly refined methods could help to determine whether SETs yield weaker scores on communication and more negative comments on language abilities based solely on visible, audible, and textual (last name) markers of a marginalized race-ethnicity.

For racialized instructors especially, but also other equity-seeking groups, biases in SETs add to the “cultural tax” (OCUFA 2019) or “identity tax” (Hirshfield and Joseph 2012, cited in Eagan and Garvey 2015) that they incur as they are prevailed upon inordinately for additional labour toward equity goals in the institution and community. A glance at the comments section on online essays critical of SETs reveals there are still defenders of the status quo use of the quantitative scores for summative purposes, in some cases dismissing findings of bias as inconclusive and or contradicted by other studies, and recommending small tweaks such as mentoring newer faculty to cope with and improve their ratings. But an increasing consensus is emerging that evidence of any bias at all supports a call for radical overhaul of how closed-ended questions are designed and used (OCUFA 2019). Age figures much less prominently in the literature examining correlations of SETs scores with ascribed characteristics of instructors (Arbuckle and Williams 2003; Basow and Martin 2012). It is widely

understood that western, wealthy societies have a cultural anti-elder bias, but predicting how age factors into assessments of teaching effectiveness is not straightforward. Arbuckle and Williams’s quasi-experimental study finds that age interacts with gender in numerical ratings on various aspects of teaching style, with being both young, and male, conferring an advantage (2003). Basow and Martin (2012) extrapolate to predict that instructors who are older and female likely face lower SETs ratings given society’s sexism and ageism. However, this probably depends on what exactly is being assessed; for example, being young and female could negatively affect perception of knowledgeability of subject matter.

“
It has been shown that far from measuring pedagogical performance, SETs numerical ratings are better indicators of, for example, the pleasantness of the weather on the date of the SETs survey the provision of tasty treats prior to survey completion, and grade expectations based on grades already received.”

“
There is simply too much evidence that SETs scores reflect discrimination toward equity-seeking groups to warrant their use for summative and comparative purposes, if universities claim to respect and uphold human rights of all their members.”

The Growing Concerns with SETs Comments

Regardless of the format in which they are delivered, written appraisals of a person's performance and abilities have a greater potential impact on emotion, both positive and negative, than do numerical ratings alone. Simply put, there are an infinite number of ways that judgements can be verbalized. On SETs, answers to the free-form questions do not always correspond with results on the quantitative scaled questions; negative commentary is often provided alongside relatively good scores, and vice versa (Freishtat 2016). Written feedback also has the potential to veer away from the pedagogical objective of the survey item. Examining the responses on the qualitative component of SETs allows an understanding not just of the frequency of certain types of comments that appear, but also the themes that students themselves consider relevant in appraising their instructors.

While less prevalent in the literature, scholarly attention to the qualitative components of SETs surveys has been increasing in the past decade. Undoubtedly this owes to growing concern that with the shift away from paper surveys completed in the classroom, and the explosion in popularity and influence of the anonymous, online, unofficial platform RateMyProfessors.com (RMP), the content and tone of the comments are deteriorating. Obviously, it is impossible to determine the precise effect of digitization on SETs comments over time, which has at any rate occurred alongside the rise of anonymous rating platforms and social media. Certainly, critical

comments that are unconstructive, cruel and even obscene predate the shift (Lindahl and Unger 2010) and may be explained in part by an increasing consumerist view of education on the part of students forced to pay more for it. In fact, the content of hostile commentary has been found to resemble customer dissatisfaction with a product (Davison and Price 2009; Lindahl and Unger 2010).

But academics consulted or surveyed on this issue have largely observed that the frequency of "abusive and bullying" trolling-style content is increasing (NTEU 2018; OCUFA 2019; Vasey and Carroll 2014). Sixty per cent of Australian professors surveyed in 2018 by the National Tertiary Education Union (NTEU) whose SETs had open-ended sections reported receiving abusive comments. Of these, 30% were about appearance; 27% were about religion, culture, sexuality, or disability; 14% were about English ability. In the words of one Canadian academic, SETs comments today function as "institutionally sanctioned instruments of harassment" (OCUFA 2019), an interpretation that undoubtedly reflects the nearly ubiquitous shift to online delivery.

The RateMyProfessors.com (RMP) platform itself is often analysed as a proxy for university administered SETs. This is justified on several grounds: both the numerical ratings and the comments in RMP have been found to align with official SETs scores (Reid 2010; Sonntag, Bassett and Snyder 2009); it is a massive database of responses to questions that are

conveniently standardized; students heavily use the site, which can have career implications if they avoid courses rated unpopular on the site. RMP also has implications for scholars' broader reputation, given that one's RMP standing is often among the first items to emerge in Google searches, ahead of publications and other achievements (Mitchel and Martin 2018). While most RMP studies look at scores rather than comments (Subtirelu 2016), a few do both (for example Davison and Price 2009; Kowai-Bell et al 2012; Storage et al 2016; Subtirelu 2016). Obviously, there is no replacement for analyzing the real comments on university-administered SETs, since these are what instructors are obligated to engage with. But as a well-established forum for unaccountable opinions, RMP may be feeding into a broader trolling culture that increasingly influences the comments in official SETs (Freishtat 2016; Flaherty 2015). The site has allowed and encouraged the rating of professors based on physical attractiveness, which amounts to anonymous sexual harassment and contributes to a hostile work environment. ²

Some studies of abusive comments blend into the discussion of a more general category of aggressive behaviour directed toward faculty in and outside the classroom. For example in Deo's study (2015), SETs comments mirror verbal confrontations that female professors of colour experience from students both in and outside of the classroom. For example in Deo's study (2015), SETs comments mirror verbal confrontations that female professors of colour experience from students both in and outside of the classroom, ranging from challenges to

their competence, to outright disrespect. This is often framed as contrapower harassment, a phenomenon that affects more women in academe than men (Lampman et al 2016). A subtype of this to which scholars have recently turned attention is student cyberbullying of faculty, encompassing email, entries in RMP, and the use of social media. Small sample studies suggest a gendered pattern in which far more women than men are targets of this kind of online behaviour (Blizard 2016; Cassidy et al 2014).

It is not simply the frequency of approving versus negative, hostile comments, that is of interest. The themes that arise in comments also say something about the criteria by which professors are judged either positively or negatively in accordance with stereotypes and cultural scripts. With regard to gender, Mitchell and Martin's (2018) quasi-experimental study of two online courses found that students used different words to evaluate male and female professors; the latter are more often referred to as "teachers" (vs. "professors") and receive more comments about personality and appearance, while male "professors" receive more comments on intelligence. A similar finding on the interaction of race and gender is seen in a content analysis of terms utilized by users of RateMyProfessors.com: the words "genius" and "brilliant" appeared more often in entries on men than on women, and there are far fewer African-Americans and women in the disciplines where these descriptors are most often used (Storage et al 2016).

² In July 2018 the physical attractiveness item on which RMP invites ratings of looks, symbolized by a hot chili pepper, was removed in response to collective online advocacy by mainly female faculty in the U.S (Flaherty 2018).

SETs Comments and Emotional Well-being

The degeneration that has almost certainly been occurring in the tone of SETs comments throws into question the appropriateness of mandated engagement with them as a guidepost to pedagogical improvement. There are also emotional and mental health implications. The broader topic of mental health in relation to occupational stressors remains understudied for the professoriate (CAUT 2018a; Flaherty 2015). One of the few studies of this type, a UK-based national survey, finds that despite their comparatively high degree of control and autonomy over what they do, one third of the 14,000 respondents had unacceptable levels of stress (Kinman and Wray 2013). This is echoed in a recent small scale (n=5) study focused on a B.Sc. Nursing program in Canada in which instructors reported “feeling distressed, overwhelmed and tormented by negative comments” and described SETs comments as “an avenue to bully instructors.” (Moralejo et al 2019: 6) *italics in original*). There was also an array of physical symptoms that some or most of the participants experienced, including muscle tension, difficulty sleeping, excessive eating and alcohol consumption. They tended to blame and stigmatize themselves for negative feedback, and did not talk about it with others (Moralejo et al 2019). Not all types of occupational stressors have the same effects; for university faculty, those that constitute threats, such as discrimination and hostile environments, are found to diminish productivity, in contrast to those which combine challenges with fulfillment (Eagan and Garvey 2015). Broad structural transformations

in higher education are likely to generate increased anxiety and depression in this occupational group: the reduced availability of stable employment; a consumerist orientation on the part of students paying higher fees as public subsidies decline; and the increased use of metrics for quality assurance of entire institutions (Ehrenberg 2011). Digitized technologies in teaching also have a bearing; the increasing number of courses being taught online, often by contingent faculty who are already less networked than their tenure-track peers, carries the mental health hazard of isolation (Reeley 2016). Though none of these studies cite SETs as a factor, SETs scores are among the metrics for assessing professors’ achievements, and the consumerist orientation noted by several experts has implications for both scores and comments.

“

The degeneration that has almost certainly been occurring in the tone of SETs comments over the past decade throws into question the appropriateness of mandated engagement with them as a guidepost to pedagogical improvement.

”

We know of only one systematic, large scale study of SETs that looks for the impact of hostile and bullying SETs comments on psychological and physical well-being. The Australian survey mentioned above reports that of the instructors who received abusive comments, 70% experienced negative emotions, while 11%, 38.5%, and 34%, respectively, experienced loss of appetite, inability to sleep, and anxiety and depression. For these negative health outcomes, percentages were higher for some equity-seeking group such as women, LGBTQ, Aboriginal faculty, faculty with disabilities, and those whose first language is not English (NTEU 2018).

Given that negative student comments, regardless of whether they are constructive, are found to lower the recipients' sense of self-efficacy in teaching (Boswell 2016), we would expect this effect to be heightened by comments that verge into bullying and harassment. Indeed, 45% of those in the Australian survey reported that abusive comments made them want to avoid teaching (NTEU 2018), again with higher percentages for women and LGBT respondents. A similar impact is also found in a small sample study in the US (Lindahl and Unger 2010). Though recipients of such comments feel demoralized, hurt, and angry, they generally tend not to discuss them with colleagues out of embarrassment (Lindahl and Unger 2010). In light of studies on the unequal distribution of abusive comments across particular identity groups, we are disturbed by even suggestive evidence that this may prompt exiting an academic career. If university administrations are to live up to the claim of valuing diversity, and in Canada, the goal of reconciliation with Indigenous peoples,

more attention is needed into how members of equity-seeking groups are affected by SETs comments in the current climate.

SETs' Impact on Teaching and Learning

The question of whether and how instructors use SETs to modify their pedagogy is under-researched (Omer et al 2020 forthcoming). A few studies from the 1970s and 80s, before the massive shift to online SETs, found that most instructors do use them to improve their teaching (Murray 1997). Clearly, the manifest function of SETs is to improve teaching by signaling to instructors what elements of their practice to change or retain. But a number of studies taken together suggest a paradoxically opposite effect. Stroebe (2016) employs psychological theory to analyse prior scholarship on SETs, drawing causal links between empirical findings on SETs scores and students' grade expectations, SETs scores and student learning, and instructor tailoring of courses with SETs in mind. In one set of studies, SETs scores are found to vary proportionately with the grades students expect, based on grades already received and the amount of work required to attain them. Another cluster of studies suggests that student learning is perversely hindered by SETs: the higher students rate a course in the SETs (which in turn reflects a more lenient, less challenging standard that they tend to reward), the worse they perform in more advanced levels of the same subject. There are also studies showing instructor awareness that making a course easier, or grading more leniently, leads to better SETs scores, and that they tailor courses accordingly. For Stroebe (2016), universities'

increasing reliance on SETs scores for summative purposes since the 1970s explains why undergraduate GPAs and letter grades have been rising over the last 30 years, while the hours they expend in studying has declined just as markedly over that period. In other words, SETs contribute to grade inflation and lower levels of learning.

These studies focus exclusively on the pedagogical effects of SETs ratings, but what about the comments? We have seen no literature on this, no doubt because the comments are seen as less influential than scores in the administrative processing of SETs. However, the fact that instructors are found to make more use of SETs comments than scores (Omer et al 2020 forthcoming; Smith and Wellecker-Pollack 2008) suggests that they provide useful and constructive information. In light of survey evidence that professors are increasingly receiving personalistic, pedagogically irrelevant comments as SETs have shifted online, it is possible that instructors make choices to avoid abusive content, even as they use the feedback to improve pedagogy. It is also possible that their effort to engage with the feedback in search of constructive information comes at a cost to their emotional well-being.

SETs and Contingent Faculty

In light of the immense growth in the proportion of courses taught by contingent faculty over the past twenty years in Canada and elsewhere, and the array of disadvantages they face as precarious workers (Foster and Birdsell Bauer 2018), it is critical to consider how they are affected by SETs. Throughout this report we use the terms “contingent faculty” and “contract academic staff” interchangeably. We borrow the latter term from Foster and Birdsell-Bauer

(2018), who designate it to encompass a variety of employment status titles used at different institutions, all of which have in common the fact that “they do not work in tenure-stream positions, they are usually paid only to teach (not to research or do administrative work), and their jobs are not, in any robust sense, permanent” (Foster and Birdsell-Bauer 2018: 14).

There is very little systematic study of contingent professors’ experiences with SETs generally, let alone the open-ended feedback.³ This is surprising given that teaching is the primary component in their performance evaluation, and that SETs are the main or only metric for this (Langen 2011; Murray 2019). A crucial difference between the contingent and full-time faculty experience is that for the former, there are no concrete rewards for outstanding SETs ratings or comments, and yet poor SETs results can be grounds for job loss (Lewontin 2014). Cases in both Canada and the US in which SETs comments appear to have been used by administrators to dismiss contingent faculty members substantiate this (Smele et al 2020, forthcoming); Reichman 2020). In the US case, the professor had been active in unionizing fellow part-timers (Reichman 2020). This suggests that SETs contribute to reducing contingent faculty’s access to standard, on-paper protections of academic workplace freedoms.

In the only in-depth study we know of that examines precariously employed faculty members’ experience of SETs, an interview-based study of women faculty at one Ontario University (Smele et al 2020, forthcoming), respondents are overwhelmingly dissatisfied with SETs use as tool for evaluating their teaching. They feel enormous pressure to relax standards

³ For an exception, see (Smele et al 2020, forthcoming), who analyse the experiences of contingent women faculty with SETs, both the ratings and the comments.

for grading, deadline adherence, and plagiarism because of SETs, echoing anecdotal accounts from the US (Lewontin 2014). It is perhaps not surprising that SETs are found to have considerable and longlasting ill effects on emotions and mental health of contingent faculty, encompassing in some cases clinical depression (Smele et al 2020, forthcoming). These findings and observations call for immediate reconsideration of the use of SETs, for both the scores and the comments, for this group.

Institutional Responses

We have come across a number of studies that minimize or dismiss concerns with hostile SETs comments and their impacts. Methodologically, some of these utilize content analysis of massive numbers of raw comments and quantify those deemed problematic, rather than surveying instructors regarding their experiences (see for example Alhija and Fresko 2009; Brockx et al 2012). These tend to find minuscule percentages of abusive comments. Other studies attribute problematic comments or their psychological impact to the newness and inexperience of the instructor. A variant of this signals gender differences in cognitive and affective processing of SETs results as the problem -- women get basically the same kinds of comments and scores as men do, but interpret them more self-critically (Kogan et al 2010). The solution commonly recommended in these studies is for universities to facilitate additional mentoring of the younger or newer faculty, especially women, to cope better with the comments and to improve their practice to attain better SETs results going forward (Floden 2017; Kogan et al 2010; Lutovac et al 2017; Wong and Moni 2014). While we consider such measures to be superficial and avoidant of the glaring need for deep overhaul of how SETs comments are collected and used, it is a shortcoming of the universities that they provide minimal supports to

women and minority faculty targeted by abusive comments (OCUFA 2019). This is especially the case for sessional instructors who generally enjoy less access to institutional resources (Eidinger 2017). An additional solution that has been discussed, but to our knowledge not yet implemented, is the deployment of staff to filter out toxic comments before the instructor receives them. But this would deprive the instructor of the ability to demand and utilize procedures to hold aggressors accountable and pre-empt further abuse. In Ontario, and no doubt other jurisdictions, this may also contravene provincial health and safety legislation (OCUFA 2019).

The June 2018 decision at Ryerson University to no longer use SETs scores to evaluate faculty performance for tenure and promotion was part of a wave of resolutions to long-standing struggles over the use of SETs, at least across North America, and probably the most significant. Shortly prior, several US institutions were already moving away from using SETs for summative and comparative evaluation, and more have continued to do so (ASA 2019; Flaherty 2019; OCUFA 2019). Since June 2018, there has been increased pressure on administrations from faculty bodies to move further in this direction; for example, eighteen professional associations in the US alone have issued calls for replacing SETs with more holistic means of assessment, some explicitly citing the Ryerson case (Flaherty 2019). While the ground is shifting in the use of quantitative scores, student input on teaching will still certainly be solicited, at least for formative purposes. If the techniques for obtaining useful qualitative feedback from students are to be overhauled, is it imperative to deepen our understanding of how the open-ended questions on SETs surveys -- as they are currently constructed and accessed -- affect the emotional well-being, sense of safety, and professional morale of all instructors.

Hypotheses

H1. A majority of sociology instructors engage with their qualitative SETS comments and use them to improve their teaching practice.

H2. A majority of sociology instructors make pedagogical choices to preempt hostile feedback.

H3. A majority of sociology instructors will have experienced more negative than positive effects of SETs comments on their self-esteem; self-confidence in their ability to teach; morale; motivation for teaching; and emotional well-being.

H4. A majority of sociology instructors will receive comments that reflect pedagogically irrelevant, personalized criteria.

H5. Higher percentages of faculty from equity-seeking groups will receive comments that do not reflect pedagogical criteria, and that are hostile and abusive.

H6. Higher percentages of faculty from equity-seeking groups will report negative effects of SETs comments on well-being.

H7. We anticipate intersectional effects: a higher proportion of sociology instructors who belong to more than one equity-seeking group will experience these negative effects on well-being.

H8. Contingent instructors will report more negative impacts of SETs comments on well-being and professional self-esteem.

It is complicated to predict any single effect of faculty age on their experience with SETs comments. The literature provides little guidance, as there is even less scholarly attention to comments than to scores when it comes to age. Age may combine with other ascribed traits, and its effects may be curvilinear, such that, for example, both younger and older female-identified instructors are least praised and recognized for being knowledgeable. Thus, we refrain from any specific predictions about this variable.

Methodology

Study Participants

Our target population were sociologists across Canada employed as professors or instructors in sociology departments or departments which combine sociology with another discipline. By limiting the disciplinary background, we aimed to control for the substantive content respondents taught, enhancing confidence in comparisons across demographic variables. Further justifying a single-discipline framework, evidence suggests that instructors in STEM disciplines differ from those in the social sciences and humanities in their satisfaction with SETs, with more of the former regarding them as useful for improving pedagogy (Omer et al 2020, forthcoming).

“
In light of survey evidence that professors are increasingly receiving personalistic, pedagogically irrelevant comments as SETs have shifted online, it is possible that instructors make choices to avoid abusive content, even as they use the feedback to improve pedagogy.
”

Data Collection, Design and Procedure

The on-line survey consisted of 26 questions which focused on the qualitative comments of SETs rather than on the numerical scores. The sampling frame for the survey was based on publicly available e-mail addresses of faculty listed on Sociology department websites of all Canadian universities. We included tenured and tenure-track faculty, as well as contingent or contract academic staff. The Canadian Sociological Association (CSA) also provided e-mail addresses of faculty from their member list, which we included and cross-referenced for duplicates. The sampling frame was collected in the summer and fall of 2018. The survey (see Appendix A) was emailed to 1,179 faculty in August 2019 and was self-administered online using the survey platform Qualtrics in English and French. The response rate was 26.4% (n=311), of whom 93% of participants completed the entire survey. After deleting individuals who did not complete the survey and also an individual who indicated their department did not administer SETs, the final sample size was 288 individuals.

This study received ethics clearance through the Research Ethics Boards (REBs) from the University of Guelph and the University of Toronto.

Limitations of the Study

We wish to address three noteworthy limitations of our study. Firstly, our sample is non-randomly selected, as it comprises respondents who self-selected to participate over the data collection period. Relatedly, our sampling frame – drawn from available e-mail addresses of faculty – is likely incomplete. With an incomplete sampling frame and without known population parameters, post-stratification weights could not be applied to adjust for selection bias. While our results and analyses do not attempt to generalize or make causal arguments about faculty experiences with SETs, we acknowledge these limitations of the data. Secondly, as a limitation of our research scope and in the interest of creating a parsimonious survey instrument, we did not include questions about teaching and research culture at the department, faculty and university levels. Arguably, post-secondary institutions of different sizes and compositions will foster teaching and research cultures that come to bear on students' assessments of teaching effectiveness. In doing so, faculty must contend with feedback from SETs in the context of local academic norms, which may vary widely. Therefore, we wish to identify a third study consideration – the assessment of perception. Our study set out to uncover how Sociology faculty process, understand, use and reflect on qualitative feedback from SETs – pedagogically and personally. In doing so, we capture faculty evaluations of students' evaluations; or more aptly, the social construction of SETs. We wish to note this as a study consideration rather than a limitation, as our study bears insight to nuanced experiences beyond that in currently published SETs discourse.



“

One solution commonly recommended is for universities to facilitate additional mentoring of the younger or newer faculty, especially women, to cope better with the comments and to improve their practise to attain better SETs results going forward. We consider such measures to be superficial and avoidant of the glaring need for deep overhaul of how SETs comments are collected and used.

”

Data Analysis

Descriptive results are presented from quantitative and qualitative components of our survey. A sample characteristics table for respondents in the study are presented in Appendix B. It should be noted that we cannot comment on the representativeness of our sample because there are no comparable statistics about Canadian faculty employed in Sociology departments. In the quantitative analysis, aggregate results when $n < 10$ were not reported, as is common practice by Statistics Canada. A series of cross-tabulations were performed to examine the relationship between variables of interest and the socio-demographic variables, mainly gender, race, sexual orientation, and age. Chi-square tests were not reported as the total sample size for the survey was too small ($n=288$) for chi-square tests, which are influenced by both very large and very small samples. Moreover, it was our view that this report should focus on the size and direction of effects rather than their statistical significance, consistent with the American Statistical Association's statement on statistical significance and p-values (Wasserstein & Lazar, 2016). The statistical package STATA 12.0 was used for all quantitative analyses.

Our coding of gender, sexual orientation and race endeavored to preserve marginalized experiences, rather than erase them. Unfortunately, in the absence of sufficient sample size to analyze experiences across the spectrum, we made the following decisions regarding the quantitative data: 1) Gender

is presented dichotomously (i.e. differences between men and women) as the number of non-binary respondents ($n=15$) were too few to yield meaningful statistical comparison⁴; 2) We use the term LGBTQ to capture sexual minority faculty experiences⁵ ($n=51$) and compare them to heterosexual respondents; and 3) we use the term racialized faculty to capture all racial minority faculty respondents⁶ ($n=40$) and compare them to non-racialized (i.e. White) faculty.

Qualitative data from three open-ended questions on our survey instrument were analyzed; each delineates a dimension of experience with or reflection on qualitative feedback from SETs.⁷ Firstly, of the respondents who answered Q5, 26 respondents communicated they do not read qualitative comments from SETs (~9%), and were asked to provide feedback as to why. Secondly, all respondents were asked to share their most memorable/impactful qualitative comments received from past SETs, of whom 122 contributed feedback (~39%). Lastly, all respondents were encouraged to share anything about student evaluations of teaching that we did not incorporate in our survey instrument; 152 respondents contributed feedback (~49%). Responses from each of the three questions were open coded using NVivo 12 Plus Qualitative Analysis Software. In our analysis, below, respondents' names have been replaced with pseudonyms and identifying information removed to protect confidentiality.

⁴ Nevertheless, as per above, qualitative data from all respondents were analyzed.

⁵ Specifically, the experiences of lesbian, gay, bisexual, sexually fluid and queer respondents were combined.

⁶ We combined those respondents who indicated they were Chinese, South Asian, Black, Filipino, Latin American, Southeast Asian, Arab, West Asian, Korean, Japanese into one category. This category also included those respondents who indicated they were white and another racialized group.

For question wording, see Appendix A: Survey Instrument, questions 5a, 26 and 27. Response rates for the qualitative data are calculated with the full dataset (i.e. before it was cleaned; $n=311$), as all qualitative data from respondents were analyzed.

Quantitative Data

Experience of the Administration and Use of SETs

Respondents were asked to indicate the type of SETs their department or institution administers. A little over half (51%) of the respondents indicated that their department or institution only administers SETs online while a quarter of respondents indicated that their department or institution administers paper student evaluations. Twenty-three percent experience a combination of online and paper SETs, while 1% indicated “other”. The majority of respondents (64%) indicated that their department does not make the results of their SETs public, while 23% said they were able to opt out of making the results publicly available. However, 14% of respondents indicated that their department makes the results publicly available without giving them the choice of opting out.

When respondents were asked how important are SETs when it comes to promotion, salary increases or tenure recommendations, 48% of respondents indicated that quantitative scores were important or very important while 27% indicated that qualitative comments were important or very important. The overwhelming majority (92%) of respondents indicated that they read the qualitative comments on their SETs.

“ Respondents were asked to indicate the frequency with which they received positive comments regarding how caring, empathetic, or sympathetic they are towards students. 83% of female respondents received such comments in comparison to 61% of males. ”

Impact of SETs Comments on Sense of Well-being

In this section respondents were asked questions that focused on the impact of SETs on their emotions, morale, self-esteem, motivation and sense of safety.⁸ Respondents were asked to report which emotional responses they have experienced after reading the qualitative comments on their SETs. They were given a list which included five positive and five negative emotions (see Kogan et al., 2010). Respondents were also given space to list other emotional responses they have experienced.

⁸ We chose not to base this on clinical symptoms of depression and anxiety, as seen for example in Reevey and Deason (2014), because our intent was to capture recollections of emotional states over the course of respondents' careers, not components of clinical diagnoses.

Figure 1: Which emotional responses have you experienced after reading the qualitative comments on your SETs

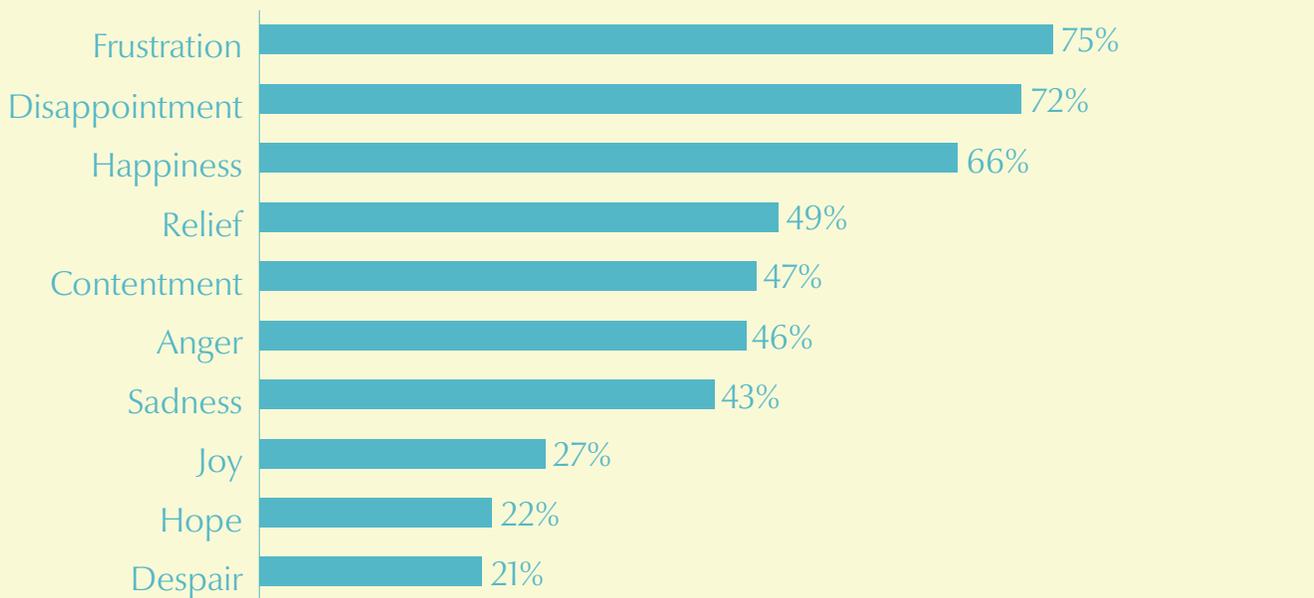


Figure 1 shows that respondents most frequently indicated experiencing frustration after reading the qualitative comments on their SETs, followed by disappointment, and happiness. Respondents least frequently reported experiencing despair, followed by hope and joy.

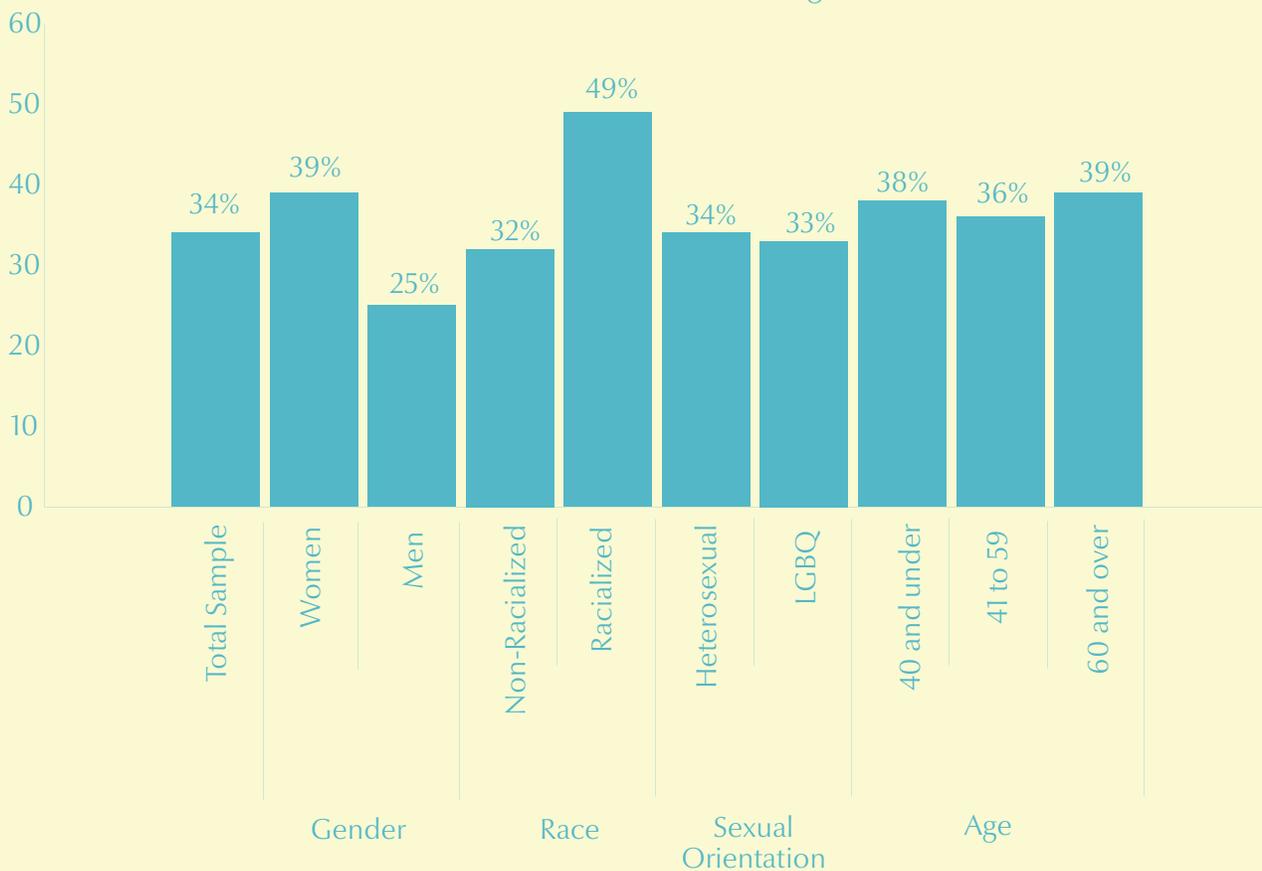
Figure 2: Qualitative comments on SETs greatly increases ...



Figure 2 above shows the extent to which qualitative comments on SETs increase respondents' self-esteem and confidence in a number of areas. The question was scaled from 1 to 5, with 1 being "not at all" and 5 being "a great deal". We combined responses that were "a lot" and "a great deal" in the figure below. Overall, only one-fifth to just under one-third of respondents indicated that qualitative comments on SETs increased their self-esteem or confidence. Analyses by socio-demographic sub-groups indicated numbers that were too small to report.

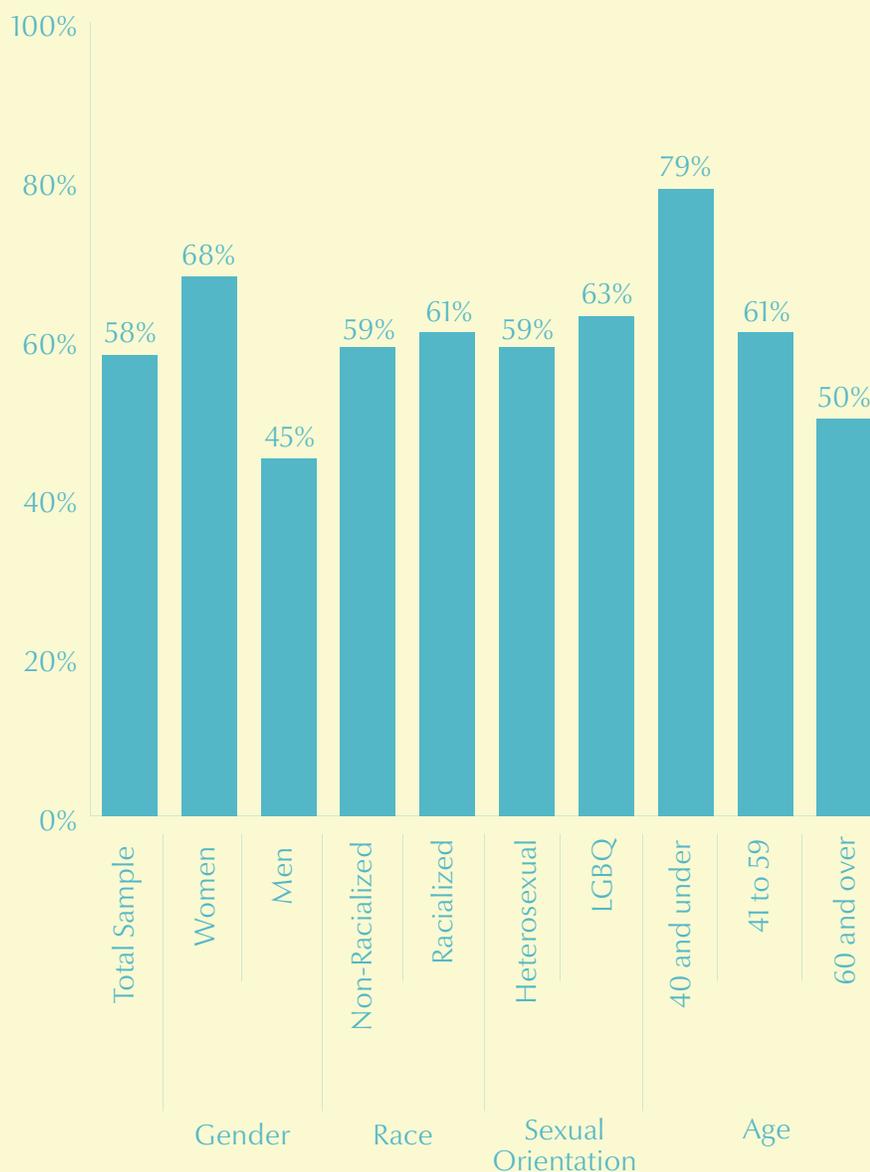
Figure 3, below, shows the proportion of respondents who indicated that qualitative comments from their SETs have had a somewhat or very negative impact on their sense of well-being. Overall, 34% of the sample indicated that the comments from their SETs have had a negative impact on their well-being. These results are commensurate with the findings of recent studies that have looked into emotional implications of SETs comments, including a large survey study of Australian faculty (Boswell 2016; Lindahl and Unger 2010; NTEU 2018; Smele et al 2020, forthcoming). Looking at the breakdown by sociodemographic variables, we can see that women more often reported that the qualitative comments have had a negative impact on their well-being in comparison to men (39% versus 25%). Moreover, 49% of racialized faculty reported that SETs comments have had a negative impact on their sense of well-being in comparison to 32% of non-racialized respondents. There were no meaningful differences by age or sexual orientation.

Figure 3: Qualitative comments from SETs have had a negative impact on sense of well-being



Survey respondents were asked whether the student evaluation period provokes anxiety due to fear of negative feedback. Figure 4 shows that 58% of the sample does experience anxiety during the evaluation period. The proportion was higher for female respondents (68%) than for male respondents (45%). Racialized and LGBTQ faculty were slightly more likely to indicate feeling anxiety during evaluation time. The overwhelming majority of younger faculty (i.e. those who are 40 and under) indicated that they faced anxiety during evaluation time due to negative feedback.

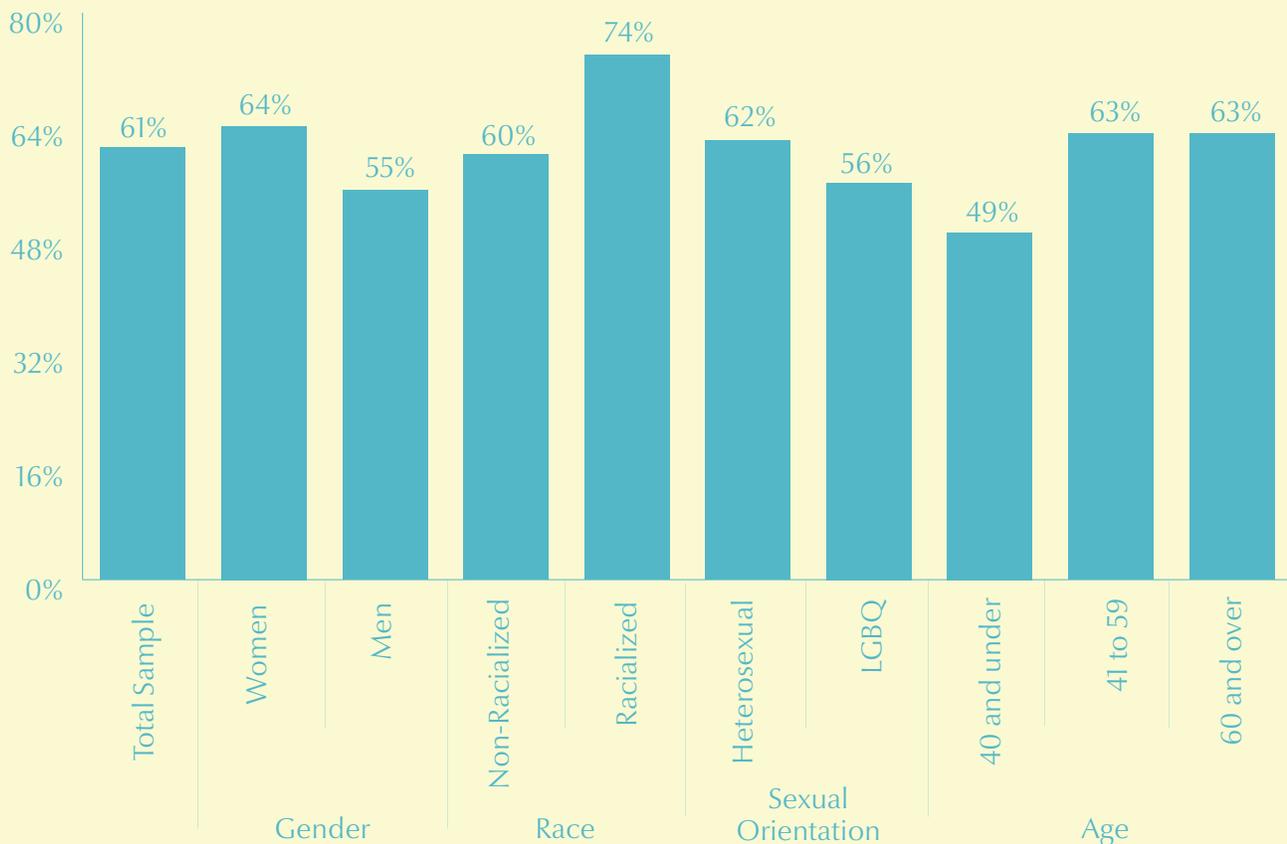
Figure 4: Anxiety during evaluation time due to negative feedback



Prior scholarship suggests that the negative emotional impact of SETs comments cause a sizeable minority of faculty, 32.6%, to wish they could leave teaching (NTEU 2018). We asked our respondents if they had ever considered leaving academia due to negative SETs comments. Eighty-four percent indicated "no", and 16% "yes". The analyses for particular socio-demographic sub-groups indicated numbers too small to report.

Respondents were also asked if they have ever received qualitative comments in their SETs that they perceived as abusive, hostile or bullying. Figure 5 shows that the majority of respondents in the survey indicated that they had received such comments. Noticeably, 64% of women and 74% of racialized faculty reported having received comments on their SETs that they viewed as abusive, hostile or bullying, in comparison to 55% of men and 60% of non-racialized respondents. LGBTQ respondents and those age 40 and under less frequently indicated that they have received such comments. These patterns accord with the findings of the large Australian faculty survey (NTEU 2018).

Figure 5: Have received comments that were perceived as abusive, hostile or bullying



Of those 61% of respondents who indicated that they received qualitative comments on their SETs that they perceived as hostile, abusive, or bullying, a little over a quarter (27%) indicated that they have received comments that have made them feel unsafe. This varied strikingly by race, with 44% of racialized respondents, and 23% of non-racialized respondents, being made to feel unsafe. Analyses by other demographic sub-groups indicated numbers too small to report.

Criteria that Students Apply to SETs Comments

Prior studies provide reasons to expect that non-pedagogical criteria figure heavily in students' evaluations of their professors, and that this is strongly affected by gender and race (Freishtat, 2016; Mitchell and Martin 2018; Stark et al 2016; Storage et al 2016). Our respondents were asked a set of questions that focused on the general utility of SETs comments for pedagogy, and the impact the comments have had on their teaching practices. Survey respondents were asked to indicate agreement with a number of statements about the criteria derived from prior studies, that SETs comments reflect. Figure 6 shows the proportion of those who agreed or strongly agreed with the statements. It is evident that there is a high level of agreement on many of the statements, which indicate that the majority of respondents believe that qualitative comments on SETs are based on factors other than pedagogy, most of which are ascribed characteristics of the instructors themselves.

Figure 6: Qualitative comments on SETs reflect ...

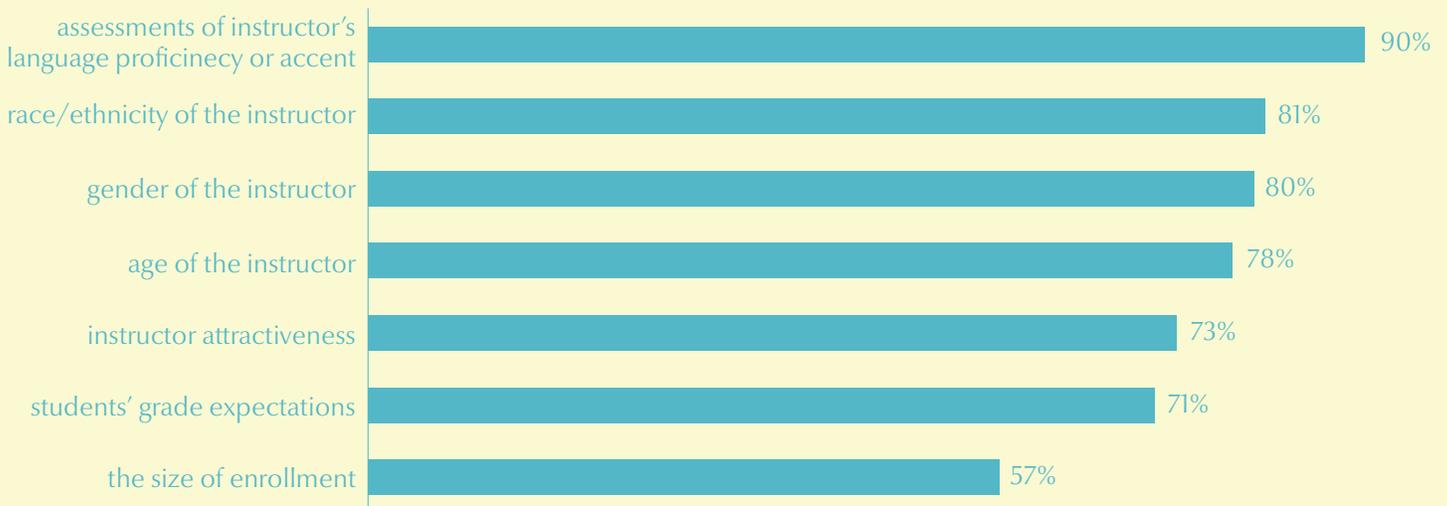
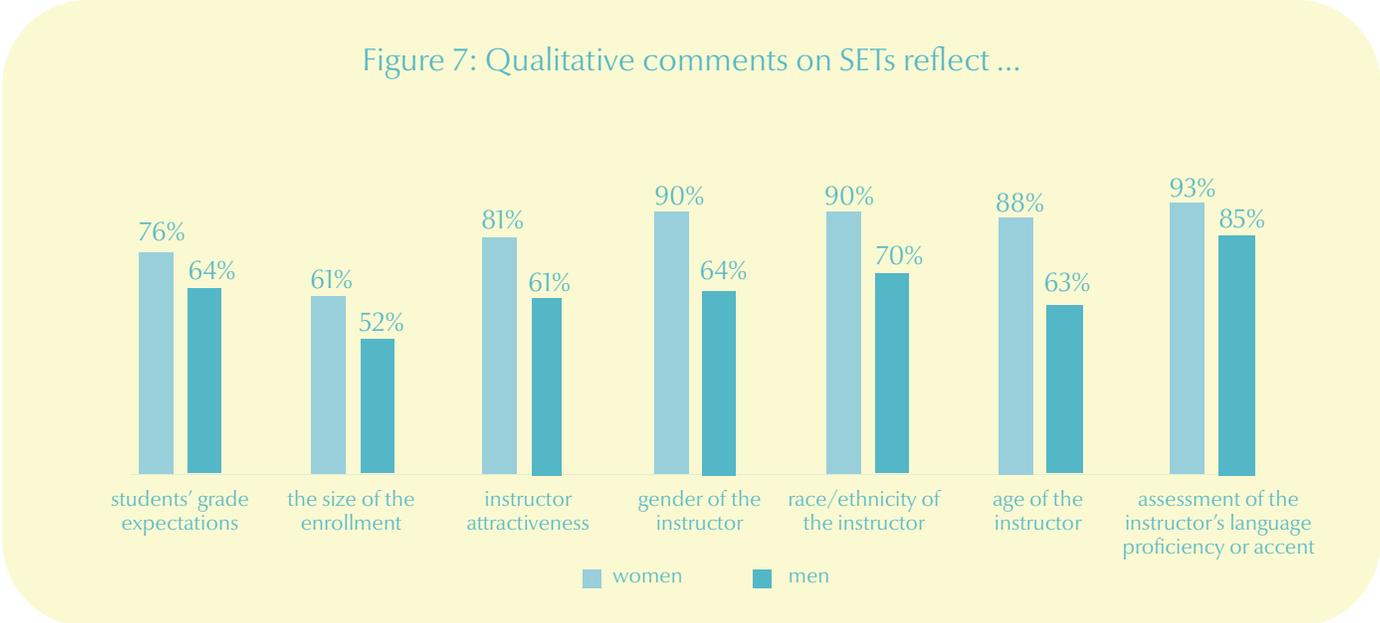


Figure 7 shows how this breaks down by gender. Female respondents are much more likely than their male counterparts to agree or strongly agree that qualitative comments are based on factors other than pedagogy.



In Figure 8, we also looked at how this broke down by sexual orientation. Similar to the findings on gender, LGBTQ respondents in the survey are much more likely than heterosexual respondents to agree or strongly agree that qualitative comments on SETs are based on criteria other than their actual pedagogy.

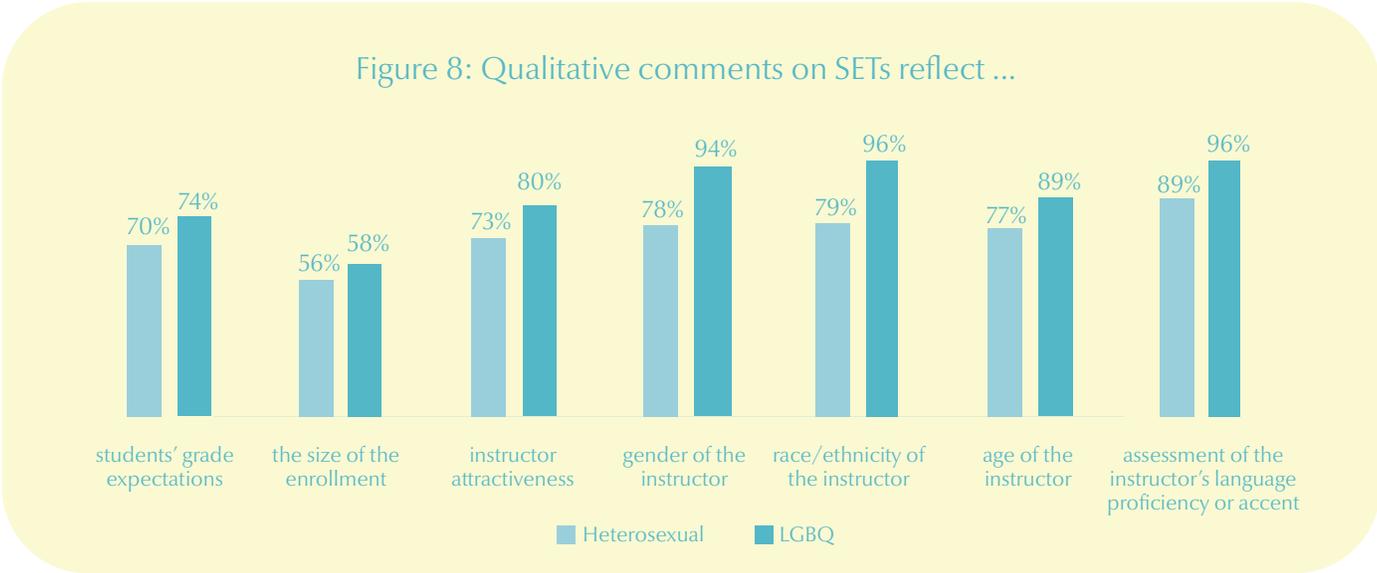


Figure 9 shows the impact of professors' age on their experience of non-pedagogical criteria in their SETs comments. For the most part, younger respondents in the sample are much more likely than older respondents to report that qualitative comments on SETs are based on factors other than pedagogy. Differences between racialized and non-racialized persons for this question were not meaningful.

Figure 9: Qualitative comments on SETs reflect ...

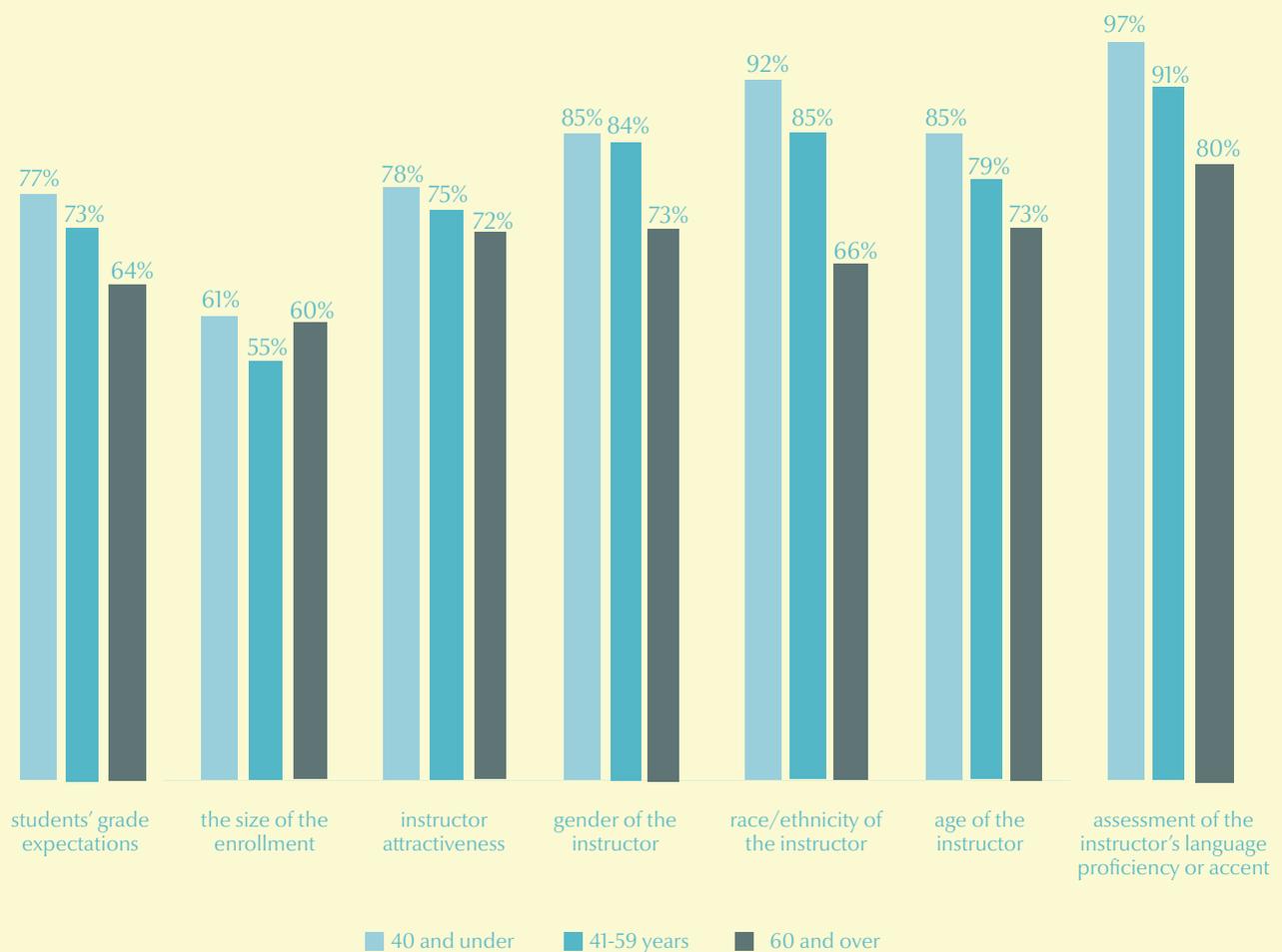
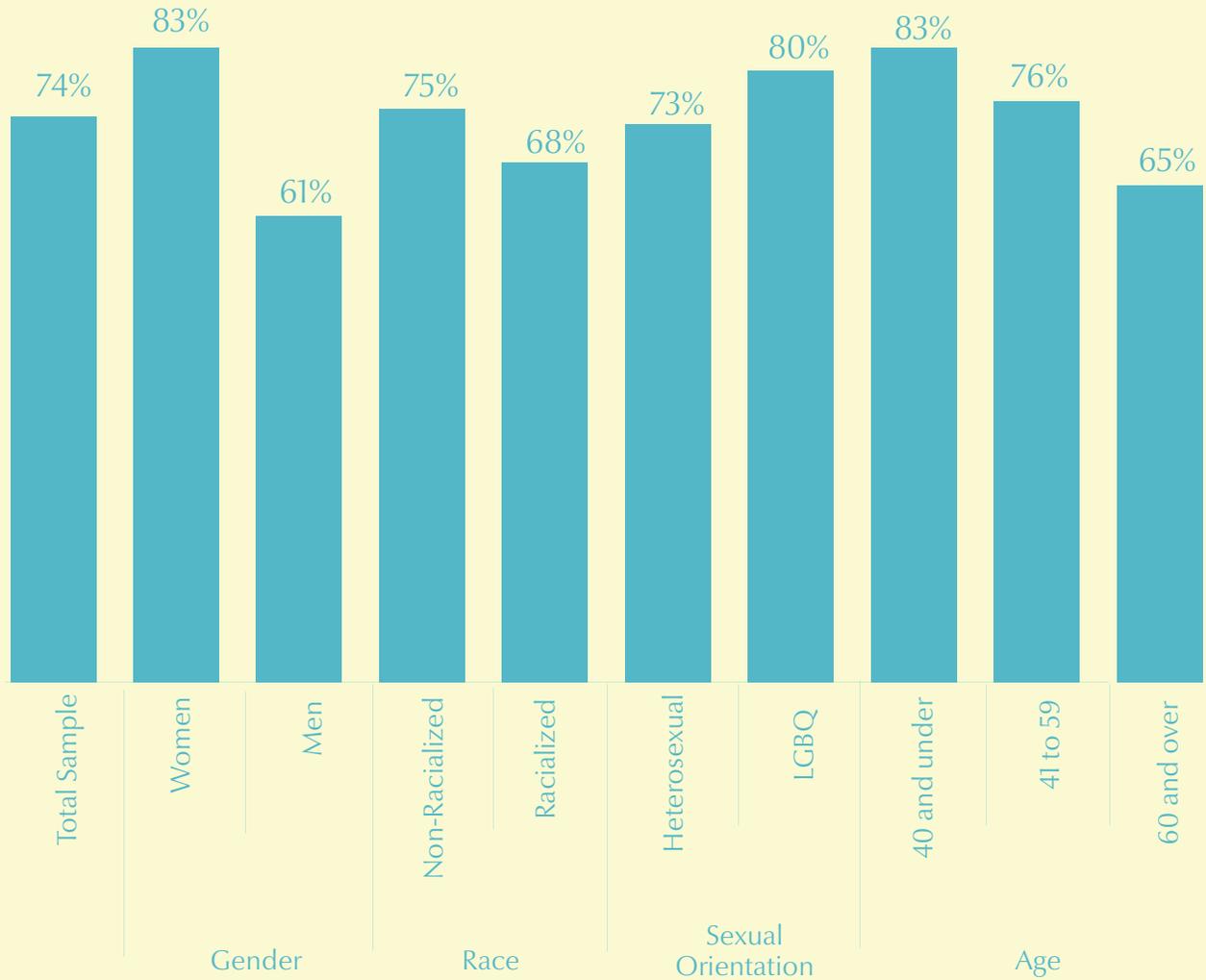


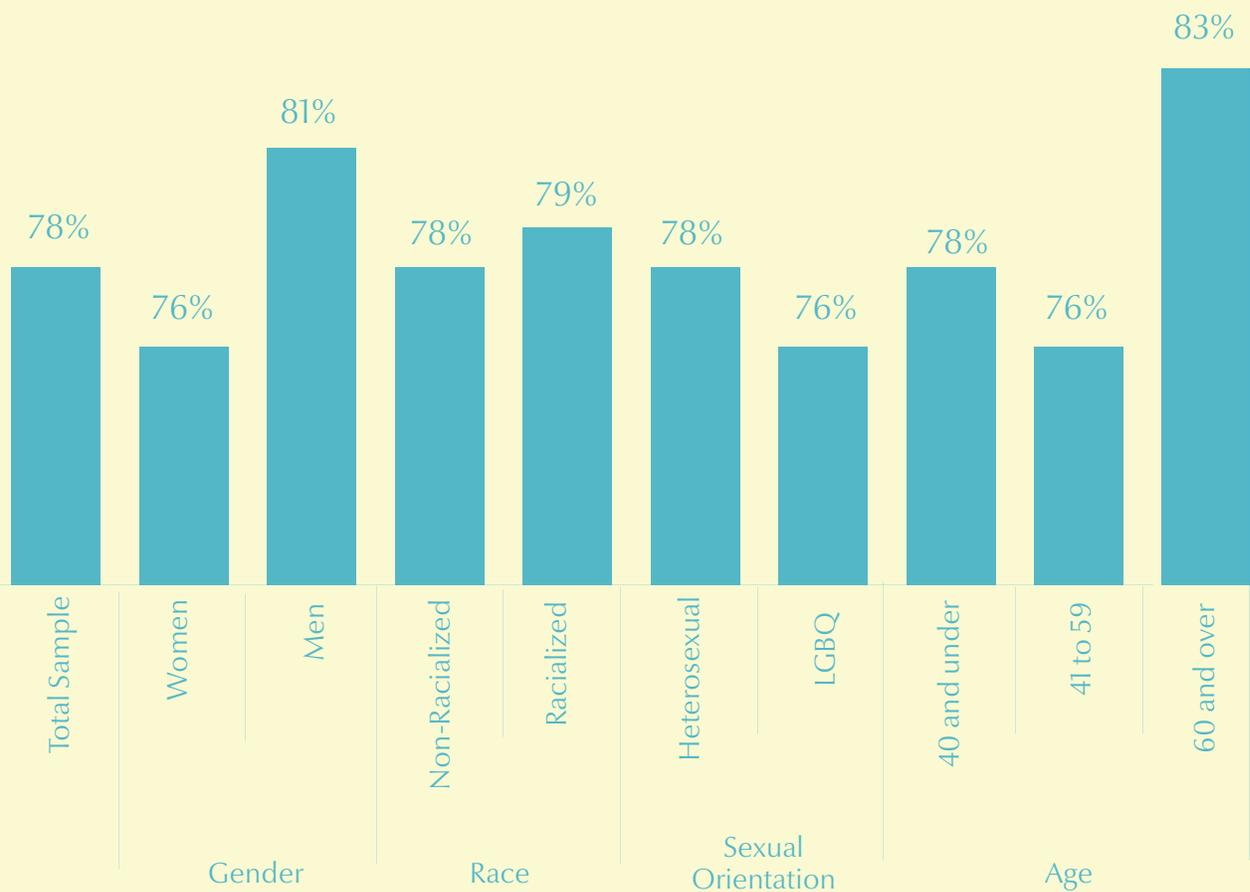
Figure 10: Often or always receives comments regarding how caring, empathetic, sympathetic you are towards students



Respondents were asked to indicate the frequency with which they received positive comments regarding how caring, empathetic, or sympathetic they are towards students. Figure 10 shows that the majority of survey respondents often or always receive positive comments on those criteria. Not surprisingly in light of previous scholarship (Mitchel and Martin 2018; Storage et al 2016), 83% of female respondents received such comments in comparison to 61% of males. Racialized respondents received such positive comments less often than non-racialized, similar to older respondents compared to their younger peers. LGBQ received such comments more often than heterosexual respondents.

Respondents were also asked to indicate the frequency with which they received positive comments regarding how knowledgeable they are on the subject matter they teach. Figure 11 shows the proportion of respondents who often or always receive such positive comments. Again, it is evident that the majority of survey respondents reported they often or always receive positive comments on the knowledgeability criterion. But in this case, 81% of male instructors and professors reported receiving such comments in comparison to 76% of female instructors and professors. While the differences for race and sexual orientation were not meaningful, we do see a difference by age: those instructors who were 60 years and older more often received positive comments on their knowledgeability compared to their younger counterparts.

Figure 11: Often or always receives positive comments regarding how knowledgeable you are about the subject matter being taught



Impact of SETs Comments on Pedagogy

Respondents were presented with a number of statements on the effect that qualitative comments have had on their pedagogical practices. It should be recalled that the vast majority of our sample report that they engage with, rather than ignore, their SETs comments. Figure 12 shows the proportion of respondents who agreed or strongly agreed with each statement. Only 32% of the sample see SETs comments as an effective means of determining teaching effectiveness. Over one third (38%) agreed that SETs comments have resulted in a lowering of academic standards, echoing empirically supported argument that SETs ratings have a perverse effect on learning (Stroebe 2016). However, almost two-thirds find qualitative comments helpful in improving course design and also the way they conduct lectures or seminars (62% and 61% respectively). We will have more to say on these apparent contradictions below.

Figure 12: Level of agreement on pedagogical practices as a result of qualitative comments on SETs

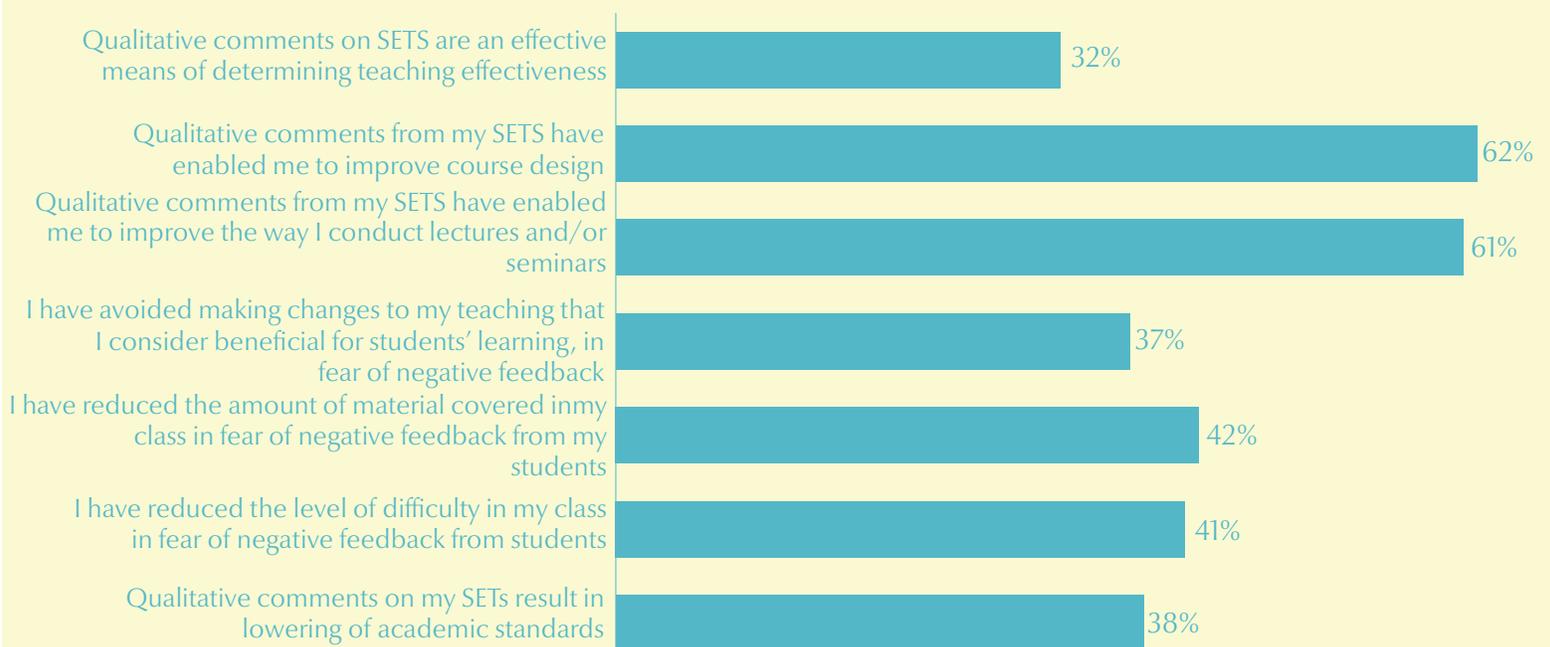


Figure 13 shows how these views break down by gender. More male instructors consider SETs comments an effective means of determining teaching effectiveness than female instructors. Likewise, more men find SETs comments helpful for improving course design, lectures, and seminars, than women. Conversely, more female instructors report having modified their pedagogical practices out of fear of negative feedback. Moreover, almost half of female respondents (45%) believe that SETs comments lower academic standards, compared to 29% of male respondents.

Figure 13: Level of agreement on pedagogical practices as a result of qualitative comments on SETs by gender

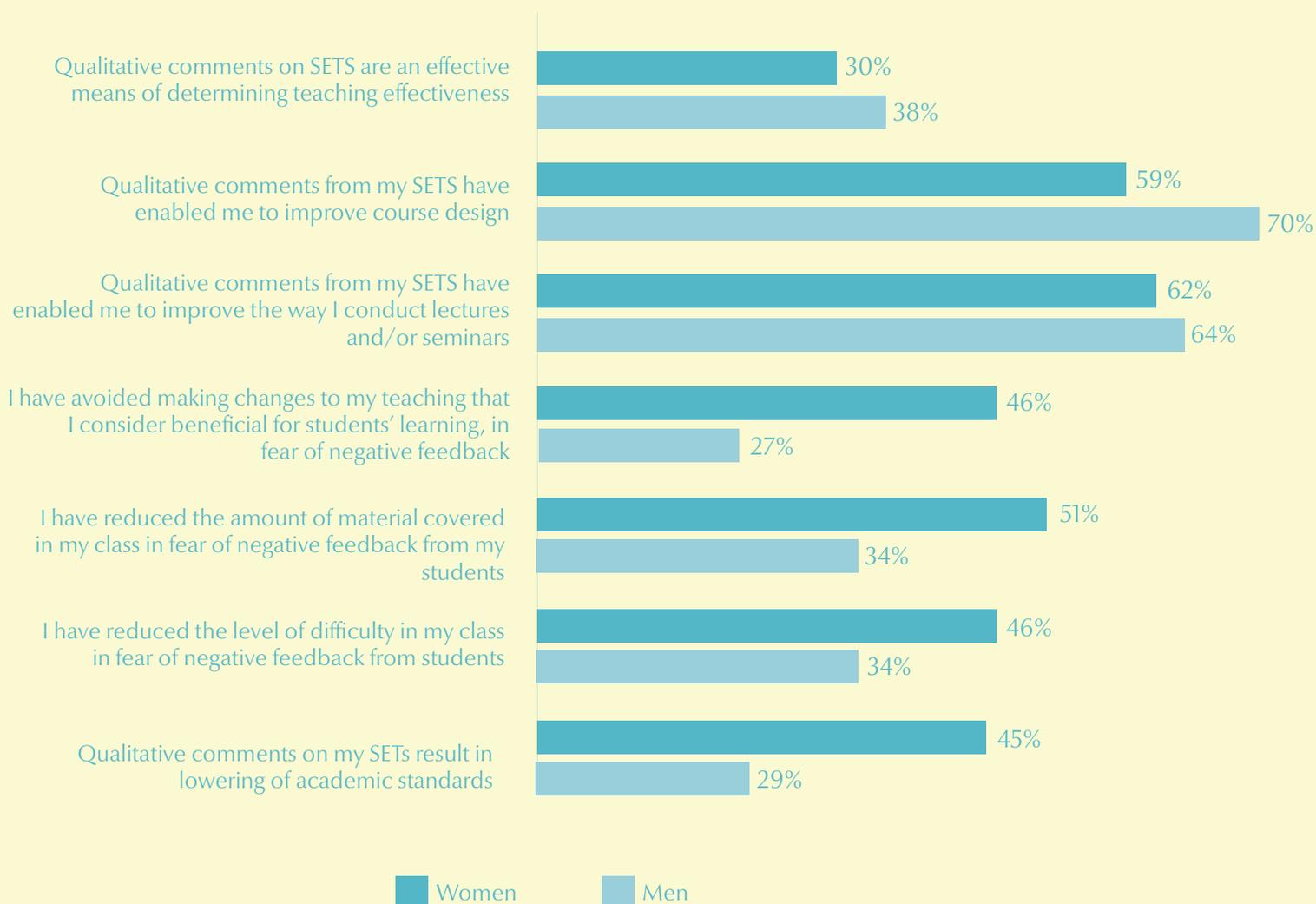
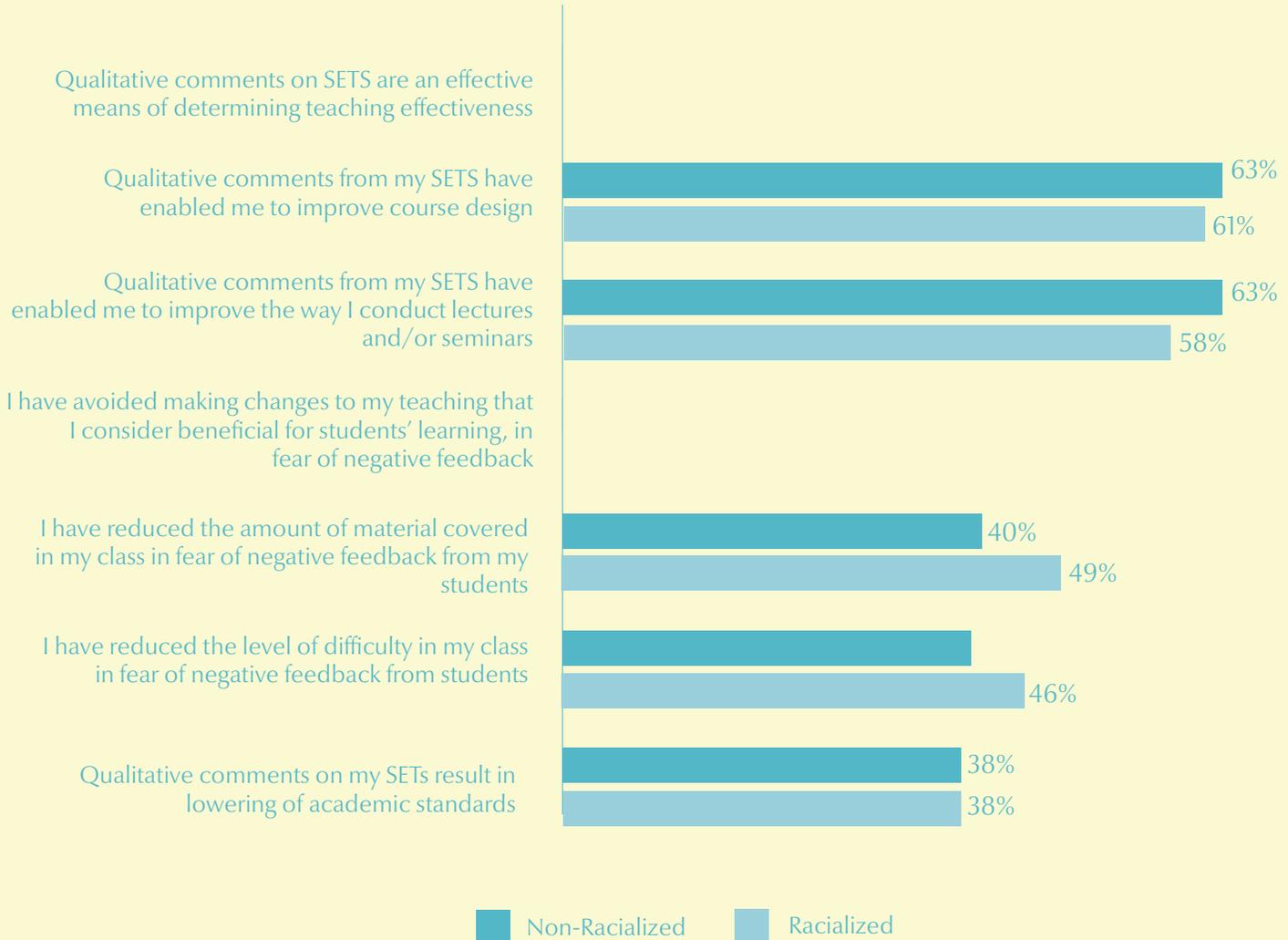


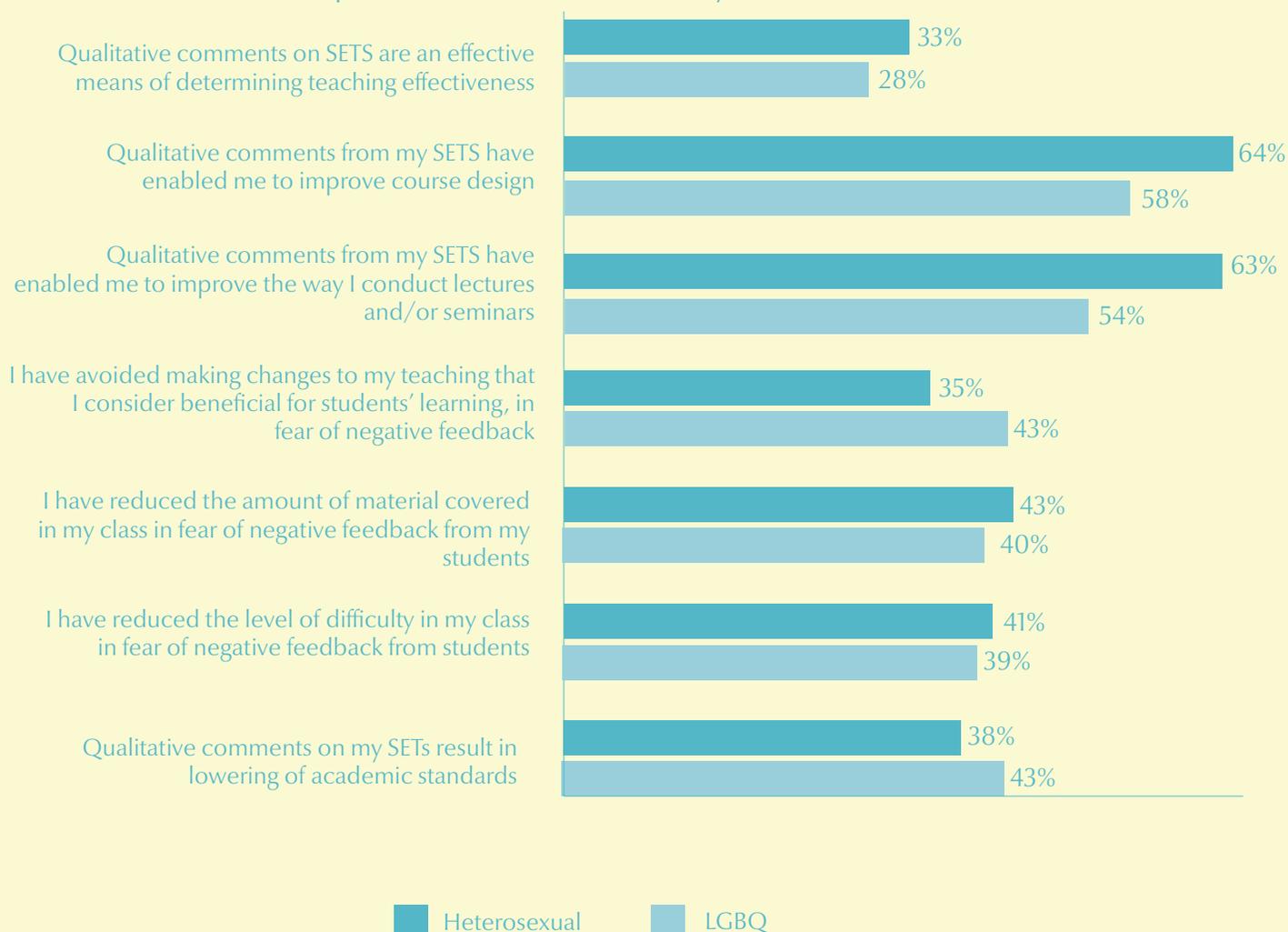
Figure 14 below presents differences between racialized and non-racialized instructors on how SETs comments affect pedagogy. Responses where n<10 were not reported. Fewer racialized respondents than non-racialized find SETs comments helpful for improving course design, lectures, and seminars. As well, a larger percentage of these faculty modified their pedagogical practices in fear of negative feedback. For example, 49% of racialized instructors agreed that they have reduced the amount of material covered in class in comparison to 40% of non-racialized instructors.

Figure 14: Level of agreement on pedagogical practices as a result of qualitative comments on SETs by race



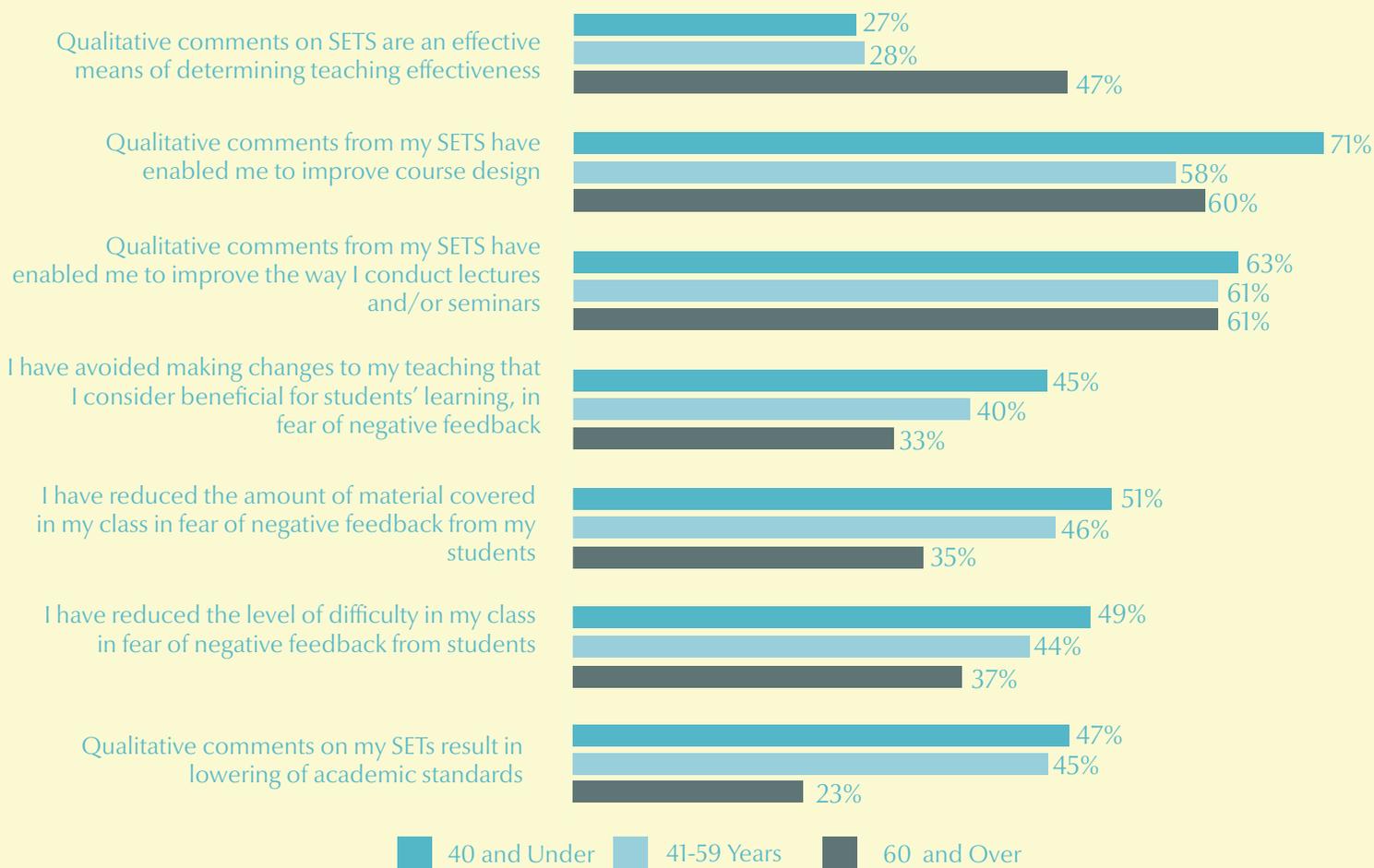
We also examine how SETs comments' effects on pedagogy breaks down by sexual orientation. Figure 15 shows that LGBTQ respondents less often agreed that SETs comments are an effective means of determining teaching effectiveness and that the comments have helped them improve their course design and the way they conduct lectures and seminars. LGBTQ respondents also more frequently agreed that qualitative comments have resulted in the lowering of academic standards.

Figure 15: Level of agreement on pedagogical practices as a result of qualitative comments on SETs by sexual orientation



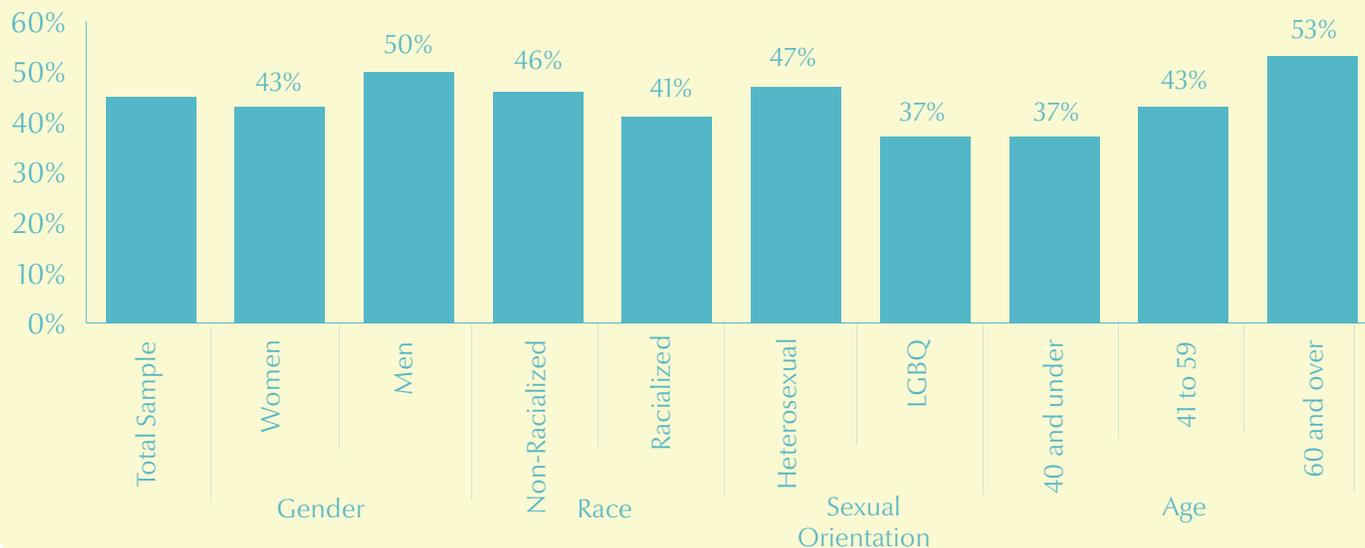
Finally, Figure 16 shows how instructors' age correlates with views on the impact of SETs comments on pedagogy. A greater proportion of younger faculty find SETs comments useful for improving course design, lectures, and seminars. But they less often modified their pedagogical practices in fear of negative feedback. Despite less frequently making use of SETs comments to improve pedagogy, a larger percentage of older faculty than younger faculty agreed with the statement that SETs comments are an effective means of determining teaching effectiveness. Again, we will comment on these contradictory findings below.

Figure 16: Level of agreement on pedagogical practices as a result of qualitative comments on SETs by age



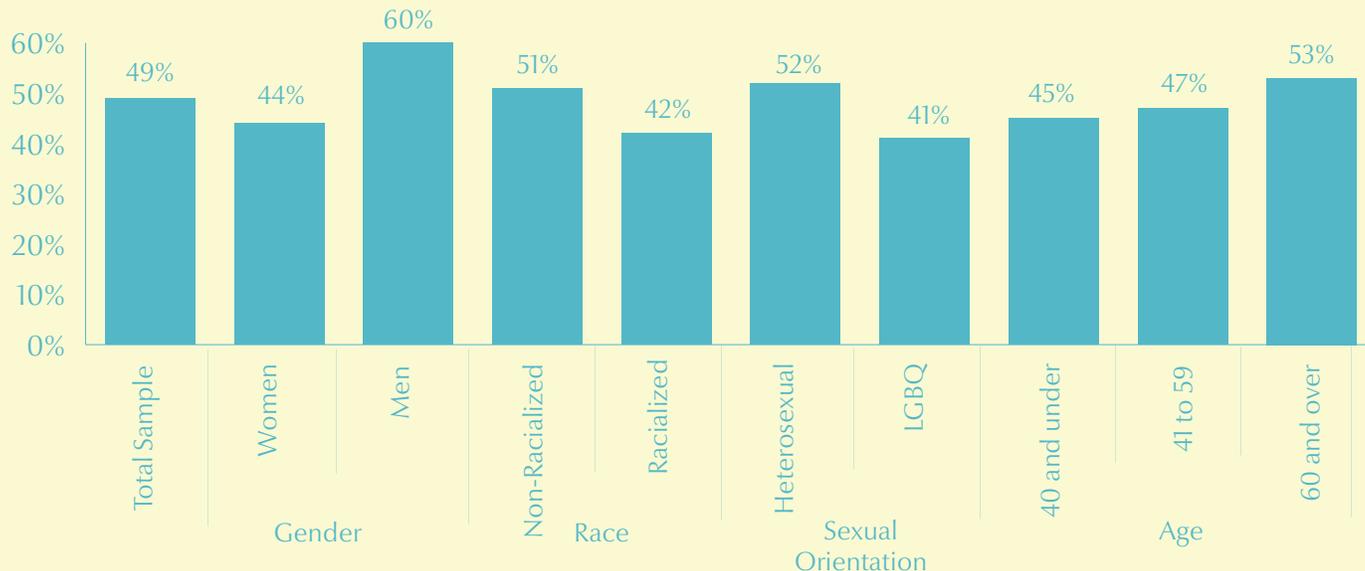
Respondents were asked whether SETs comments should remain part of the evaluation of teaching. Figure 17 shows that fewer than half (45%) agreed that they should. The proportions are lower for female instructors (43% compared to 50% of male instructors), and for racialized instructors. It is also lower for LGBTQ instructors (37%) than their heterosexual counterparts (47%), and lower for instructors 40 and under.

Figure 17: Qualitative comments on SETs should be part of the evaluation of teaching effectiveness



Respondents were also asked whether SETs comments have helped them become a better instructor. Figure 18 shows the proportion of respondents who agreed or strongly agreed with the statement. Again, just under half of the respondents in the survey (49%) agreed, but with female instructors less commonly agreeing than male instructors (44% versus 60%). Racialized and LGBQ instructors were also less in agreement with this statement. With respect to age, younger faculty agreed less than older faculty with the statement.

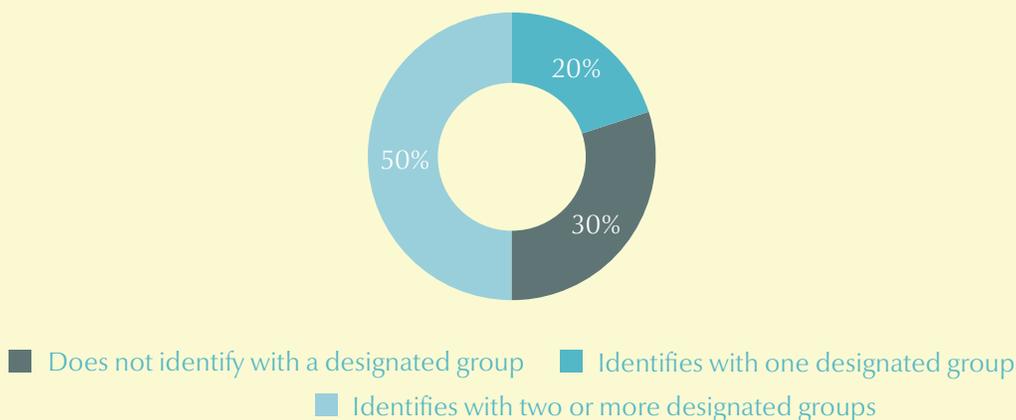
Figure 18: Qualitative comments on SETs have helped me become a better instructor



Intersectional Effects of SETs Comments

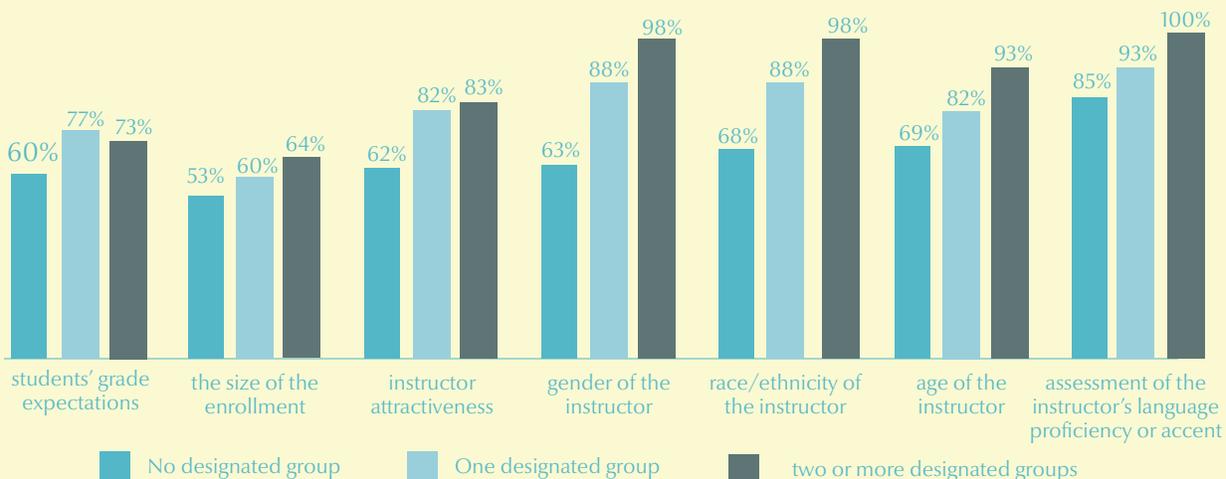
Due to the limited size of our sample, we could not analyze how specific intersectional identities correlate with respondents' views and experiences of SETs comments. Instead, we looked at whether a respondent belonged to one or more equity-seeking groups (gender, race, sexual orientation) as a proxy for intersectionality. Figure 19 shows that 30% of the overall sample consisted of white, heterosexual male respondents while just over half of the sample belonged to one equity-seeking group and 20% of the sample belonged to two or more equity-seeking groups.

Figure 19: Number of intersections across designated groups



In Figure 20 it is evident that respondents who belonged to two or more designated equity-seeking groups were more likely than their counterparts to agree or strongly agree that qualitative comments on SETs are based on factors other than pedagogy, most of which are ascribed characteristics of the instructors themselves. This is especially evident when it comes to statements about SETs comments being based on personal characteristics such as gender, race, age and language proficiency.

Figure 20: Qualitative comments on SETs reflect ...



We also examined intersectionality in the effects of SETs comments on sense of well-being. Respondents who identified with two or more designated groups were much more likely (41%) to indicate that qualitative comments have had a negative impact on their sense of well-being in comparison to those who identified with one designated equity-seeking group (38%) and those who identified with none (21%). Similarly, when looking at whether instructors feel anxiety during evaluation time due to fear of negative feedback, 67% of those who identified with two or more equity-seeking groups answered "yes" in comparison to 65% of those who identified with one designated group and 46% that did not identify as belonging to any equity seeking group. Over a quarter (26%) of those who identified with two or more equity-seeking groups indicated that they have considered leaving academia. The number of those who did not identify with a designated equity-seeking group was too small to report the percentage. When asked whether respondents have received SETs comments that they perceive as abusive, hostile or bullying, there were no meaningful differences between those who identified with one and those who identified with two or more equity-seeking groups. However, these two groups more commonly reported receiving comments that were abusive, hostile or bullying when compared to those individuals who did not identify with any equity-seeking group. Moreover, it should be noted that 40% of those who identified with two or more designated equity-seeking groups who indicated they had received qualitative comments that were abusive, hostile or bullying indicated that they the comments made them feel unsafe. The number of those who did not identify with a designated equity seeking group was too small to report the percentage.

Contingent Faculty and SETs Comments

While many of the differences between contingent versus tenured or tenure-track faculty were small, there are some areas that were worth noting. When it came to the questions on how SETs comments affect their sense of well-being, we found that contingent faculty more often reported that they have considered leaving academia due to negative comments on their SETs (33% versus 13% of tenured/tenure track). Moreover, 74% of the contingent instructors reported having received SETs comments that they perceived as hostile, abusive or bullying in comparison to 58% of tenured/tenure-track. Finally of those reporting that they received hostile, abusive or bullying comments, 41% of the contingent faculty reported feeling unsafe as a result of receiving such comments in comparison to 24% of tenured/tenure-track.

“

Respondents who identified with two or more designated groups more often 41% indicated that qualitative comments have had a negative impact on their sense of well-being in comparison to those who identified with one designated equity-seeking group (41% vs. 38%) and those who identified with none (21%). A similar pattern is seen in instructors' experience of anxiety during evaluation time.

”

Qualitative Data Analysis

Our survey included three open-ended questions, each delineating a dimension of experience with or reflection on qualitative feedback from SETs. The results show respondents reflecting on and contextualizing their experiences with these comments. They describe student feedback as multifaceted accounts, with commentary on everything from student's academic development, to the instructor's intellect, physical appearance, teaching ability, course design, integrity, organization, and support. In summarizing the most memorable/impactful comments from their teaching careers, respondents highlight **four major SETs considerations** – pedagogy, personality, personhood, and perceived identity.

Pedagogy

Instructors routinely receive comments about their pedagogy, including style and innovations in content delivery and teaching effectiveness. The most favourable are students' appreciation of feedback on assignments, individual check-ins, sustained classroom engagement, passionate content delivery, and innovative course design. The least favourable recount classes (or instructors themselves) as boring, incomprehensible, repetitious, worthless, and biased. However, providing descriptive nuance to our quantitative findings, respondents most frequently recall a combination of strikingly favourable and critically disapproving feedback – across the teaching career, in individual courses, about specific pedagogic tools, and at times, from individual students – what we call, mixed sentiments. Alongside these highs and lows are the emotional rollercoaster respondents describe as they attempt to make sense of the qualitative feedback.

Of the memorable comments respondents received in SETs, the most common were mixed sentiments about course pedagogy – i.e. polarizing favourable and disapproving feedback about teaching style, effectiveness, pedagogic tools and the like.

In most cases, respondents struggled to make sense of this feedback; at times revealing differences of opinion between students, or incongruencies between students' and instructor's assessments of pedagogic design. Showcasing the former, Carl (54), sees mixed sentiments as an issue with SETs as a whole; "students who bother to write comments either love or hate my strategies for delivering content and to engage them." It is noteworthy that this was a common sentiment amongst our respondents. In fact, as Lily (46) points out, this likely reflects the self-selection of students who are most motivated to provide feedback on SETs – a concern exacerbated by their transition online,

Most evaluations have been positive, but sometimes the negative ones outshine the positive. I find that in the qualitative portion of SETs, the students who bother to type something in that section are either the students who are really happy with the instructor/course or really upset with the instructor/course. The students who fall in the middle often don't comment and it would be nice to read what the average student thinks as well.

For Pat (39), mixed sentiments reflected students' incongruent assessments of her pedagogic design:

Overall, I have gotten overwhelmingly positive feedback from students who consistently refer to my ability to explain difficult ideas in simple terms, my organization, and enthusiasm for my subject matter. However, I recently taught an experiential-learning based course, which I put a ton of effort into planning and organizing. Student feedback was baffling: the course was designed around a lot of in-class workshops that gave students time to work on their major research projects, but they reported that I simply 'read to' them for three hours and didn't give them enough time to work on their projects in class as I'd promised.

Like many of our respondents, Pat receives quite positive student feedback on the whole, but finds it difficult to reconcile comments in direct conflict with the design of pedagogic tools and her own experience delivering them. Given the vast majority of respondents read their qualitative comments (92%), and use them to improve course design (62%) and/or lectures/seminars (61%), it is noteworthy that this challenging feedback was so common. However, we note greater import than the frequency of mixed sentiments. Our data point to an emotional toll that respondents reflect in their recollections. For example, Pat (39) describes coming to terms with the feedback from her experiential-learning course, “[it] bothered me a lot because of how much care I’d put into ensuring that they got time in class to work with me and their peers.”

In many ways, students' mixed sentiments are a source of disappointment when instructors thoroughly invest

in course planning, and even more so, when students and instructors disagree on the pedagogic value of these investments. However, some respondents' noted outright inconsistencies between their recollections of pedagogy and students' SETs feedback; these garnered more emotionally-laden reflections. As Carla (52) recollects, “I regularly have students lie -- saying that I only read verbatim from the slides, that I don't give them instructions.”

In this manner, our data illuminate an important consideration; while qualitative feedback from SETs are a summative exercise for students, they are formative for instructors' teaching. Most pronounced for contingent and specialized faculty groups – e.g. untenured, teaching-stream, and graduate students – instructors are expected to engage, utilize and justify SETs feedback in light of their teaching practice. Yet, the extent to which feedback is truly formative depends on its validity, reliability and overall quality. Mixed sentiments of such polarizing differentiation make that difficult to ascertain. More strikingly though, is the potential risk in doing so. As over one-third (37%) of our respondents communicate, there is perceived risk in making changes to course pedagogy due to fear of backlash in SETs comments (Figure 12). And as our Figures 13 through 16 show, these perceived risks may come at a great cost to women, racialized, LGBQ¹⁰ and younger faculty.¹¹ Shedding further light, some respondents commented on the careful negotiation and surveillance of course pedagogy in light of course evaluations. As Danni (57) recounts:

Others felt that I couldn't wait for the class to end given that I would watch the time. (Actually, I was

⁹ Emphasis is the respondents.

¹⁰ Findings for the LGBQ respondents are mixed, but they more frequently avoided innovation to pre-empt negative feedback, and more commonly consider SETs comments a contributor to the erosion of standards.

¹¹ While our quantitative data are limited for racial and ethnic minority faculty, we expect this risk to similarly apply based on the qualitative data (more on this in the Personality, Personhood and Perceived Identity sections).

watching the time to make sure that they [sic] class ended on time, Now, to ensure that they don't see me watching the time, I bought a large faced watch and when I look down at my notes, I take an unknowing glance at my watch. Other times, when watching a video, one claimed that I was sleeping. So, now I make sure that I quickly blink and move around in my chair. Other comments indicate that they don't like the amount of material that they have to cover. (Little do they know, that I have cut numerous chapters). Others don't like the exam structure, even when I have resorted to giving them the questions before the exam. I could go on and on with many more examples...

While itself a cause for concern, mixed sentiments about pedagogy jeopardize more than the potential for new and innovative teaching enterprise. They also jeopardize the upholding of academic standards (38%), the volume of content coverage (42%) and the rigor of teaching practice (41%).¹² As Danni's reflections demonstrate, they may motivate overzealous performativity to compensate for, or pre-emptively avoid, penalties on SETs feedback; an emotional burden borne disproportionately by women, racial/ethnic minorities and younger faculty. We will say more on this in the sections that follow.

Highlighting the emotional import of SETs comments, respondents demonstrate marked consistency in how they make sense of mixed sentiments.

In nearly all cases, positive feedback is simply listed, superficially examined or regarded as a "fast ego boost" (Natasha, 57) – largely due to feedback that is affirming, but lacking pedagogic depth. Conversely, respondents routinely provide context or analysis to justify or explain negative comments. We make particular

note of this pattern considering our survey question asked respondents about their most memorable/ impactful SETs comments, but nothing more. As Arthur¹³ recalls, "I have received many comments to the effect that I have a broad and comprehensive knowledge of the subject, which is gratifying. Some students from time to time dislike the critical approach I take and feel it is "biased", despite my efforts to present both sides of issues in many cases." Or as Michaela (68) reflects,

Ones that tell me I am engaging, caring, have given a new perspective on social issues, knowledgeable and fair (I get these 2 all the time), challenging. Ones that say I am boring are more frequent than I would like, unfortunately. [The] amount of work expected can be an issue - though I can teach the same class on 2 different occasions and get quite different comments. [...] It often feels that the student evaluations are more of a popularity contest than a genuine helpful evaluation, though some students do indeed write very helpful comments about strengths and weaknesses of the course and of the instructor.

By most accounts, the net effect of students' mixed sentiments is, at best, pedagogically dubious in the formative sense. That notwithstanding, for select respondents, their most memorable SETs comments reflect students' careful and critical consideration of teaching pedagogy:

Overall[,] I receive very positive comments. I put huge amounts of time and effort into my teaching/grading and I do find the positive response validating. I particularly appreciate when students find the content useful and state that it changed their perspective for the better -Patricia, 35.

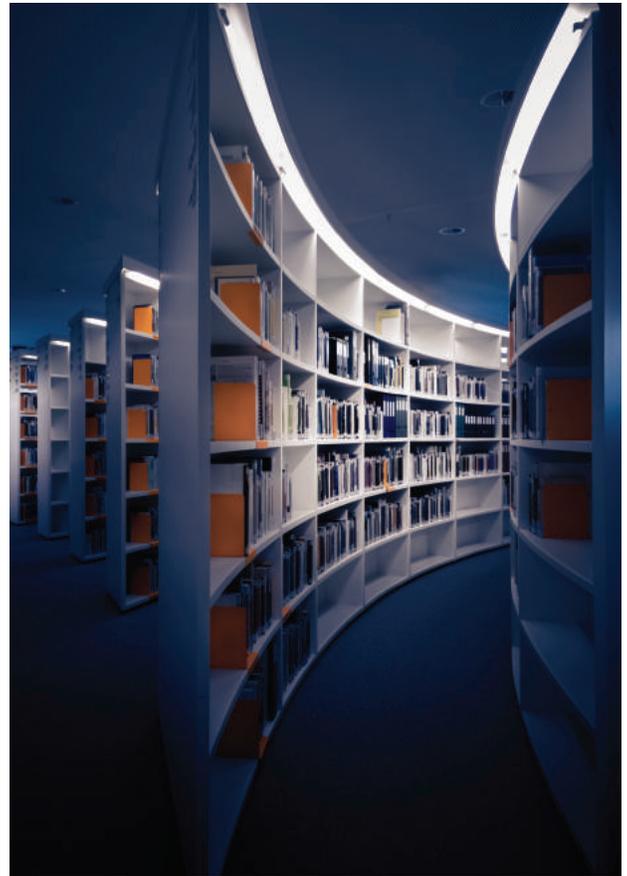
¹² As seen in Figure 12.

¹³ No age provided.

Well, I do feel good when students write: 'The best teacher I ever had' or especially comments like; 'This was a really tough course, but I learned so much and I see the world differently now', because they acknowledge that there is significant reading and writing -- but that they learn – Nila, 44.

In general, most instructors describe making sense of mixed sentiments regarding pedagogy with negative emotions, including frustration, confusion, and irritation. We note the parallel between these emotions and those most frequently experienced after reading qualitative comments from SETs (75% frustration and 72% disappointment; as seen in Figure 1). Others, like Hanna (43), were resigned, “I’ll say that my most common comments are[,] on the positive side[,] I’m really enthusiastic[,] and on the negative[,] that I can be disorganized. Both are fair assessments, I’d say.” While others still suggested there is little more you can do than simply laugh; as Kyla (48) recalls a student’s comment: “‘this course sucked much less than I expected it to’....still chuckle at that one....or ‘I stayed awake for every lecture so you definitely accomplished something.’”

But respondents’ most memorable SETs comments reveal nuance and emotional import beyond the mixed pedagogic assessments described above. For example, numerous instructors reported attacks on their personality, personhood and perceived identity in SETs feedback. In the sections that follow, we more closely examine these examples.



“

In nearly all cases, positive feedback is simply listed, superficially examined or regarded as a “fast ego boost”, largely due to feedback that is affirming, but lacking pedagogic depth. Conversely, respondents routinely provide context or analysis to justify or explain negative comments.

”

Personality, Personhood and Perceived Identity

Many of our respondents also indicate reading and understanding SETs comments in light of, or in spite of, the social bias evident within the comments. Startling, however, was the frequency with which these biases appeared amongst respondents' most memorable comments. We note a consistent pattern of memorable comments from students relating to instructor's personality, personhood and perceived identity – collectively termed, non-pedagogic factors. The patterns we observe here are threefold: 1) visible minority groups – including women, people of colour, and ELL/ESL¹⁴ instructors – frequently recall student feedback about or seemingly influenced by non-pedagogic factors; 2) irrespective of socio-demographic background, respondents broadly acknowledge and are critical of social biases in SETs (both in their own and in fellow colleagues' evaluations); and 3) relative to comments about pedagogy, respondents more clearly articulate the emotional toll they experience from assessments about personality, personhood and perceived identity.

The consistency with which these comments are recalled and explained, and the depth and breadth of their content, underscore their emotional import. Further, comments regarding personality, personhood, and perceived identity are difficult to disentangle from content about pedagogy. Below, we note unique patterns in the way comments about these non-pedagogic elements are recalled and understood by respondents.

i. Personality

The most common reflections on personality were those describing instructors as caring, enthusiastic, passionate, and engaged. Some were tangentially related to pedagogy – for example, students' lauding "enthusiasm for the material being taught" (Robin¹⁵) or noting a teaching style of "authentic engagement, knowledge, and caring" (Carmen, 38). But a greater proportion were distanced from any relation to course pedagogy. For example, an inordinate number of memorable comments highlighted how nice instructors were (Anita, 44; Abigail, 53; Parker, 54; Chris, 66 and more), or how they were the best or favourite instructor students had come across in their undergraduate careers (Delores, 63; Nila, 44; Kim, 30; Aaron, 39; Jordan, 45; Bill, 49 and more). These were laden with gender bias – for example, women

were most commonly regarded as nice; a highly gendered metric of willingness to perform emotional labour (Smele et al 2020, forthcoming). Respondents of all genders received feedback about being their students' best or favourite professor; but only one man was regarded as brilliant – Nick (42). While disturbing, it is an altogether consistent finding. The more startling pattern is how acutely respondents verbalized how they (or their colleagues) were/are (dis)advantaged because of these biases.

As many respondents also point out, SETs are a popularity contest, reflecting socially biased perceptions and expectations of teaching personality, at times in lieu of critical feedback. As Natasha (57) delineates,

¹⁴ By visible minority group, we refer to respondents' personal identity with a visible minority group(s) or perceived identity (by students) by virtue of SETs commentary. In this way, respondents could be 'visible' by virtue of any socio-demographic factor students perceived and overtly identified in their course evaluations – e.g. accent.

¹⁵ Respondent did not provide her age.

Generally I find that 'satisfied' students do not offer valuable feedback. Students inspired by a course can give me a fast ego boost which also does not really offer much in the way of pedagogical growth for me. The disgruntled students either don't show up to do the evaluation in class, or make general comments that legitimate their frustration. So this also doesn't really provide much in the way of pedagogical growth for me.

In many ways these experiences illuminate data from Figure 12 providing a glimpse of the types of student comments that faculty consider in trying to improve teaching.

As data from Figure 10 supports, student evaluations reward instructors who adhere to stereotypical representations of expertise and emotional labour. This throws further into question the effectiveness and utility of SETs, and not just with regard to gender.¹⁷ Respondents from a broad spectrum of socio-demographic backgrounds grapple with whether SETs can meaningfully distinguish between evaluations of personality and effective teaching. As Sean (57) deliberates, “I am a highly popular, engaging, fun, thoughtful, empathetic, insightful even entertaining prof --is that enough?? are they getting it or am a [sic] just a good performer. Now being a good performer is essential and teaching is performance but is it enough. It could be worse I suppose. So the evaluations tell me all that good stuff...but am I actually a good teacher!!!”

Relatedly, it is also nearly impossible to clarify the context within which SETs comments (including those about personality) are communicated. In some cases, the institution, department, classroom

facility or course design provide the necessary context to make sense of student feedback. For Danni (57), substantive content of the course is of particular relevance: “They make up stories, they lie, and they are just down right nasty. It was worse when I was teaching courses that they didn't like, such as statistics and theory. Even when I was knowledgeable, one student felt that since I knew so much, that it was time for me to retire.” In this case, course content (statistics and theory), and its intersection with gender, turn traits such as knowledgeability into detriments. For Claudette (70), it is a mix of institutional factors – abysmal response rates for online SETs¹⁷ and identity (her age)¹⁸ – that provide the context for her most memorable comments. She is, “too old, disrespects students, lies, [and] comes to class late” – Claudette (70).

An acute awareness of social bias in SETs comments about personality does not exempt instructors from their emotional toll. As Figure 4 demonstrates, marginalized faculty more frequently report experiencing anxiety during the SETs administration period – most notably women (68%), racial/ethnic minorities (61%), and sexual minorities (63%). Pat (39) and Nicole (50) reflect below:

To be clear about how anxious I get when I get my SETs, I usually start physically shaking when I see the email land in my inbox saying that evaluations are in. I feel like I'm about to get ambushed by a barrage of negative evaluations. If I see negative feedback, particularly that which seems unfair, it occupies my thoughts for days
– Pat

¹⁷ As Claudette notes, the response rate at her university is “usually less than 10% of the class.”

¹⁸ More information on patterns related to socio-demographic factors in the section on identity below.

I am not the kind of professor who needs to be liked by students, but if I were, I think this would be extremely detrimental to my mental well-being. Teaching is difficult enough without having to read student evals, which some students use as an opportunity to lash out at professors. I have been subject to sexist remarks, and it both irks and bewilders me that students are not made accountable for such remarks. I have colleagues who have reported racist, homophobic, and fat-phobic comments. This is particularly stressful for graduate students who are teaching courses and who (think they) need extremely positive evals to get jobs – Nicole

The affective responses to assessments of personality are extensive – anxiety, sadness, hurt, confusion, annoyance, disappointment, resentment and the like – and while not all comments are inherently negative (as in the examples of brilliant Nick and the numerous nice faculty amidst our sample), few respondents detail long-lasting positive, emotional impact:

The term that I had some students complain about my asthma-related coughing, and some others complain about my apologies for coughing, was the term that my attention to SETs began to wane significantly. I find there is little of value--and much that has the potential to hurt and disappoint--that comes out of giving hundreds of people an anonymous platform to say anything they want about me with impunity. I feel this way despite the fact that, overall, I largely get positive reviews. The negative comments are the ones that stick with you. – Phil, 44

For Phil (44), Pat (39), Lily (46) and select others, negative comments are particularly pervasive, in the sense that they ‘stick’ and ‘occupy the mind for days’.

Nevertheless, only 34% of respondents reported a negative impact on their sense of well-being as a result of SETs comments (see Figure 3). In conjunction with Figure 4, this suggests that respondents distinguish between the acute anxiety surrounding SETs administration and long-standing psychological impact. While the empirical support for this interpretation is not within the scope of our data, these are important questions for future research.

Overall, respondents’ most memorable comments demonstrated numerous patterns regarding assessments of personality: Assessments are 1) at times, indistinguishable from genuine pedagogic feedback; 2) intertwined with feedback of genuine pedagogic value, when present; 3) reflections of gendered social biases, including expectations of emotional labour and assessments of content expertise; 4) recognizably biased to respondents themselves; and 5) routinely decontextualized from external factors necessary to make sense of them. All the while, instructors continue to (and are expected to) engage with SETs feedback, good or bad, incurring the emotional toll.

ii. Personhood

In the section above, we note numerous instances of cross-over between assessments of personality, personhood and identity. For example, gendered expectations of emotional labour bridge assessments of identity (gender) and personality (caring). Below, we explore examples where faculty reflect on student assessments of personhood – i.e. personal or professional character – and their various intersections with personality and identity. In doing so, we reveal nuanced insight regarding their emotional import; namely that these comments are exclusively negative and have the potential for greater prevalence amongst marginalized instructors' SETs. This experience may be unique to faculty in social science disciplines as it pertains to course content of close relevance (and arguably critique) to students' lives.

Overall, assessments of personhood typically took the form of assumptive liberties by students. For example, some respondents recalled students conflating their approach to substantive content – such as teaching critical race theory – with personal or professional politics. In doing so, students voiced their concerns – or in many cases discontent – about the appropriateness of pedagogic content and approaches, by means of negative feedback about faculty. Consider how teaching in the substantive area of race and ethnicity play out differently for two of our respondents. Lara (41), a Southeast Asian woman, is accused of reverse racism in her course evaluations. "In my large intro to race and ethnic relations course, students have actually commented that I was "racist against whites" because I teach topics such as white privilege, white fragility, racial

battle fatigue, settler colonialism, anti-Black and anti-Indigenous racism, etc." Whereas Tim (38), a white man, "got called racist because, despite care to emphasise the social inequities causing it, [...] ¹⁹ presented race/crime statistics in a 2nd year criminology class that showed some minority groups with elevated levels of offending."

In both cases, students' express discontent for the way substantive content is presented (or perhaps the content itself). But a more subtle, noteworthy mechanism is also at play. Rather than describe the content, department or discipline as racist (or reverse racist), students filter their understanding of the substantive content through the (perceived) identity of the instructor, to make assumptions about their personal or professional politics. In this way, students are, at first glance, evaluating pedagogy, but more overtly assessing instructors' identities and personhood. It is worthwhile to note that the line between feedback of pedagogic value and personal/professional attacks is easily blurred, because it is distinctly relative. In other words, many instructors would find it difficult to look past the acerbic and intimate nature of attacks on personhood (especially if frequent) for the potential "good learning experience" (Rick, 54) – as may be the case for faculty with intersecting marginalities. For faculty, these comments will range from innocuous to threatening, contingent upon a myriad of factors – including the frequency of their occurrence, the identity of the instructor, the rank and precarity of the teaching position, the content of the course etc. Nevertheless, the highly personal nature of these attacks likely explain why they were

¹⁹ Ellipses refers to the pronoun "I" which was removed for grammatic clarity.

memorable. In another example, a student reflects on the pedagogic choices of an instructor (rather than the content itself), to assess their personhood. Rick (54) recalls an instance at the beginning of his teaching career,

I invited a colleague to do a guest lecture in the first course I taught. One student wrote on the evaluation that "Inviting a Black friend to do a lecture for you doesn't show the class that you are not a racist". I often think back to that class for what it revealed to me about the race politics in the classroom. In retrospect, I think that it was the only way for the student to feel like they could express how they felt about me in the role of instructor for the class. At the time, I didn't know what to think. I was trying my best and I felt attacked, and like I failed. It was a good learning experience.

Taken together, student assessments of personhood reveal four central takeaways: 1) that students comment on assessments of personhood in deliberate and motivated ways when they perceive an overlap between the instructor's identity and the substantive content; 2) even when content is not concerned, students can make assessments of personhood; 3) all assessments of personhood are negative (e.g. no respondents recalled students commending them on keeping their politics 'out of it' or using a particularly 'balanced' approach); and 4) assessments of personhood have the potential to more frequently occur for instructors with marginalized identities.

iii. Perceived Identity: Gender

Our respondents' most memorable comments revealed student assessments in five dimensions of identity – gender, race/ethnicity, age, language/accents, and sexuality – most of which are limited, or entirely absent, in current literature on SETs comments. Their experiences largely align with normative hierarchies we see in the quantitative results – privileging white, middle-aged men – and are widely criticized by respondents. Nevertheless, the breadth of students' evaluations is as vast as respondents' understandings and emotional confrontations with them, ranging from nuisance and flattery to feeling threatened and anxious. Below, we provide select examples to illustrate two major findings from our data: 1) readily perceivable identities (such as gender, race, age etc.) were more pronounced in memorable student feedback from our respondents and, 2) assessments of identity are understood and summarized by respondents (and sometimes students) in relative terms, highlighting a comparative (and potentially competitive) use of SETs. As these ideas are not yet fleshed out in current SETs literature, our findings point to fruitful avenues for future research on intersectional privilege and disadvantage for faculty in higher education.

As previous examples establish, gendered expectations of personality, including performativity (Butler 1990) – the defining and maintaining of identity (real and perceived) through speech acts and non-verbal communication – are embedded in students' reflections on pedagogy and personality. We see this, for example, in expectations of emotional labour and flexibility in exercising course policy. However, gender is also activated in students'

assessments of physical appearance/attraction and presentation of self. For respondents who identified as women and genderqueer/genderfluid, their personal style and desirability (i.e. attractiveness) were most commonly addressed, mirroring the findings seen in Figures 7 and 8. As Anita (44) notes, "in general students say nice things about me however my appearance has often been discussed by males [-] "a Milf" "sexy af for an older chic." For Mara, the same age as Anita, comments of this kind were more common at the beginning of her career, when students memorably remarked that she was "easy to listen to and easy to look at." In both cases, student evaluations are used as a mechanism for the male gaze (Mulvey 1975) – representations of women as objects of pleasure shaped by the lens of masculinity and heterosexual desire. Note for example that the comments Anita receives are highly sexualized. In fact, she directly addresses that her appearance has often been discussed by male students, but the assumption here needn't be that male/men students provided these comments, only that these comments reinforce the idea that women are objects of desire in systems of patriarchy. In this way, student evaluations of gender represent norms or expectations of physical appearance/attraction and presentation of self, not simply descriptive accounts. Nevertheless, independent of the students' identities, or their intentions, comments of this kind are as Anita describes, "demeaning."²⁰

While of no pedagogic value, and by some accounts "wildly inappropriate," (Evelyn, 46) these evaluations were among the most memorable for respondents; and in some cases, not uncommon (Jordan, 45). We do note, however, that Evelyn understands and makes sense of her feedback by comparing her

experiences to male professors: "I've received comments about my appearance, that I sometimes had a stain on my shirt, and other comments that I am certain male professors would not receive" – Evelyn, 46. This was a routine interpretive strategy by faculty grappling with feedback about personality and bears further investigation.²¹

Among respondents who identified as men, four recounted assessments of physical appearance/attraction and presentation of self. Ash (53) and Martin (69) remember comments about their attractiveness and sex appeal, respectively:

This one truly runs the gamut: "Your exams made me want to vomit. If I had to define "crime", I'd write [your] exams. Also, I found them to be inconsistent [sic]. Sometimes, what you advice [sic] to study is heavily tested (which is good), other times, your questions seem to come out of nowhere. However, you're a very good prof- in the sense that you teach well. Your [sic] also very attractive, I don't know why." – Ash

He is sexy in a bookish kind of way... – Martin

As with the bookish sex appeal of Martin (69) above, these comments are simplistic in nature, and barely reflected upon by the instructors who receive them. With so few examples, it is difficult to observe sufficient patterning. However, the lack of engagement/reflection with these comments by our male/men respondents point attention to the need for further examination – for example, are gendered differences in faculty engagement with comments about identity related to the frequency and severity of comments themselves?

²⁰ While it is likely that men provided these comments, there is no way to link evaluations to students' themselves.

²¹ See quotes from Gia, 58; Nick, 42; Patricia, 35 (below) etc. See also section on Race/Ethnicity below with further investigation into the 'relative' nature of assessments of identity.

iii. Perceived Identity: Race and Ethnicity

Reflections on race came from faculty in both privileged and marginalized groups, with respondents critical of their bias in the qualitative comments from SETs, consistent with our finding that fully 81% of all respondents think that the comments are biased by this variable (see Figure 6). For some, they noted the liberty they received to cover controversial course content with little to no resistance from students:

“I do feel good when students write: 'The best teacher I ever had' or especially comments like; 'This was a really tough course, but I learned so much and I see the world differently now', because they acknowledge that there is significant reading and writing -- but that they learn. On the other hand, I always take these with a grain of salt because I am the white daughter of a professor and I know that my class-cultural-capital confidence and whiteness mean -- despite teaching anti-racisms, and a strong focus on Indigenous scholarship in all of my classes -- that I get a 'free ride' in evaluations that many racialized colleagues do not. – Nila, 44.

In fact, as Nila (44) addresses, this is likely due to the combination of race, class, and cultural privilege; although class privilege and cultural competency are arguably (and literally) less visible. It is the noteworthy relative assessment that we wish to address here. On the one hand, Nila is addressing the 'free ride' she receives relative to her racialized colleagues (we return to this idea shortly). On the other hand, while SETs are meant to provide feedback to faculty on content, delivery, and related aspects of course design, they are undoubtedly procured by students

who make comparisons to other classes they're enrolled in. By doing so, information that should have no bearing on the assessment of teaching effectiveness (such as the popularity of the course topic) may become integrated in the evaluation holistically. In light of this, numerous faculty are at a disadvantage, long before assessments of identity are at play.

In several respondents' accounts, teaching unpopular, controversial or simply challenging material produces poorer comments – an empirical finding substantiated by quantitative evaluations as well. These are particularly fraught when evaluations of course content are filtered through the identity of the course instructor, such that students “reject both message and messenger, projecting their frustrations and emotions about this topic onto instructors” (Boatright-Horowitz and Soeung 2009). These reflections also illuminate the socio-psychological toll that negative comments from SETs, however rare, can have on faculty, even when they are an overall useful tool for pedagogic improvements. No clearer is this illustration than in the seemingly contradictory findings of Figure 12 which highlight strong support for SETs as a tool for pedagogic improvement, alongside concrete steps to reduce workload, diminish scholastic rigor and lower academic standards.

As students routinely reflect on personality and personhood through the lens of instructor's perceived identities (some of which we've already explored), intersectionally marginalized faculty are at risk of receiving the least helpful (pedagogically) and most psychologically damaging feedback. A particularly visceral example comes from Alanna (39), who

belongs to multiple marginalized groups and is untenured, “the one I always remember describe[s] my breasts. I always think about how that is in my permanent, professional file.” Summarizing their cumulative impact, she continues “the evaluations are BRUTAL in terms of gender, race, my body, etc. I used to read them when I began even though they were brutal, but now I have found other ways to get what I need from evals without having to read violence against me.”

While the experiences of racially marginalized faculty needn’t all be negative, as in the case of Genevieve (35), whose students mention her race and gender “in a positive sense,” the modal experience is negative, with a commensurate emotional toll. As such, many (irrespective of racial background) describe weariness in their approach to understanding and use of qualitative feedback from SETs; notwithstanding those electing to disregard them altogether, and those, for whom disregard is not an option, seeing SETs as “a tool for perpetuating institutional racism” (Lara, 41).

As students routinely reflect on personality and personhood through the lens of instructor’s perceived identities (some of which we’ve already explored), intersectionally marginalized faculty are at risk of receiving the least helpful (pedagogically) and most psychologically damaging feedback. A particularly visceral example comes from Alanna (39), who belongs to multiple marginalized groups and is untenured, “the one I always remember describe[s] my breasts. I always think about how that is in my permanent, professional file.” Summarizing their

cumulative impact, she continues “the evaluations are BRUTAL in terms of gender, race, my body, etc. I used to read them when I began even though they were brutal, but now I have found other ways to get what I need from evals without having to read violence against me.”

While the experiences of racially marginalized faculty needn’t all be negative, as in the case of Genevieve (35), whose students mention her race and gender “in a positive sense,” the modal experience is negative, with a commensurate emotional toll. As such, many (irrespective of racial background) describe weariness in their approach to understanding and use of qualitative feedback from SETs; notwithstanding those electing to disregard them altogether, and those, for whom disregard is not an option, seeing SETs as “a tool for perpetuating institutional racism” (Lara, 41).

“**They make up stories, they lie, and they are just down right nasty. It was worse when I was teaching courses that they didn't like, such as statistics and theory. Even when I was knowledgeable, one student felt that since I knew so much, that it was time for me to retire.” In this case, course content, and its intersection with gender, turn traits such as knowledgeability into detriments.**”

iii. Perceived Identity: Age

While gender and race were the most pronounced identities recounted by faculty in qualitative feedback from SETs, reflections on age, language/ accent and sexuality were also present. As seen in Figure 6, a large majority of our respondents believe that SETs comments are biased according to age (though we did not ask them about the direction of this perceived effect). The respondents' comments about age – both in early and late stages of respondents' teaching careers – were exclusive to women, were intertwined with sexist remarks, and reflected discrimination toward both younger and older faculty depending on the category of assessment. Jordan (45), a self-described “young’ish woman of colour,” regularly receives comments about her “wardrobe and personal style,” highlighting the presumptive liberties that students take to evaluate young women on their looks and presentation of self. Whereas Georgia (58) recounts the early years of her teaching career, noting the propensity for students to include inappropriate or unprofessional comments on SETs – “particularly as a younger faculty member, repeated comments about, ‘shoving feminism down peoples’ throats’.” As she summarizes, this is “an exceptionally violent and sexualised critique.”

In the case of Claudette (70) and Gloria (51), they report explicitly ageist remarks, and personal experience with increasingly negative comments on SETs over their teaching careers. As Claudette summarizes, “I have noticed that the older I get, the frequency of the negative comments have increased,” with some students calling her “too old” outright. Whereas Gloria reflects “I am still the same

teacher I always was but as I have aged -- no chili peppers for me on ratemyprofessor.com anymore -- my evaluations have gone downhill. I have also noticed that the comments often reflect my “failure” to be maternal towards my students, and that this expectation, and my failure to do so, has increased as I have aged.” Once again, we see an intersectional approach to discrimination, such that expectations of gendered presentation and performativity – personal style and maternal affect – are contingent upon age.

Taken together, we see the mobilization of male gaze during the early years of women’s careers, and a greater upholding of hegemonic femininity in later years (Mulvey, 1975). We take this as an indication that the effect of age on qualitative comments from SETs is likely curvilinear. As these are preliminary conclusions drawn with the limited data, we recommend further research in these areas to address the prevalence and consistency of “ageist and sexist nonsense,” as Doreen calls it.

iii. Perceived Identity: Language/Accent/Voice

While few in number, some respondents addressed and critiqued the comments they personally received from students regarding their language, accent or voice. Still, in keeping the fact that 90% of all respondents believe that SETs comments are biased by students' perceptions of these traits, numerous others acknowledged the existence of accent bias in broader concerns about SETs instruments (Amber 44; Gia 58; Patricia 35). James (35) recalls, “very negative comments about [his]²² accent,” but says nothing more. Whereas Callie (45)

²² Pronoun changed from “my” to “his.”

says she receives “unnecessarily positive comments on [her]²³ accent (British),” making her aware of how “people with less prestigious accents are probably negatively perceived” – perhaps akin to the experience of James. She later details that while receiving “some thoughtful comments,” there are “a lot of bewildering comments that I just don't know how to address, e.g. “condescending” (maybe that's about my accent, too...)” For Delores (63), some students have said that they “can't understand a single word she says,” although it is unclear whether this references language proficiency, accent, pitch, or some other dimension. However, with Harlow (51), she has received “many positive comments and some that have been helpful in teaching but the one [she] will never forget was about how the student couldn't stand [her] “grating high pitched voice.””²⁴

Taken together, these examples illuminate students' application of dominant stereotypes and other norms/expectations of behaviour. They demonstrate that, consistent with current literature, some accents are interpreted as signals of intelligence, knowledgeability and even condescension, while others are not. It is worth noting that social understandings of accent are often closely intertwined (and complicated) by issues of race and ethnicity, corresponding with perceptions of intelligence, physical attractiveness and trustworthiness (Anderson et. al. 2007). In the context of university classroom settings, the socially constructed nature of perception, and student evaluations of teaching, these mechanisms warrant greater empirical investigation.

²³ Pronoun changed from “my” to “her.”

²⁴ Pronouns changed from “I” and “my”, respectively” to “she” and “her.”

iii. Perceived Identity: Sexuality

A final pair of examples illuminate two respondents' experiences receiving comments regarding their sexuality. In both cases, respondents are openly gay and of similar age, but receive widely divergent feedback from students. In her own words, Carla (52) says, “I've been called names and called unqualified. One student suggested I be immediately fired. A couple of times people have been upset that I am openly gay. I also get many positive comments, but those don't stick with me the way the negative ones do.” In contrast, Sean (57) has had overwhelmingly positive feedback from students on his SETs feedback, “I'm an openly gay professor who shares stories about his life; [...]-- I have consistently found this level of openness; about the complexity, the day & light of life--has always been welcomed. I have had students tell me that I [am] profoundly kind & welcoming, approachable & have made a real impact on them--what more could I want.” Here, we make note of Carla and Sean's emotional responses, respectively. For Carla, student expressions of upset regarding her sexuality are part and parcel of a multitude of negative feedback. Whereas for Sean, openness about his sexuality is part and parcel of a pedagogic approach, which has been largely successful in his teaching career. It would be premature to argue that the reasoning behind these divergent student reflections is gender alone, but by virtue of its import to nearly all other areas of student evaluation, it clearly warrants greater inquiry.

Discussion

We now summarize the findings in light of our hypotheses.

H1. A majority of sociology instructors engage with their qualitative SETS comments and use them to improve their teaching practice.

This was confirmed in our findings: over 90% reported that they read their comments. To the extent that the comments contain content that is belittling, demeaning, and driven by biases ranging from gender and race stereotypes to disgruntlement with low grades and tough or controversial material, faculty remain willing to extract pedagogically relevant ideas from them.

H2. A majority of sociology instructors make pedagogical choices to preempt hostile feedback.

A set of questions in our survey concern the use that faculty make of their SETs comments for pedagogical purposes. These yielded some seemingly contradictory findings. Only one-third of the overall sample regard SETs comments as an effective means of determining teaching effectiveness, and 38% believe that the comments have contributed to eroding academic standards. The percentage of those who reduce various kinds of challenges and innovations hovers around 40%. How can it be, then, that almost two-thirds of our respondents report that SETs comments are helpful in improving course design? We think that answer may lie in how people define improvement: it may well be that faculty think of this in terms of achieving better SETs scores and comments in future courses. In the case of the comments, they may be aiming to pre-empt psychologically wounding,

abusive feedback that is ignited when they hold students to standards which have been dropping over the past few decades. Further research should employ more finely tuned questions to understand this.

H3. A majority of sociology instructors will have experienced more negative than positive effects of SETs comments on their self-esteem, self-confidence in their ability to teach, morale, motivation for teaching, and emotional well-being.

The strongest confirmation of this hypothesis in relation to the overall sample is that only a minority of respondents -- one-fifth to just under one-third -- indicated that SETs comments positively affect these key aspects of professional self-esteem. This should be considered together with the fact that, according to our qualitative results, hostile comments are more impactful and memorable than positive ones.

Only 34% of the total sample say that SETs comments have a negative impact on overall sense of well-being. However, when asked if anticipation of SETs comments provokes anxiety, a much higher percentage, 58%, reported affirmatively. This suggests that questions about well-being need to be framed in rather specific terms.

Our question about forms of affect that are activated by reading SETs comments yielded a mixed picture, with frustration and disappointment being the most

commonly reported, but with happiness, relief, and contentment more often reported than anger, sadness and despair.

H4. A majority of sociology instructors will receive comments that reflect pedagogically irrelevant, personalized criteria.

This hypothesis is confirmed for our sample overall, with percentages in the 70s, 80s, and 90s agreeing that the comments reflect criteria that have nothing to do with pedagogy. These largely have to do with ascribed characteristics of the instructors themselves. Though our question did not ask whether respondents themselves experience these biases, it can be inferred that this at least partly informs these responses.

Our qualitative findings reflect these quantitative results. Faculty's most memorable SETs comments are of almost no pedagogic value, even when they are related to elements of the course. This is exemplified in one respondent's observation after many years of teaching, that SETs comments (and perhaps SETs overall) are a "popularity contest." These open-ended responses also indicate that in the absence of commentary focused on pedagogy, students make routine assessments of personhood. From what was reported, all of these assessments are negative, and are often personal attacks masked as comments about substantive content. Taken together these findings illuminate a marked incongruence between what SETs are designed to do and what they actually do. We are alarmed to see the breadth and depth of SETs commentary as bearing little relationship to course pedagogy, even when comments are about the course.

H5. Higher percentages of faculty who are members of equity-seeking groups will receive comments that do not reflect pedagogical criteria, and that are hostile and abusive.

This hypothesis is largely confirmed.

Women, more than men, report receiving SETs comments with abusive, hostile or bullying content. Women, more than men, experience SETs comments as biased by factors unrelated to pedagogy, find SETs comments less helpful to improving their pedagogy, and make more modifications to their teaching practice (reducing the difficulty level) to avoid such unconstructive feedback. They were also more praised for adhering to gender stereotypes of nurturance, and less commended for intellectual abilities.

We also found that racialized respondents more often received abusive and hostile content in the comments. When it comes to the pedagogical impacts of the comments, more racialized than non-racialized faculty have pre-emptively reduced the difficulty level of their material.

A higher proportion of LGBTQ respondents consider SETs comment to be reflective of non-pedagogical criteria. In regards to the pedagogical impact of the comments, the main difference is that LGBTQ faculty consider them less useful for improving their teaching.

H6. Higher percentages of faculty who are members of equity-seeking groups will report negative effects of SETs comments on well-being.

This hypothesis is confirmed in regards to women and racialized faculty. Compared to their male peers, women faculty more often reported that SETs comments negatively affect their well-being, and that SETs provoke anxiety due to fear of negative feedback. Racialized respondents also experience more negative effects on well-being from SETs comments than their non-racialized peers.

The findings were less clear cut on this hypothesis for faculty who are sexual minorities: there were only small differences in effect on overall well-being and anxiety, and a difference in the opposite direction for receiving abusive or bullying comments compared to their heterosexual peers.

H7. We anticipate intersectional effects: a higher proportion of sociology instructors who belong to more than one of these identities will experience these negative effects.

Our findings confirmed this hypothesis. SETs comments have a more negative impact on sense of well-being, and were more anxiety-provoking, for respondents who belonged to two or more equity-seeking groups, than those who belonged to only one or none. They also more commonly see SETs comments as reflective of non-pedagogical factors, especially gender, race, age and language proficiency, and more commonly received abusive comments. Fully 26% of those who identified with two or more equity-seeking groups have considered leaving academia.

H8. Contingent instructors will report more negative impacts of SETs comments on well-being and professional self-esteem.

This hypothesis was confirmed for only one of our indicators of well-being and professional morale: SETs comments have prompted twice the proportion of contingent faculty, who make up 18.6% of our total sample, to consider leaving academia, compared to tenured/tenure track (33% versus 13%). That we did not find a bigger disparity in these impacts may mean that the deteriorating tone of SETs comments over recent years spares neither tenured/tenure-track nor contingent faculty. It is also possible that a larger sample with a higher proportion of contingent faculty would yield a more nuanced result. However, we also found that contingent faculty more often receive abusive comments than do tenured/tenure-track counterparts (74% vs. 58%), and that the abusive comments they receive are more often experienced as a threat to safety. These are results we did not anticipate based on prior literature.

“ Respondents who identified with two or more designated groups more often 41% indicated that qualitative comments have had a negative impact on their sense of well-being in comparison to those who identified with one designated equity-seeking group (38%) and those who identified with none (21%). A similar pattern is seen in instructors’ experience of anxiety during evaluation time. ”

Age

We did not generate specific hypotheses for age. But we did find meaningful differences between our recoded age groups in some key outcome variables. Being relatively younger correlated with several negative impacts, and with impacts that highlight perverse effects of SETs comments on pedagogy, in contradiction of the ostensible objective of SETs. When it comes to anxiety as an effect of anticipating SETs comments, the younger the faculty, the more anxiety they experienced. This is not surprising considering that some faculty in this group may be in early stages of their career, and feel that the comments, like SETs overall, weigh more in how they are assessed. A similar linear correlation appears with the perception of SETs comments as biased according to non-pedagogical criteria. When it comes to impact on pedagogy, a greater proportion of younger faculty find SETs comments useful for improving their teaching. But they also much more often reduced the difficulty level of their courses in fear of negative feedback. The younger the faculty, the less they considered SETs comments a valid reflection of teaching quality. This seeming contradiction between eagerness to use SETs comments to improve teaching, alongside a regard for SETs as invalid, and a tendency to reduce course difficulty to pre-empt negative feedback, mirrors what we see in the overall sample. We speculate that the contradiction hinges on how faculty (and in this case, younger faculty) define improvement: avoiding negative comments, especially of the unconstructive kind, may be a benchmark. The entrenchment of the summative and comparative purposing of SETs scores across universities would certainly contribute

to this kind of “gaming” of course content, and given the increasingly abusive content of SETs comments with the online delivery of the survey, it can be a rational means of emotional self-protection to steer course difficulty downward.

There were also some surprising and contradictory findings in regard to age. Though as mentioned, younger faculty experience more anxiety around SETs comments, the older the faculty, the more often they received comments that were abusive, hostile, or bullying. Yet older faculty more often regard SETs comments as a valid reflector of teaching effectiveness, and also want to keep comments as a part of SETs. We can only speculate that part of what is going on is that older faculty are more emotionally immune to abusive content because the comments are seen as less weighty in how they are evaluated. But further research is needed to understand ways that older and younger faculty feel penalized by the comments. Our qualitative results suggest that there are particular kinds of harassing and bullying comments for both younger and older women faculty, and this should be explored more fully.

Recommendations

Because this project was conducted under the auspices of the Canadian Sociological Association (CSA), we address our recommendations firstly toward Sociology departments, though we believe that they have relevance to departments housing other disciplines and multi-disciplinary programs. In that regard, we acknowledge again the financial support we received from the Ontario Confederation of University Faculty Associations (OCUFA), whose own recent report on SETs, which is not discipline-specific, informs various aspects of our study. Our prescriptions here derive from the key quantitative and qualitative findings of our survey, woven together with recent literature in the scholarship on innovations in teaching evaluation where these studies echo or reinforce our comments. It is noteworthy that our open-ended survey questions generated views we did not specifically solicit for rethinking the procedures, logistics, mechanics, purposing, and processing of students' commentary on teaching.

We note that the June 2018 Ryerson decision paves the way for university administrations to reconsider the weight given to SETs scores in decisions about tenure and promotion. But although "ripple effects" across North American universities were predicted shortly after the ruling (CAUT 2018b), we know of no overview or synopsis of the extent to which this is in fact happening. Furthermore, the Ryerson decision does not specifically address SETs comments -- their biases, and their implications for emotional well-being of faculty who are obliged to engage with them. The question of whether and how other

institutions are instructing evaluative committees in light of the Ryerson decision, or revisioning the way that SETs generally, and the comments specifically, are solicited and assessed, merits future research.

The principle twin harms of SETs comments for faculty well-being and for human rights and equity in those well-being impacts, is the increasing prevalence of hostile, abusive, harassing comments that reflect a troll culture in the broader milieu of mediated communications. Among the solutions that have recently been proposed by various sources to mitigate these effects, there are two that we think should not be entertained. One is the proposal to have third parties (perhaps an AI mechanism) vet or filter student comments to quarantine those which are toxic, so that the faculty member does not have to engage with them. Apart from the ambiguities of many borderline comments that would plague such an approach, we concur with OCUFA that this would deprive faculty and administration of the opportunity to hold verbal abusers to consequences whereby to prevent repeat offenses (OCUFA 2019), and also to reeducate them. Related to this latter point, filtering would also absolve administrations of the obligation, as we see it, to adequately educate and acculturate students a priori to a standard of professional, constructive feedback as part of a well-rounded post-secondary experience. To continue sidestepping this responsibility tacitly confirms and encourages a pattern of verbal bullying that can extend beyond the university milieu and beyond the undergraduate phase of students' lives and careers.

Secondly, we consider a non-starter the notion that universities only need to provide proper mentoring or counselling of faculty to cope better psychologically with verbal abuse in their SETs comments. This proposal implies that faculty are misreading and over-reacting to the upsetting content, and that the faculty fail to (or need training to) separate the unhelpful from the pedagogically useful components of the feedback comments. Like the proposal to filter out abusive comments, this would also enable administrations to continue to sidestep a serious educational responsibility toward students.

The societal tumult occasioned this year by the novel coronavirus COVID-19, and the resultant transformations to education at all levels, the full reach of which cannot yet be tallied, are sure to have implications for SETs design and use that we cannot fully predict. From the emergency remote teaching that virtually all North American universities implemented for the final weeks of the Winter 2020 semester, to fully online learning in the Summer semester, most of these institutions had declared or were contemplating a longer-term foray into fully or mostly online teaching for Fall 2020, with pivoting flexibility, as this report was being finalized. Online and hybrid approaches are likely to endure for as long as the pandemic remains unchecked by a well-distributed vaccine, without a doubt well into 2021 (Steele 2020).

Clearly there will need to be particular ways of interpreting and weighing SETs scores and comments for courses where students may not enjoy equitable access to online materials due to pre-existing

inequalities in household/family resources, not to mention unreliable or absent high-speed internet, and who may never meet their professor in person. We acknowledge that there will be a need to tailor the mechanics and content of surveys of student feedback on teaching depending on the specific combinations of course delivery mode and the inherent limitations of each. But notwithstanding the transformed context in which SETs will likely continue to be administered through this uncertain period, and the fact that no single study can yield definitive solutions for all institutional contexts, we believe our recommendations are general enough to have relevance for courses delivered online, in person, or some hybrid of the two. The long overdue revision of the purposing, design, and processing of SETs comments are as necessary in the present period of flux as they were before. We hope these recommendations inform productive dialogue in Sociology departments and circles across the country, moving us toward a transformation that takes into account intersectional identities of faculty, and differing institutional cultures (for example, those which are research- vs. teaching-intensive), to better serve both students and instructors.

1. The overarching intentionality of soliciting open-ended commentary from students on the quality of teaching must be formative only, not summative-comparative.

That is, SETs comments must be regarded, and designed, solely as a tool for instructors -- not the committees evaluating their teaching quality/performance (in comparison with the performance of others) -- to improve specific aspects of course design and delivery. In this regard, we echo prescriptions for change from a number of other sources (OCUFA 2019; Omer et al 2020 forthcoming). Open-ended feedback/commentary from students has great potential as a formative tool, as seen in several studies since the 1980s suggesting that faculty regard SETs comments as more pedagogically useful/informative than numerical ratings (sources cited earlier + Omer et al 2020 forthcoming). In our own study, a key finding in the responses to our open-ended questions is that relevant aspects of the context of either negative or positive comments that evaluative committees cannot know. This undermines their comprehensibility as indicators of whether an instructor did well or poorly in some area of their teaching.

2. As corollary of 1, comments on teaching should be solicited as the course is underway, and not at the end of the semester as has long been the convention.

This will obviate the impact of poor grades received and anticipated on the kinds of unconstructive and sometimes abusive comments that faculty perceive as punishment for the standards to which they have held students (we note that a large percentage of our respondents believe that SETs comments do reflect the way they have been graded in the course).

3. Also as a 2nd corollary of 1, faculty must be allowed to decide on the content of the open-ended questions for student feedback on their courses. The content of the questions should not be standardized and predetermined by departments or higher bodies.

For SETs comments to have formative use, faculty should be able to solicit specific, targeted feedback on particular innovations that they introduce, or on particular components of their teaching. In combination with a mid-semester administration of the survey, this will ensure that the feedback that is requested is focused on elements that instructors themselves know they need to, and still have time to, revise before the course finishes. Especially if these are administered in class, this should also mitigate the tendency we found for faculty to receive comments that are too few in number and too reflective

of the extremes of only high approval and high disapproval, to make constructive use of any given set of SETs comments. To ensure that students feel free to express critical perspectives without being identified by hand-writing in open-ended questions, they can be asked not to sign their questionnaires, and/or can be given the option of submitting typed responses online during the class. For purposes of evaluation of their performance, faculty can provide copies of the surveys they have designed, and how they responded, to evaluative committees to demonstrate their dedication to constant improvement of teaching.

4. Before being asked to complete SETs surveys, whether these are mid-semester or at end of term, students should be trained regarding how SETs are used both as an instrument for pedagogic improvement but also as a measure of teaching effectiveness in faculty's tenure and promotion files.

Here we echo prescriptions for change from a number of sources (Omer et al 2020 forthcoming; Lindahl and Unger 2020). This innovation will mitigate the twin harms of the current mechanisms surrounding SETs comments -- on faculty well-being (considering the anxiety that is occasioned by the arrival of SETs comments particularly for faculty in equity-seeking groups), and on the human rights and equity aspect of the comments' content and impact. Both our quantitative and qualitative findings point to faculty being punished and rewarded according to stereotypes based on gender, age, and race-ethnicity, and to some extent sexual orientation. Women were sexually harassed in their comments. There is also intersectional discrimination: harassing comments were worse for women of colour, and both younger and older women faculty recalled abusive, unconstructive, pedagogically irrelevant comments that are gender- and age-specific. Also, of particular relevance to sociology and closely related disciplines, many of our respondents reported ways that students direct their anger about, or rejection of, content related to social inequality, into attacks on faculty character. Furthermore, recent experimental research demonstrates that faculty receive comments that are less hostile and more pedagogically relevant when their students are provided such instruction. Even when comments are critical, they are constructive – specific, contextualized, and with concrete suggestions to improve (Moralejo et al 2019; Tucker 2014). We quote one of these studies at length because it captures one of the most important reasons that a transformation in the SETs comments are solicited is urgently needed: “Providing professional, thoughtful, mindful, and considerate feedback as well as understanding the potential negative influences of disingenuous or condescending feedback are essential skills for all post-secondary students. While students receive and deliver an abundance of feedback throughout their education, little or no time is allotted to teaching students how to deliver formal and informal feedback to others in a respectful and effective manner” (Moralejo et al 2019: 17).

5. University administrations should invest in an array of alternative or supplementary methods of evaluating teaching to replace end-of-term surveys of students' opinions.

We know of several universities, including the University of Guelph, that have established mechanisms for faculty to optionally request that peers from their own or other departments provide feedback on their teaching. These procedures should be systematic and routinized while still allowing customization to the nature of the course content and delivery mode, with adequate recognition and reward to faculty who volunteer to provide this service to their peers. There are also models of teaching evaluation in which students are hired and trained to undertake observation in classrooms of courses in which they are not enrolled, enabling both faculty and their student observers to understand and influence one another's perspectives (Cook-Sather and Motz-Story 2016).

6. End of term SETs, both ratings and comments, should be suspended immediately for contingent faculty.

They should be replaced with alternative means of obtaining student input, including mid-semester formative surveys designed by the contingent instructors, peer evaluations, and other methods. This is informed by recent literature findings on the impact of SETs ratings and comments on those teaching in a contingent status, as well as our own survey's finding that these faculty more often received abusive comments, and more often considered leaving the profession as a result.

References

Algozzine, Bob; Beattie, John; Bray, Marty; Flowere, Claudia; Gretes, John; Howley, Lisa; Mohanty, Ganesh; Spooner, Fred. 2004. "Student Evaluation of College Teaching: A Practice in Search of Principles." *College Teaching*. Vol. 52(4):, 134-141.

Alhija, Fadia Nasser-Abu., Fresko, Barbara. 2009. "Student evaluation of instruction: What can be learned from students' written comments?" *Studies in Educational Evaluation* 35: 37-44.

American Sociological Association (ASA). 2019. "Statement on Student Evaluations of Teaching." September.

Anderson, Samantha, Samuel D. Downs, Kaylene Faucette, Josh Griffin, Tracy King, and Stacy Woolstenhume. 2007. "How accents affect perception of intelligence, physical attractiveness, and trustworthiness of Middle-Eastern-, Latin-American-, British-, and Standard-American-English-accented speakers." *BYU Undergraduate Journal of Psychology* 3 (2007): 5-11

Arbuckle, Julianne and Benne D. Williams, 2003. "Students' Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations." *Sex Roles* 49(9-10) 507-516.

Basow, Susan A. & Julie L. Martin. 2012. "Bias in Student Evaluations." in M. E. Kite (Ed.), *Effective Evaluation of Teaching: A Guide for Faculty and Administrators*, pp. 40-49.

Bavishi, Anish., Madera, Juan M., Hebl, Michelle R. 2010. "The Effect of Professor Ethnicity and Gender on Student Evaluations: Judged Before Met." *Journal of Diversity in Higher Education* 3(4): 245-256.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review*, 94 (4): 991-1013.

Blizard, Lida Marie. 2016. "Faculty Members' Experiences of Cyberbullying by Students at One Canadian University: Impacts and Recommendations." *International Research in Higher Education* 1(1): 107-124.

Boring, Anne., Ottoboni, Kellie., Stark, Philip B. 2016. Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness "ScienceOpen Research: DOI: 10.14293/S2199-1006.1.-SOR-EDU.AETBZC.v1.

Bothwell, Ellie. 2018. "Tides turning against using student evaluations to rate staff: Experts predict ripple effect from ruling against Canadian university. Times Higher Education: July 26.

Boswell, Stefanie S. 2016. "Ratemyprofessors is hogwash (but I care): Effects of Ratemyprofessors and University-administered teaching evaluations on professors". Computers in Human Behavior 56(1): 155-162.

Braga, Michela., Paccagnella, Marco., Pellizzari, Michele. 2014. "Evaluating students' evaluations of professors" Economics of Education Review 41: 71-88.

Brockx, Bert., Van Roy, K., Mortelmans, Dimitri. 2012. "The student as a commentator: students' comments in student evaluations of teaching" Procedia – Social and Behavioural Sciences 69: 1122 – 1133.

Cassidy, Wanda; Chantal Faucher, and Margaret Jackson, Margaret. 2014. "The Dark Side of the Ivory Tower: Cyberbullying of University Faculty and Teaching Personnel." Alberta Journal of Educational Research 60(2): 279-299.

Canadian Association of University Teachers (CAUT). 2018a. "Academic Anxiety". CAUT Bulletin 5 May.
<https://www.caut.ca/bulletin/2018/05/academic-anxiety>

Canadian Association of University Teachers (CAUT). 2018b. "The End of Student Questionnaires?". CAUT Bulletin 11 November.
<https://www.caut.ca/bulletin/2018/11/end-student-questionnaires>

Canadian Association of University Teachers (CAUT) 2018c. "Underrepresented and Underpaid: Diversity and Equity among Canada's Post-Secondary Education Teachers."
https://www.caut.ca/sites/default/files/caut_equity_report_2018-04final.pdf

Ehrenberg, Ronald G. 2011. "Independent Colleges and Universities in a Time of Transition." Cornell Higher Education Research Institute (CHERI).
<https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1164&context=workingpapers>

Davison, Elizabeth & Jammie Price. 2009. "How do we rate? An evaluation of online student evaluations". *Assessment & Evaluation in Higher Education*. 34(1) 51-65.

Deo, Meera E. 2015. "A Better Tenure Battle: Fighting Bias in Teaching Evaluations." *Columbia Journal of Gender and Law* 31(1): 7-43.

Eagan, M. Kevin, Dr. and Jason C. Garvey. 2015. "Stressing Out: Connecting Race, Gender, and Stress with Faculty Productivity." *The Journal of Higher Education* 86(6): 923-954.

Eidenger, Andrea. 2017. "She's Hot: Female Sessional Instructors, Gender Bias, and Student Evaluations." *Active History*. March 30.

<http://activehistory.ca/2017/03/shes-hot-female-sessional-instructors-gender-bias-and-student-evaluations/>

Ewing, Vanessa Lynn; Arthur A. Stukas, and Eugene P. Sheehan, 2003. "Student prejudice against gay male and lesbian lecturers." *The Journal of Social Psychology* 143(5): 569-79.

Farr, Moira. 2018. "Arbitration Decision on Student Evaluations of Teaching Applauded by Faculty." *University Affairs* 18 August.

<https://www.universityaffairs.ca/news/news-article/arbitration-decision-on-student-evaluations-of-teaching-applauded-by-faculty/>

Flaherty, Colleen. 2019. "Speaking Out Against Student Evals." *Inside Higher Ed* 9 September.

<https://www.insidehighered.com/news/2019/09/10/sociologists-and-more-dozen-other-professional-groups-speak-out-against-student>

Flaherty, Colleen. 2015. "Professor's reflections on his battle with depression touch many at recent disciplinary meeting." *Inside Higher Ed*. February 25.

Flaherty, Colleen. 2018. "Bye Bye Chilli Pepper." *Inside Higher Ed* 2 July.

<https://www.insidehighered.com/news/2018/07/02/rate-my-professors-ditches-its-chili-pepper-hotness-quotient>

Flodén, Jonas. 2017. "The impact of student feedback on teaching in higher education" *Assessment & Evaluation in Higher Education* 42(7): 1054-1068.

Foster, Karen and Louise Birdsell Bauer. 2018. *Out of the Shadows: Experiences of Contract Academic Staff*. Canadian Association of University Teachers, 2018. https://www.caut.ca/sites/default/files/cas_report.pdf

- Freishtat, Richard L. 2016. Expert Report on Student Evaluations of Teaching (SET). Prepared for the Ryerson Faculty Association and the Ontario Confederation of University Faculty
- Kimmel, Krista M. & Jennifer L. Fairchild. 2017. "A Full-Time Dilemma: Examining the Experiences of Part-time Faculty" *The Journal of Effective Teaching* 17(1): 52-65.
- Kinman, Gail and Siobhan Wray. 2013. "Higher Stress: A Survey of Stress and Well-Being among Staff in Higher Education." For University and College Union.
https://www.ucu.org.uk/media/5911/Higher-stress-a-survey-of-stress-and-well-being-among-staff-in-higher-education-Jul-13/pdf/HE_stress_report_July_2013.pdf
- Kogan, Lori R., Schoenfeld-Tacher, Regina., Hellyer, Peter H. 2010. "Student evaluations of teaching: perceptions of faculty based on gender, position, and rank" *Teaching in Higher Education* 15(6): 623-636.
- Kowai-Bell, Neneh., Guadagno, Rosanna E., Little, Tannah E., Ballew, Jamie L. 2012. "Professors are people too: The impact of informal evaluations of professors on students and professors". *Social Psychology of Education* 15(3): 337-351.
- Lampman, Claudia., Crew, Earl C., Lowery, Shea D., Tompkins. 2016. "Women Faculty Distressed: Descriptions and Consequences of Academic Contrapower Harassment". *NASPA Journal About Women in Higher Education* 9(2): 169-189.
- Lewontin, Max. 2014. "For Adjuncts, A Lot is Riding on Student Evaluations." *Chronicle of Higher Education*. 6 October.
<https://chroniclevitae.com/news/741-for-adjuncts-a-lot-is-riding-on-student-evaluations>
- Lindahl, Mary W., Unger, Michael L. 2010. "Cruelty in Student Teaching Evaluations". *Collage Teaching* 58(3): 71-116.
- Lutovac, Sonja., Kaasila, Raimo., Komulainen, Jyki, Maikkola, Merja. 2017. "University lecturers' emotional responses to and coping with student feedback: a Finnish case study" *European Journal of Psychology of Education* 32(2): 235-250.
- MacNell, Lillian, Adam Driscoll, and Andrea N. Hunt. 2015. "What's in a Name: Gender Bias in Student Ratings of Teaching." *Innovative Higher Education* 40(4): 291-303.
- Mitchell, Kristina, M., Martin, Jonathan. 2018. "Gender Bias in Student Evaluations". *PS: Political Science & Politics* 1-5.

Murray, Darrin S. 2019. "The Precarious New Faculty Majority: Communication and Instruction Research and Contingent Labor in Higher Education." *Communication Education* 68(2): 235-63.

National Tertiary Education Union (NTEU). 2018. *Staff Experience of Student Evaluation of Teaching and Subject/Units*. Melbourne, Australia. June.

OCUFA. 2019. *Report of the OCUFA Student Questionnaires on Courses and Teaching Working Group*. February.

<https://ocufa.on.ca/assets/OCUFA-SQCT-Report.pdf>

Ogier, John. 2005. "Evaluating the Effect of a Lecturer's Language Background on a Student Rating of Teaching Form." *Assessment and Evaluation in Higher Education* 30(5): 477-488.

Omer, Kamel, Shoshanah R. Jacobs, Karl Cottenie, William Bettger, John Dawson, Steffen Graether, Coral Murrant, John Zettel, Genevieve Newton. Forthcoming, 2020. "Perceptions of student teaching evaluations and their formative value: does academic background matter?"

Smele, S., Quinlan, A. & LaCroix, E. 2020. (Under review). "It's an albatross": Precariously Employed Academic Women's Experiences of Student Evaluations of Teaching.

Reeley, Stan G. 2016. "Online Teaching May Be Perilous To Your Health: A Research and Reflection Essay on the Not-So-Well-Being of College Professors in 21st Century Classrooms". *Journal of Marketing Development and Competitiveness* 10(2): 19-23.

Reichman, Hank. 2020. "The Vulnerability of Contingent Faculty and Student Evaluations." *The Faculty Lounge: Conversations about Law, Culture, and Academia*. 22 January.

<https://www.thefacultyounge.org/2020/01/the-vulnerability-of-contingent-faculty-and-student-evaluations.html>

Reid, Landon 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.

Reeve, Gretchen R., and Grace Deason. 2014. "Predictors of Depression, Stress, and Anxiety among Non-Tenure Track Faculty." *Frontiers in Psychology* 6: 1-17.

Schueths, April M., Gladney, Tanya., Crawford, Devan M., Bass, Katherine L., Moore, Helen A. 2013. "Passionate pedagogy and emotional labor: students' responses to learning diversity from diverse instructors". *International Journal of Qualitative Studies in Education* 26(10): 1259-1276.

Singh Kang, Lakhwider, and Harpreet Sidhu. 2015. "Identification of Stressors at Work: A Study of University Teachers in India" *Global Business Review* 16(2) 303-320.

Smith, Bettye P., and Billy Hawkins. 2011. "Examining Student Evaluations of Black College Faculty: Does Race Matter?" *The Journal of Negro Education* 80(2): 149-162

Sonntag, Michael E., Jonathan F. Bassett, and Timothy Snyder. 2009. "An Empirical Test of the Validity of Student Evaluations of Teaching Made on RateMyProfessors.com." *Assessment and Evaluation in Higher Education* 34(5): 499-504.

Steele, Ken. 2020. "Near-Term Impacts of COVID-19." *Eduvation Blog* 27 March. <http://eduvation.ca/2020/03/covid-impacts2/>

Storage D, Horne Z, Cimpian A, Leslie S-J (2016) The Frequency of "Brilliant" and "Genius" in Teaching Evaluations Predicts the Representation of Women and African Americans across Fields. *PLoS ONE* 11(3): e0150194. <https://doi.org/10.1371/journal.pone.0150194>

Subtirelu, Nicholas Close. 2015 "She does have an accent but...": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessor.com." *Language in Society* 44: 35-62.

Uttl, Bob., White, Carmela A., Wong Gonzalez, Daniela. 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.

Vasey, Craig & Linda Carrol. 2016. "How Do We Evaluate Teaching?: Findings from a survey of faculty members". For the American Association of University Professors (AAUP). <https://www.aaup.org/article/how-do-we-evaluate-teaching#.WljEw9JKjIU>.

Wasserstein, Ronald L, and Nicole A. Lazar. 2016. "The ASA Statement on p-Values: Context, Process, and Purpose." *The American Statistician* 70(2): 129-133. DOI: 10.1080/00031305.2016.1154108

Wong, Wai Yee & Karen Moni. 2014. "Teachers' perceptions of and responses to student evaluations of teaching: purposes and uses in clinical education" *Assessment & Evaluation in Higher Education* 39(4): 397-411.

Appendix A: Survey Instrument

Introduction Page

Faculty Experience with Student Evaluations of Teaching (SETs)

You are invited to participate in a survey on the experiences that faculty have with student evaluations of teaching (SETs), particularly the qualitative comments found in SETs. The survey will take approximately 15 minutes to complete. Please be assured that your participation in this survey is voluntary and you may skip any questions you do not feel comfortable answering. You may also terminate the survey at any time. All responses are confidential and no identifiable information will be collected. Any results will be aggregated when being reported. The project has been reviewed and approved by the University of Guelph Research Ethics Board. By clicking on the "start" button you consent to participating in the survey and have read and understood the information letter attached to the invitation email.

Faculty Experience with Student Evaluations of Teaching (SETs) Survey

I. Administration and use of SETs

Questions in this section focus on how SETs are administered and utilized at your university.

1. Which of the following best describes the type of student evaluations of teaching (SETs) your department/institution administers?

Online student evaluations

Paper student evaluations

Both online and paper student evaluations

My department/institution does not administer SETs

Other, please specify _____

[If answered "My department/institution does not administer SETs" to question 1]

1. a) Have you ever worked in a department/institution that administers student evaluations of teaching (SETs)?

Yes

No

Don't Know

[If answered YES to question 1a then skip to question 6]

[If answered NO to question 1a then person gets directed to the end of the survey where it thanks them for their participation]

2. Does your department/institution make the results of your student evaluations of teaching (SETs) publicly available?

Yes, but I am able to opt-out of doing so

Yes, and I am unable to opt-out of doing so

No

Don't Know

3. In your opinion, how important are the quantitative scores on student evaluations of teaching (SETs) when it comes to promotion, salary increases, or tenure recommendations?

- 1 - Not at all important
- 2 - Slightly important
- 3 - Moderately important
- 4 - Important
- 5 - Very important
- Don't Know

4. In your opinion, how important are the qualitative (i.e. open-ended) comments on student evaluations of teaching (SETs) when it comes to promotion, salary increases, or tenure recommendations?

- 1 - Not at all important
- 2 - Slightly important
- 3 - Moderately important
- 4 - Important
- 5 - Very important
- Don't Know

II. Your engagement with student evaluations of teaching (SETs) comments

In this section, we are asking about your own engagement with the qualitative comments on student evaluations of teaching (SETs).

5. Do you read the qualitative (i.e. open-ended) comments on your student evaluations of teaching (SETs)?

- Yes
- No
- Don't Know

[If answered NO to question 5 to 5a; if YES to question 6]

5. a) Please tell us why you do not read the qualitative comments on your student evaluations of teaching (SETs)? For subsequent questions, please reflect on your recollection of student evaluations of teaching (SETs) when you did read the qualitative comments.

III. Pedagogical relevance of student evaluations of teaching (SETs)

Questions in this section ask your opinion on the general utility of student evaluations of teaching (SETs) for pedagogy, and the impact they have had on your own teaching.

6. Please indicate your view on each statement using a scale from 1 (strongly disagree) to 5 (strongly agree): 1 - Strongly disagree 2 - Somewhat Disagree 3 - Neither agree nor disagree 4 - Somewhat Agree 5 - Strongly agree - Don't Know

- a) Qualitative comments on student evaluations of teaching (SETs) reflect students' grade expectations
- b) Qualitative comments on student evaluations of teaching (SETs) reflect the size of the enrollment
- c) Qualitative comments on student evaluations of teaching (SETs) reflect instructor attractiveness

9. Please indicate your view on each statement using a scale from 1 (strongly disagree) to 5 (strongly agree): 1 - Strongly disagree 2 - Disagree 3 - Neither agree nor disagree 4 - Agree 5 - Strongly agree - Don't Know

- a) In general, qualitative comments on student evaluations of teaching (SETs) should be part of the evaluation of teaching effectiveness
- b) In general, qualitative comments on student evaluations of teaching (SETs) have helped me become a better instructor

IV. Impact of student evaluations of teaching (SETs) qualitative comments on your sense of well-being. Questions in this section focus on the impact of SETs on your emotions, morale, self-esteem, motivation, and sense of safety.

10. Which of the following emotional responses have you experienced after reading the qualitative comments on your student evaluations of teaching (SETs)? (Check all that apply)

sadness

disappointment

despair

frustration

anger

happiness

relief

contentment

joy

hope

Other, please specify _____

11. Please indicate to what extent the qualitative comments on your student evaluations of teaching (SETs) increases the following (1 - Not at all; 2 - A little; 3 - A moderate amount; 4 - A lot; 5 - A great deal):

- (a) confidence that I would be able to teach my future classes well
- (b) confidence that I would have good rapport with future classes
- (c) self-esteem
- (d) confidence that I should continue being a professor
- (e) confidence in my ability as a professor
- (f) confidence that the students like me
- (g) enthusiasm to teach again

12. In general, what impact have qualitative comments from student evaluations of teaching (SETs) had on your sense of well-being?

Very positive

Somewhat positive

Neither positive nor negative

Somewhat negative

Very negative

13. Do you feel anxiety during evaluation time due to fear of negative feedback?

Yes

No

Don't Know

14. Have you ever considered leaving academia due to negative qualitative comments you received on student evaluation of teaching (SETs)?

Yes

No

Don't Know

15. Have you ever received qualitative comments in your SETs that you perceive as abusive, hostile, or bullying?

Yes

No

Don't Know

[If yes to question 15, move to Q15a. If answered No or or Don't Know to Q15, move to Q16]

15. a) Have the qualitative comments in your SETS ever made you feel unsafe?

Yes

No

Don't Know

V. Employment Status

Questions in this section focus on your employment status and years in the profession.

16. What is your current academic position? (check all that apply)

Sessional (course instructor)

Limited term appointment (all ranks)

Instructor

Researcher

Assistant professor

Associate professor

Full professor

Other, please specify _____

17. What is the status of your current academic position?

Full-time

Part-time

Other (please specify)

18. How many years have you been teaching at the university level? _____

19. What is your tenure status at your current institution?

Tenured

On tenure track, but not tenured

Not on tenure track

No tenure system for my faculty status

No tenure system at this institution

VI. Socio-demographic information

In this final section we would like to gather some demographic information about you.

20. In what year were you born? _____

21. Which of the following best describes your gender identity? (Check all that apply)

Woman

Man

Transgender

Genderfluid/Genderqueer

Non-binary

Two-spirit

Agender

An identity not listed, please specify _____

Prefer not to say

22. Do you consider yourself to be:

Heterosexual or straight

Gay or lesbian

Bisexual

An identity not listed, please specify _____

Prefer not to say

23. What is your current marital status?

Married

Living common-law

Widowed

Separated

Divorced

Single, never married

Prefer not to say

24. Do you identify as an Indigenous person (i.e. First Nations, Métis or Inuit)?

Yes

No

Prefer not to say

25. Which of the following best describes your racial/ethnic/national identity. (Check all that apply)

White

Chinese

South Asian (eg. East Indian, Pakistani, Sri Lankan, etc.)

Black

Filipino

Latin American

Southeast Asian (e.g. Vietnamese, Cambodian, Malaysian, Laotian, etc.)

Arab

West Asian (e.g. Iranian, Afghan, etc.)

Korean

Japanese

Other group, please specify _____

Prefer not to say

26. Please feel free to share the most memorable/impactful qualitative comments you have received from your student evaluations of teaching (SETs) in the past.

27. Please feel free to add anything else about student evaluations of teaching that you feel we have not asked.

Exit Page

Thank you for taking the time to complete our survey. A summary of the analysis of the aggregated results will be made available on the website of the Canadian Sociological Association (CSA) once the study has been completed. <https://www.csa-scs.ca>

If you have any questions or concerns about our study, please feel free to contact Lisa Kowalchuk: lkowalch@uoguelph.ca.

If you have any concern about ethical aspects of your participation in this survey, please contact Director, Research Ethics; University of Guelph; reb@uoguelph.ca; (519) 824-4120 (ext. 56606).

Appendix B: Sample Characteristics (n= 288)

	n	Percentage
Current academic position		
Sessional/LTA/instructor	52	18.6
Assistant professor	52	18.6
Associate professor	104	37.1
Full professor	72	25.7
Status of current academic position		
Full-time	246	85.7
Part-time	39	13.6
Years teaching at the university level	Mean = 16.8	SD = 9.5
Tenure status		
Tenured	184	64.6
On tenure track, but not tenured	47	16.5
Not on tenure track/no tenure system at institution	54	19.0
Age		
40 and under	41	16.5
41 to 59 years	159	63.9
60 and over	49	19.7
Gender Identity		
Women	163	56.8
Men	102	35.5
Non-binary	15	5.2
Racial Identity		
Non-racialized	224	84.9
Racialized	40	15.2
Sexual Orientation		
Heterosexual	216	80.9
LGBQ	51	19.1
Marital Status		
Married or common-law	212	75.5
Widowed/separated/divorced	27	9.6
Single, never married	29	10.3

Note: percentages may not add up to 100% due to rounding or the decision to not report when n<10. It should also be noted that not every respondent answered every question.

CSA Canadian Sociological Association
Société canadienne de sociologie

