Journal of Women and Minorities in Science and Engineering 29(6):v–xii (2023)

TEACHING EVALUATIONS AS A SITE OF INSTITUTIONAL INTERVENTION TO IMPROVE CULTURE

Robin Fowler,[1,*] Trevion Henderson,[2] & Katie Snyder[1]

[1]*University of Michigan, Michigan Engineering, 2609 Draper Dr, Ann Arbor, MI 48109, USA*

[2]*Tufts University School of Engineering, Mechanical Engineering, 212 College Avenue, Medford, MA, 02155, USA*

*Address all correspondence to: Robin Fowler, University of Michigan, Michigan Engineering, 2609 Draper Dr, Ann Arbor MI 48109, USA; Tel.: +(734)615-3095, E-mail: rootsr@umich.edu

Student evaluations of teaching are a controversial medium; there's some meaningful information and some bias encoded in the data. Many institutions have recognized this and adjusted in terms of how quantitative information is used in promotion and retention decisions, but they are (mostly) ignoring the harms that student evaluations can have on instructors. The anonymous open-ended feedback students provide is specifically a place that enables trolling. Trolling has negative effects on instructors, especially instructors of color and women in male-coded disciplines. There are some things institutions can do about it.

"I used to love technical writing, but ROBIN FOWLER has ruined it for me."

"I used to love technical writing, but (other instructor) has ruined it for me." (But Robin felt better seeing that this same comment was applied to someone else/it wasn't personal!)

"Now I understand why all the faculty say she's just here because of her husband."

"I hope she pops that baby out soon, because she looks miserable and it makes me uncomfortable."

"I am genuinely curious as to whether she has the credentials required to teach me."

"Try to explain things better in the time you use up."

"Not sure if she is a bad professor or if the content is just bad, but probably a mix of both. I would not take her again in my life."

Student evaluations of teaching (SETs) are an important feedback forum and an opportunity for instructors to gather feedback that can be used to improve teaching. However, comments made on teaching evaluations are sometimes hugely inappropriate and intentionally hurtful. We suspect these comments are disproportionately directed at instructors of color, at women in men-coded disciplines such as those in STEM, and at any of us who do not fit students' stereotypes of "professor." Student instructors, including undergraduate lab assistants and graduate students considering careers in teaching, receive such evaluations as well. We argue that institutions owe our colleagues and student instructors better oversight. Rightly, institutions of higher education are invested in identifying and mitigating barriers to success faced by instructors. One opportunity institutions may currently overlook is better monitoring of the typically anonymous channel of open-ended feedback on SETs.

Studies of bias in SETs often focus on quantitative analysis of numerical ratings (Chávez and Mitchell, 2020; Felkey and Batz-Barbarich, 2021; Kreitzer and Sweet-Cushman, 2021; Laube et al., 2007; Morgan et al., 2016; Mengel et al., 2019; Reid, 2010; Wagner et al., 2016; Okoye et al., 2020; Sprague and Massoni, 2005; Wallace et al., 2019). This important scholarship, much of which identifies relationships between instructor racial and gender identities and numerical scores from students, has led to arguments that SETs be de-emphasized in hiring, promotion, and retention decisions (Hornstein, 2017; Read et al., 2001). Institutional response to this DEI-related concern has been to de-emphasize or even exclude quantitative SET data from important decisions, such as promotion and tenure.

However, institutions need to think beyond how teaching evaluations are used for promotion and tenure to understand the full potential harm of racism, sexism, and other biases in SETs. SETs are an important source of formative feedback through which instructors view themselves in their instructional role. Especially for new instructors, such as student instructors and some junior faculty, a single course's SET feedback can have an outsized impact on their self-efficacy in the classroom (Moore and Kuol, 2005). The fact that this institutional feedback system encodes racism and bias from larger society is not surprising, but it is concerning. Institutions have a responsibility to consider how to mitigate the harm of SETs, especially as they work to make campus more inclusive, equitable, and just. We specifically note that oversight of open-ended written feedback given as part of SETs is a critical step institutions can take that will make our institutions more welcoming and supportive for instructors from marginalized backgrounds (e.g., instructors of color, women/non-cis-men instructors).

We view feedback, including SET feedback, as varying along two axes: it can be positive or negative, and it can be constructive or nonconstructive. To our mind, all constructive feedback is valuable. Any reflective practitioner needs to learn to solicit and use feedback well. There is value in understanding student perspectives on what worked well and what did not work well to refine our pedagogical practices. We do recognize that such commentary can still be racialized and gendered, and in our "what can be done" section we describe some opportunities for institutions and for individual instructors to help to mitigate the impact of racialized and gendered biases in SET feedback.

In contrast, feedback that is not constructive, whether positive or negative, is problematic, and yet it often makes up a nontrivial amount of the feedback instructors receive. Our greatest concern is with feedback that is negative but not constructive, and especially with feedback that is hostile, abusive, and vicious in nature; we argue these comments are a form of "trolling" that anonymous SETs expose instructors to.

SETs are a form of anonymous online communication that can enable trolling comments directed at instructors, disproportionately affecting underrepresented groups but terrible for everyone. We are concerned with the trolling we have seen and heard colleagues and friends talk about and the impact such trolling can have on instructor mental health, sense of belonging and purpose, psychological and physical safety, and other factors shaping instructors' affective experience at an institution. Just as trolling can disrupt online communities, we argue that trolling in SETs can disrupt academic environments by harming instructors in material ways. Thus, we contend that institutions must develop practices that address trolling in SETs in order to protect the safety and wellbeing of instructors, as well as foster a healthy learning community for students and instructors alike. We believe that these negative experiences with trolling-type feedback are borne disproportionately by instructors who mismatch students' expectations of prototypical faculty; in STEM disciplines, this is often instructors of color and/or women.

For example, while the student comments from an evaluation in a midlevel lab course gave the first author some good thoughts regarding changes I could make for the class, I almost lost sight of them, focusing on one: "I used to love technical writing, but ROBIN FOWLER has ruined that for me." This single comment—meant by someone to express their dissatisfaction in the course—has shaken my sense of purpose at my job. It affects my confidence that I can do this work, and it affects my job satisfaction. It is also not useful—there is no information here that helps me produce a better experience for students going forward.

We do not criticize a student feedback forum lightly. We believe that students need more opportunities to voice their opinions and to be heard, but SETs may not be that space. We argue that SETs instead often function as a space for malicious actors to enact harm—trolling—on faculty and student instructors. The verb "to troll" was coined to capture a form of speech that arose with the internet; it (originally) specifically referred to the internet-linked phenomenon of the use of inflammatory messages to provoke emotional reactions and hostility, disrupting civil discourse (De Seta, 2018). While certainly not all communication that happens "off the internet" is civil, characteristics of internet communication have made trolling more prevalent. Internet communication is often low in social presence, with the communicator and the receiver both sometimes anonymous, faceless, or a name referring to a person unknown to the other, allowing "trolls" to escape accountability for hostile, abusive, and vicious comments that disrupt or even destroy online communities.

Similarly, the anonymity of SETs may empower students to disregard the norms that typically govern civil discourse. The literature on communication media often points to the ways that total anonymity informs what is communicated/what communicators are willing to communicate, with both positive and negative impacts. Social

presence theory suggests text-based communication decreases the salience of an audience, allowing a communicator to be less concerned with the impact of words on an addressee and increasing the ability to be critical (Cobb, 2009; Kehrwald, 2008; Kreijns et al., 2022). Anonymity removes some of the norms that lead to self-silencing (e.g., Christopherson, 2007; Dillon et al., 2015; Jong et al., 2013; Schlosser, 2020). This "disinhibition" related to anonymous channels of communication can have positive effects, such as allowing students to brainstorm more freely without prejudging and self-censoring (Fowler, 2015), or it can have negative effects, such as enabling online commenters to more freely troll (Huang et al., 2015). We contend that the feedback requested in SETs does not benefit from the decreased concern regarding the impact of one's words on the addressee and lessened accountability of an online, text-based, truly anonymous channel.

SET comments that provoke negative emotional responses, rather than foster instructors' reflections and improvements, can cause short- and long-term harm to faculty and student instructors (Kowai-Bell et al., 2012; Moore and Kuol, 2005). Provocative SET comments can affect instructors' sense of belonging and self-efficacy in a discipline and likely have implications for patterns of attrition in academia (Hutchins and Rainbolt, 2017). These effects would be exacerbated during the formative years of student and faculty careers, such as when junior faculty or student instructors are teaching courses for the first time. Moreover, while the affective effect of SETs on instructors has not been a major focus of empirical research, the impact of negative, provocative (i.e., trolling) SET comments may be disparately felt by faculty and student instructors from underrepresented or historically excluded backgrounds in their respective disciplines (e.g., women and/or instructors of color in engineering) (Kogan et al., 2010).

Currently, at most institutions, unfiltered SET feedback is provided directly to instructors. Given that it is often unproductive and sometimes uncivil, racist, and sexist, we argue that institutions have an obligation to interrupt this system. While the first author argued elsewhere (Fowler, 2019) that total anonymity is one factor enabling incivility, we note that there are many routes institutions could take to improve the current system.

1. Filter student comments

   In a teaching blog in the *Chronicle of Higher Education*, Supiano (2019) suggests instructors might find a trusted person to read through our SET comments and then summarize them back to us in a constructive form. We would push this suggestion further, arguing that the responsibility to filter and summarize should fall upon our institutions. The current model of providing anonymous open-ended feedback directly to faculty and student instructors perpetuates harm that institutions need to mitigate.

   Our suspicion is that any such filtering would be resource intensive. One option might be to work within an institution's pedagogical support structures (and provide appropriate resources). Consultants in teaching and learning centers could filter and summarize student feedback in productive ways as part of consultations with faculty. Department chairs could read through and provide

summary feedback for faculty. Faculty mentors could read through and provide feedback for new faculty or student instructors.

Another option might be to use language processing tools. Language processing tools are already in use across various settings to mitigate the negative social and economic consequences of hate speech (e.g., social media platforms) and other hostile acts fostered by anonymity in online commenting spaces. As such, there is a robust literature on fairly sophisticated language modeling for the purpose of mitigating the impact of trolls in anonymous online spaces (e.g., Kunupudi et al., 2020). These language processing tools range in complexity from simpler techniques, such as sentiment analysis (e.g., Birjali et al., 2021) or filtering specific hostile, negative words, to more sophisticated approaches such as identifying potentially inappropriate responses prior to submission and indicating to SET writers (i.e., students) that they should review, revise, or delete before submission. Unfortunately, we worry that the resources required to create any system of filtering and summarizing will make such opportunities untenable.

2. Remove total anonymity

We advocate removing total anonymity but note that it is important instructors do not know who wrote what feedback. Instructors might see these students again, and students should be able to provide critical feedback without worrying that their comments or ratings might affect letters of recommendation, experience in later courses, or specific interactions with the faculty member in the future. However, the anonymity that characterizes the SET channel is an enabler of uncivil responses.

Anonymity-to-instructors could be preserved while total anonymity might be removed from the system. For example, student evaluators could be told that someone (perhaps the chair of the department or the dean of the college) can identify authors of inappropriate comments. This model of evaluation might operate similar to reviewing for academic publishing: while instructors don't need to know what comments come from what reviewers, an "editor" knows who those reviewers are. In that context, while feedback could sometimes be framed more constructively (We're talking to you, Reviewer #2!), it's rarely or never intended as simple flaming. Even in academic publishing, with nonanonymity and presumably more mature reviewers, we see inappropriate comments. From our own experience chairing divisions at conferences, and from colleague-editors of journals in our field, we know that editors still find themselves filtering and/or redacting comments at times to protect authors from inappropriate hostility. Instructors deserve that same consideration.

We do want to emphasize the importance of true anonymous channels as well. Students absolutely need to have mechanisms for reporting a number of teaching-related (teaching-adjacent?) experiences, including bullying and harassment. Institutions must actively work to minimize barriers to such reporting, including maintaining anonymous channels for such reporting as well as ensur-

ing students are aware of such systems. At both of our institutions, the Ombuds office is one such mechanism, but it is one some students may not be aware of. We contend that SETs, which have the instructor as their primary audience, are not the appropriate mechanism for such reports. A system where student anonymity was not guaranteed would not eliminate unproductive negative feedback, but it might help to reduce it.

3. Train students/faculty for good feedback

   Not all negative feedback is inappropriate, and we want to maintain the student voice in SETs. At the institutional level, there is an opportunity, to help students develop the professional skill of providing critical feedback that might serve them in their professional lives. Not only will well-formulated feedback be less traumatic for instructors to read, but it will be more likely to be read as it is intended and to result in actionable improvements in instruction.

   A number of good sources summarize characteristics of constructive feedback. From an editorial in this journal (targeted at reviewers; Martin, 2020), we highlight that constructive feedback should be specific, actionable, prioritized, balanced, and contextualized. Martin also notes that this feedback should be provided in a timely manner and in language that focuses on the "work product" (for academic paper review, the journal article; in this context, the class/lectures, etc.) rather than on the person.

   Our institutions have used the opportunity of SETs to provide such instruction for students. University of Michigan's Center for Research on Learning and Teaching, for example, makes a handout available for instructors to share with students at evaluation time (University of Michigan, n.d.). We appreciate such efforts and would love to see this instruction integrated into evaluation systems. Peterson and colleagues (2019) find that adding language to SET instructions highlighting potential gender bias improved the gender gap in numerical ratings; we are curious whether it also impacted the rare, trolling-type comments we are most concerned with. In a non-SET context, Seiter and Brophy (2022) find that anonymous online forums that carefully codify norms see more prosocial behavior, suggesting that institutions that describe appropriate feedback for students might see improvements in framing of responses.

   At the same time, faculty can be supported in adopting a more productive response to critical student feedback. Cornes and colleagues (2023) provide a series of suggestions for both institutions and instructors; we would highlight especially the suggestions to "recognize and account for the challenges of innovation," "observe and reflect on one's immediate response to feedback," as well as the reminders to institutions to "provide active oversight of the evaluation process," and "develop learners' ability to provide effective feedback."

   In summary, student evaluations of teaching are a controversial medium; there's some signal and some bias encoded in the data. Institutions have recognized this and adjusted in terms of how quantitative information is used in promotion and

retention decisions, but they are (mostly) ignoring the harms that student evaluations can have on instructors. The anonymous open-ended feedback of SETs is specifically a place that enables trolling. Trolling has negative effects on instructors, especially instructors of color and women in male-coded disciplines. There are some things institutions can do about it.

REFERENCES

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134.

Chávez, K., & Mitchell, K. M. (2020). Exploring bias in student evaluations: Gender, race, and ethnicity. *PS: Political Science & Politics*, 53(2), 270–274.

Cobb, S. C. (2009). Social presence and online learning: A current view from a research perspective. *Journal of Interactive Online Learning*, 8(3).

Cornes, S., Torre, D., Fulton, T. B., Zoa, S., Teherani, A., & Chen, H. C. (2023). When students' words hurt: 12 tips for helping faculty receive and respond constructively to student evaluations of teaching. *Medical Education Online*, 28(1), 1–8.

De Seta, G. (2018). Trolling, and other problematic social media practices. *The Sage Handbook of social media*, 390–409.

Dillon, K. P., Neo, R. L., & Seely, N. (2015). Civil keystrokes: Examining anonymity, politeness, and civility in online newspaper forums. *AoIR Selected Papers of Internet Research*.

Felkey, A. J., & Batz-Barbarich, C. (2021). Can women teach math (and be promoted)? A Meta-analysis of gender differences across student evaluations of teaching. In *AEA Papers and Proceedings* (Vol. 111, pp. 184–89).

Fowler, R. (2015). Talking teams: Increased equity in participation in online compared to face-to-face team discussions. *Computers in Education Journal*, 6(1), 21–44.

Fowler, R. (2019). Protecting students or enabling trolls? Why faculty evaluations should not be totally anonymous. *Scatterplot: The unruly darlings of public sociology.* Retrieved from https://scatter.wordpress.com/2019/06/.

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1), 1304016.

Huang, N., Hong, Y., & Burtch, G. (2015). Anonymity and language usage: A natural experiment of social network integration. In *Proceedings of the 35th International Conference on Information Systems*.

Hutchins, H. M., & Rainbolt, H. (2017). What triggers imposter phenomenon among academic faculty? A critical incident study exploring antecedents, coping, and development opportunities. *Human Resource Development International*, 20(3), 194–214.

Jong, B.-S., Lai, C.-H., Hsia, Y.-T., & Lin, T.-W. (2013). Effects of anonymity in group discussion on peer interaction and learning achievement. *IEEE Transactions on Education*, 56(3), 292–299.

Kehrwald, B. (2008). Understanding social presence in text-based online learning environments. *Distance Education*, 29(1), 89–106.

Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P. W. (2010). Student evaluations of teaching: Perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, 15(6), 623–636.

Kowai-Bell, N., Guadagno, R. E., Little, T. E., & Ballew, J. 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.

Kreijns, K., Xu, K., & Weidlich, J. (2022). Social presence: Conceptualization and measurement. *Educational Psychology Review*, 34(1), 139–170.

Kunupudi, D., Godbole, S., Kumar, P., & Pai, S. (2020). Toxic language detection using robust filters. *SMU Data Science Review*, 3(2), 12.

Kreitzer, R. J., & Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, 1–12.

Laube, H., Massoni, K., Sprague, J., & Ferber, A. L. (2007). The impact of gender on the evaluation of teaching: What we know and what we can do. *NWSA Journal*, 87–104.

Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535–566.

Moore, S., & Kuol, N. (2005). Students evaluating teachers: Exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education*, 10(1), 57–73.

Morgan, H. K., Purkiss, J. A., Porter, A. C., Lypson, M. L., Santen, S. A., Christner, J. G., Grum, C. M., & Hammoud, M. M., (2016). Student evaluation of faculty physicians: Gender differences in teaching evaluations. *Journal of Women's Health*, 25(5), 453–456.

Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C., Hammout, N., Nakamura, E. L., Escamilla, J., & Hosseini, S. (2020). Impact of student's evaluation of teaching: A text analysis of the teacher's qualities by gender. *International Journal of Educational Technology in Higher Education*, 17(1), 1–27.

Peterson, D. A., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS One*, 14(5), e0216241.

Read, W. J., Rama, D. V., & Raghunandan, K. (2001). The relationship between student evaluations of teaching and faculty evaluations. *Journal of Education for Business*, 76(4), 189–192.

Reid, L. D. (2010). The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.com. *Journal of Diversity in Higher Education*, 3(3), 137.

Schlosser, A. E. (2020). Self-disclosure versus self-presentation on social media. *Current Opinion in Psychology*, 31, 1–6.

Seiter, C. R., & Brophy, N. S. (2022). Social support and aggressive communication on social network sites during the COVID-19 pandemic. *Health Communication*, 37(10), 1295–1304.

Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles*, 53(11), 779–793.

Supiano, B. (2019). One way to take the sting out of student feedback. *Chronicle of Higher Education (teaching newsletter)*.

University of Michigan Center for Research on Learning and Teaching. (n.d.). Course evaluations: Providing helpful feedback to your instructors [Handout]. Retrieved from https://crlt.umich.edu/sites/default/files/resource_files/Course%20Evaluation%20Guidance.pdf?utm_source=CRLT-Engin+Audience&utm_campaign=e284cf16e3-EMAIL_CAMPAIGN_W20_RETURNING_GSI_IA_COPY_01&utm_medium=email&utm_term=0_6ce0b4d408-e284cf16e3-119393090.

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

Wallace, S. L., Lewis, A. K., & Allen, M. D. (2019). The state of the literature on student evaluations of teaching and an exploratory analysis of written comments: Who benefits most? *College Teaching*, 67(1), 1–14.