



A Typology of Perceived Negative Course Evaluations

Heather J. Carmack¹  and Leah E. LeFebvre² 

1. Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA
2. Department of Communication Studies, University of Alabama, Tuscaloosa, AL, USA

ABSTRACT

Instructors and administrators continue to debate the merit and value of using course evaluations to assess instructor effectiveness and course outcomes, especially when students see course evaluations as satisfaction surveys where they can unload negative and/or hurtful comments directed at instructors. Little is known about instructors' perceptions of negative course evaluations. This study qualitatively examined faculty's ($N = 90$) perceptions of negative course evaluation qualitative comments. Using a grounded analyst-constructed typologies approach, three types of negative course evaluation comments were identified: *professional*, *personal*, and *performance*. These types of negative comments call into question the disconnection between what students and instructors perceive as negative comments and how instructors and administrators should evaluate their performance in the classroom. The findings offer recommendations for how instructors and administrators can better navigate how to use negative comments in performance reviews.

KEYWORDS: course evaluations, communication, department administration, faculty, microaggressions

Editor Note—*This article was previously accepted by previous editor. The authors report no conflict of interest.*

CONTACT Heather J. Carmack  • E-mail: heather.carmack1@gmail.com • Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic • 200 First St. SW • Harwick Building Second Floor • Rochester • MN 55905



Copyright 2023 Authors. Published under a Creative Commons Attribution 4.0 International (CC BY-NC-ND 4.0) license.

Introduction

Course evaluations (e.g., student evaluations of teaching, student opinion evaluations, teaching course evaluations, etc.) are a common contemporary metric utilized for assessing the quality of faculty teaching and course outcomes (Allen, 1996; Ramsden, 1991). These evaluations offer opportunities to reflect and improve; however, the process of creating, administering, evaluating, and interpreting is open for debate and rife with tension (Rice et al., 2000). Headlines in *Inside Higher Education*, *The Chronicle of Higher Education*, and *American Association of University Professors* permeate popular higher education outlets highlighting the contention and critique of course evaluations and course assessment practices (e.g., Falkoff, 2018; Flaherty, 2018; Gannon, 2018). The use of student course evaluations is well accepted, entrenched, and researched across the country (Algozzine et al., 2004). The current higher educational climate relates students to customers, and as such the emphasis on evaluation of their experience is based on their perspectives (Tricker et al., 2005). At the heart of the debate is whether students, who are often presented as higher education “customers,” are capable of evaluating instruction and pedagogy, or whether they simply see the course evaluation form as a customer service satisfaction survey (Flaherty, 2015).

Scholars in higher education have long debated the advantages, disadvantages, and controversy surrounding quantitative closed-ended ratings and rankings (see Hornstein, 2017, for a comprehensive review). Debate in the 1980s and 1990s questioned what exactly quantitative course evaluations should evaluate, as well as how they should be used in the performance evaluation of faculty (e.g., Abrami, 1989; Marsh, 1984, 1987; Ramsden, 1991). More recently, Salmon and colleagues (2005) questioned what quantitative evaluations are actually evaluating, especially since the predominantly quantitative evaluations are utilized to rate and rank faculty in terms of instructional effectiveness without explaining what instructional effectiveness is to students. An assessment of 6,000 course evaluations found that quantitative course evaluations function to provide individual instructor effectiveness; however, they do not operate as valid measurements of student-learning outcomes. Instead, students used course evaluations to express their likes and dislikes of an instructor (Salmon et al., 2005). Less than half (43.3%) of students in Kite et al.’s (2015) study said they utilized open-ended comments to provide feedback to instructors because they did not believe that their comments were likely to be read by their instructors. Open-ended questions can enrich the course evaluation responses; however, they often focus on students’ perceived negative review of the course and instructor (Tricker et al., 2005). This is telling, as research on course evaluations tends to focus on quantitative evaluation scores, not the qualitative open-ended responses, and how these are valued by students, and subsequently perceived and interpreted by instructors. A disconnect exists between how students approach course evaluations as a way to evaluate how much they *liked* an instructor and an instructor’s focus on student learning.

Although scholars know students write negative comments in the qualitative section of course evaluations, little is known about the types of perceived qualitative negative comments faculty receive. Gee (2015) called for the rigorous study of qualitative course evaluation comments as a way to examine how students and faculty process comments and use them to make changes to their courses. The purpose of this study is to identify the types of perceived negative qualitative comments faculty report receiving on their course evaluations. First, we examine the literature on course evaluations, exploring factors that impact how students rate their instructors. After explaining the

grounded analyst-construct typology approach to creating a typology of perceived negative course evaluation comments, we present a typology of comments that examines professional, personal, and performance critiques.

Course Evaluations

Course evaluations are a generally integrated and accepted tool to assess institutional accountability in higher education institutions. When used appropriately, course evaluations serve as formative *and* summative tools, with the dual purpose of “improving instructional practice and employment decisions” (Algozzine et al., 2004, p. 135). Although there is no universal standard for course evaluation question design, course evaluations questions should assess (1) student learning and value of information, (2) organization and clarity of material, (3) exams and grading, (4) assignments and readings, (5) workload difficulty, (6) breadth of material covered, (7) instructor enthusiasm, (8) classroom interactions, and (9) student/instructor rapport (Algozzine et al., 2004; Marsh, 1984, 1987). These metrics offer opportunities for multiple stakeholders to assess the quality of education from educators in classrooms, to help make employment decisions, and serve to facilitate discourse inclusive of faculty, review committees, administrators, board of regents, state legislators, alumni, and students (Rice et al., 2000; Smith et al., 2007). Faculty can utilize the results as a feedback tool for reflecting on their instructional practices and educational climate as well as utilize them for improving their teaching skills, adequately meeting course learning objectives and outcomes, and increasing their effectiveness.

Although these metrics purport to evaluate teaching effectiveness, recent research suggests these tools struggle to capture teaching effectiveness, and instead, ask students to rate instructor “likeability” (Hornstein, 2017). The focus on likeability instead asks students to use personal feelings to guide their perceived success of the class. A number of contemporary studies (e.g., Boring et al., 2016; Kornell & Hausman, 2016; Wright & Jenkins-Guarnieri, 2012) found that course evaluations were biased, and in some cases discriminatory, because they focused on likeability, not teaching effectiveness. More recently, El-Alayli et al. (2018) and Mitchell and Martin (2018) reported evidence of gender bias and microaggressions in student feedback. Ultimately, a focus on “likeability” trickles down to qualitative comments, where students are often asked to explain what they “liked” and “did not like” about the course and the instructor.

Halo Effects and Course Evaluations

Students’ evaluations, while completed with students’ best intentions (Kite et al., 2015), are still positively or negatively influenced by factors outside the instructor’s control, such as students’ interest and knowledge in the subject matter, past learning experiences, academic motivation, and student–teacher concordance (Algozzine et al., 2004; Feeley, 2002; Koermer & Petelle, 1991; Kozey & Feeley, 2009). These external factors, or halo effects, are outside forces or perceptions not related to course content that impact students’ perceptions of instructors (Feeley, 2002). There are four sources of halos (Cooper, 1981): engulfing (focusing on salient instructor features, such as

attractiveness), undersampling (evaluating based on limited interactions), insufficient concreteness (unclear course evaluation questions), and insufficient time (not enough time or energy to accurately evaluate an instructor). Halo effects can be positive or negative; instructors can benefit from students' limited exposure to them or limited time to complete course evaluations or they can be marked down if they do not have desirable salient features.

These halo effects, especially engulfing and undersampling, can be driving forces for how they evaluate instructors at the end of a course. Students who enter a course with high expectations and positive previous experiences with an instructor (engulfing) rate the instructor more favorably than students with lower expectations and negative past experiences, regardless of how students perceived the current course they had with that instructor (Koermer & Petelle, 1991). When asked about their view of course evaluations in the teaching evaluation process, students indicated they take the evaluation process seriously (Kite et al., 2015). Kite and colleagues found that students have a positive course evaluations viewpoint, and most importantly, that almost all students (95%) indicated they honestly assessed their instructor's teaching ability. However, a semester-long course may also not be long enough for students to effectively evaluate faculty, leading to undersampling. Additionally, students who were at least one semester removed from a course were more discerning and critical of instructional quality (Kozey & Feeley, 2009).

Course evaluations should highlight instructors' strengths and weaknesses; however, students may use extraneous features, such as attractiveness and vocal delivery, or halos to evaluate their learning in a course. Not surprisingly, faculty members are also evaluated on personal factors, such as attractiveness, gender differences, nationality, and race (e.g., Miller & Pearson, 2013; Salmon et al., 2005). For students, organization, enjoyableness, and intimacy (e.g., immediacy) with an instructor are more important than course difficulty and material utility (Clevenger & Todd-Macillas, 1981); the determination of an instructor's success in the classroom is their showmanship, not their mastery of the material. Often traditional course evaluations invite criticism (Tricker et al., 2005), but most research focuses on quantitative scores, not qualitative comments from students (Baker & Copp, 1997). So, what happens when faculty do not make the course enjoyable? Or when students find something unattractive about an instructor? Beyond the quantitative course evaluation scores, what kinds of negative comments do faculty believe students write on course evaluations? The following research question guided this study: *What types of perceived negative messages do faculty report receiving on their course evaluations?*

Materials and Methods

Participants

Ninety instructors completed an online qualitative questionnaire for this study. A majority of participants identified as female ($n = 81$, 90%; males, $n = 9$, 10%). The sample identified as Caucasian ($n = 68$, 75.6%), Hispanic ($n = 6$, 6.7%), Asian/Pacific Islander ($n = 5$, 5.6%), Black ($n = 3$, 3.3%), multiracial ($n = 3$, 3.3%), other ($n = 3$, 3.3%), and unidentified ($n = 2$, 2.2%). Participants' ages ranged from 23 to 71 years old ($M = 38.16$; $SD = 9.48$). Participants were native English speakers ($n = 82$, 91.1%) and English as their second language ($n = 8$, 8.9%).

Instructors' education level varied, with most participants having their doctorate ($n = 74$, 82.2%) or master's degrees ($n = 14$, 4.4%). Two participants reported having their baccalaureate ($n = 1$) or a different degree ($n = 1$). Instructors had a varied number of years of teaching experience, ranging from 1 to 35 years ($M = 10.80$ years, $SD = 7.19$). A variety of disciplines were represented including Communication ($n = 22$, 24.4%), Experimental Psychology ($n = 3$, 3.3%), Psychology ($n = 38$, 42.2%), and Social Psychology ($n = 7$, 7.8%). Only one participant represented the remaining disciplines that encompassed humanities, hard science, and social sciences ($n = 20$, 22.2%). Most participants were tenure track ($n = 36$, 40%) or tenured professors ($n = 23$, 25.6%), while others were in non-tenured positions or other positions ($n = 12$, 13.3%), such as administration or nonprofit work ($n = 5$, 5.6%). Fourteen participants were master ($n = 1$) or doctoral ($n = 13$, 14.4%) students.

Data Collection and Coding

Data collection began after the authors received Institutional Review Board (IRB) approval. Participants completed an online qualitative questionnaire via Qualtrics in spring 2018. Instructors were invited to participate via communication and psychology list-servs (i.e., COMMnotes and Society of Personal and Social Psychology) as well as through snowball sampling. Individuals who wished to participate agreed to the informed consent and then proceeded to the questionnaire. The questionnaire consisted of 25 questions about negative course evaluations; however, this article reports on the findings related to one question: "What kinds of negative messages have you received on your course evaluations?" This question asked participants to identify what they perceived to be negative comments from students. Some participants provided actual comments from their course evaluations, while others summarized.

The authors engaged in a grounded analyst-constructed typology analysis (Patton, 2015), where a typology, or an organized system of types (Collier et al., 2012), emerged out of the data. Similar to grounded theory, the grounded analyst-constructed typology approach involves identifying patterns and categories that emerge from the data without the guidance of a theory or a priori model (Kluge, 2000). Grounded typology approaches are becoming a more popular method in qualitative research (e.g., Buehler et al., 2019; Caiata-Zufferey & Schulz, 2012).

The authors conducted an iterative thematic analysis to identify commonalities and shared meaning (Braun & Clarke, 2006). The unit of analysis was an individual negative comment. In some cases, participants wrote multiple comments, so each comment was coded as an individual unit. In total, 277 units were analyzed. The authors began by reading each response to get a holistic understanding of the data. The first author and a graduate research assistant separately inductively coded 20 random comments, developing an initial coding framework based on in-vivo codes and sensitizing concepts (Patton, 2015). These codes were refined and used to create a codebook, including the code name, a brief explanation of the code, and illustrative examples of participant comments for each code. Once the codebook was completed, the authors worked with the graduate research assistant to individually code 40 participant comments, identifying the number of units of analysis in each participant comment and categorizing units. The coding team met to discuss the initial coding, discussing any differences. Based on initial coding, the codebook was refined, and two codes were added (see Table 1). Using the refined codebook, the authors collectively coded the entire dataset, identifying 14 distinct types. The authors constructed clear conceptualizations and

dimensions to each type. The authors grouped the 14 types into 4 major types and 12 subtypes. Comments are presented as written by participants, including typos, profanity, and punctuation.

Results

On average, participants identified three ($M = 3.26$, $SD = 1.61$) major perceived negative comments in their individual answers. Four types of perceived negative course evaluation comments emerged: *professional*, *personal*, *performance*, and *miscellaneous*.

TABLE 1 *Thematic Categories of Negative Course Evaluations*

Thematic Categories	Subcategories	Definitions	%	<i>n</i>
Professional		Comments related to expectations of how an instructor should teach and evaluate student work	68.6%	190
	Job Incompetence	The perceived lack of instructor excellence in job requirements	32.5%	90
	Toughness	The perceived extreme rigor and expectations outside of the students' standards	11.9%	33
	Classroom Power Dynamics	Student frustration of expected equality in the classroom	7.9%	22
	Excessive	Students' perceived superabundance of workload and requirements	6.9%	19
	(Un)fairness	Students' perceived unreasonable enforcement of course policies	4.7%	13
	Unmet Learning Expectations	Students' disappointment or dissatisfaction with the course outcome	4.7%	13
Personal		An attack toward an instructor based on discriminatory stereotypes	17.3%	48
	General Ad Hominem	Attacks directed at the instructor designed to hurt	10.8%	30
	Political/ Religious Attacks	Attacks based on perceived political and/or religious differences	2.9%	8
	Race/ Ethnicity Attacks	Attacks based on perceived race and/or ethnicity differences	2.2%	6
	Gender Attacks	Attacks based on non-confirmatory of gender roles	1.4%	4
Performance		Physical characteristics related to instructors' performance of the instructor role	11.9%	33
	Delivery of Material	Criticism of instructor vocal delivery and presence in the classroom	7.6%	21
	Physical Appearance/ Sexual Comments	Comments meant to sexualize or demean	4.3%	12
Miscellaneous			2.2%	6

Professional

The most prominent perceived negative course evaluation centered on professional critiques ($n = 190, 68.6\%$). Professional critiques focused on student comments related to expectations of how an instructor should teach and evaluate student work. Professional critiques included job incompetence, toughness, classroom power dynamics, excessive, (un)fairness, and unmet learning expectations.

Job Incompetence

This subcategory highlighted the perceived lack of instructor excellence in job requirements ($n = 90, 32.5\%$). Comments centered around two different areas: instructors need to successfully teach material and manage the classroom *and* do it in an inspiring and entertaining way. Comments related to teaching and managing the classroom focused on how an instructor was not successful in those areas or teaching in general. Student comments were often focused on instructors being “unorganized, unprepared, [having] poor class management.” As one participant stated, “They mostly suggest that I’m inexperienced, not prepared, or uninformed.” Student comments clearly communicated to instructors that they expect them to know the material well and to be able to communicate that to the students.

Sometimes lack of clarity on topics was based on the student’s lack of understanding, as one participant experienced:

I taught political psychology and in my evaluation one student said I didn’t know what political psychology was because we never talked about the mental health of political leaders, meanwhile most of the readings were from political psychology texts or the journal titled political psychology.

Additionally, student comments were generally specific in identifying what areas the instructor was lacking, such as “quality of slides,” “slow in grading,” “A/V problems with the classroom were treated as my personal failing,” and “handouts are not useful.” In some cases, however, participants just received blanket comments about their lack of fit and suggested they should never teach again, or as one participant was bluntly told, “They should have hired the other person who interviewed.”

Participants also commented that students said they felt bored or uninspired by the instructor. Comments focused on the phenomena that education should be fun, engaging, and inspiring and these instructors were the antithesis. Incompetence was conceptualized as not just knowing the material, but also presenting it in a dynamic and interesting way to keep students’ attention. For instance, many participants’ comments included statements about being bored or that “the class is boring.” Others specifically noted that their instructors were “not engaged or enthusiastic” or they were “uninspiring.” One participant shared a direct student comment: “Nothing in this class was engaging because [first name] sucks.” Other comments expanded beyond the classroom and instructor and noted their general demeanor for their profession. One participant received a comment that stated, “She rarely seemed excited about her discipline. When asked questions, she often responded with I’m unsure, you should research that. I appreciate the transparency, but it’s obnoxious when the professor can’t answer a question.”

Toughness

This subcategory showcased the perceived extreme rigor and expectations outside of students' standards ($n = 33$, 11.9%). This subcategory was composed of student comments about instructors being "too hard," "too harsh," "cruel," or "strict" in their grading practices. Students' course evaluation comments communicated a perceived lack of flexibility and too high of expectations for what it takes to get an A in the course, as one participant explained, "Complaints about harshness of the grading, about my strictness with enforcing course policies, about the difficulty of the exams, about the difficulty of the class relative to other sections or other sections in previous terms." For many instructors, their perceived "toughness" was related to unfair or excessive workload practices, as it was for this participant: "Most negative comments center on the difficulty of the classes, the perceived high workload, and my inflexibility in exceptions to class policies (due dates, missed assignments, etc.)." Many student comments to instructors were vague about why they perceived the instructors to be too tough. In some cases, however, students used what they believed to be the standards of upper-division or graduate level work to point out the discrepancy, as noted by this participant who shared a direct quotation from a course evaluation:

This class was way to hard . . . They don't prepare you for the final and essays. They grade essays way to hard as if we are in some graduate program. It feels like you can't succeed in this class. By far the worst class Ive ever taken. So much wasted time spent in the class for the poor grade I got.

Additionally, several comments focused on the other extreme and commented on how instructors were not tough enough. For instance, "They generally regard [me as] not being 'harsh enough.'" Overwhelmingly instructors interpreted students' perceived self- or peer-determined expectations for what they thought the course and subsequent workload should consist of, and that the instructor(s) knowingly and purposively opted to impose and inflict harsh standards on students.

Classroom Power Dynamics

This subcategory exemplified perceived student frustration of expected equality in the classroom ($n = 22$, 7.9%) and focused on instructors' desire to maintain a power difference in the classroom, which was perceived as a negative by students. "Condensing" was common for instructors' comments in this subcategory. For example, as one participant stated, "My use of colloquialisms like 'kiddos' (when referring to children, not the students) make the class feel 'less professional.'" Instructors' language choice drove the communication climate in the classroom and the "wrong" word choices could create tension and, ultimately, decrease an instructor's credibility in the classroom. For some participants, this critique often resulted in a power dispute, as one participant explained,

Comments about me talking like a kindergarten teaching [sic], and a comment that claimed I raged at students regularly in class and didn't ever let anyone ask questions (I think the student must have had me confused with a different and very colourful prof; the evaluation was done online so there was no way to know if it was genuine).

For this participant, even though she believed it was a faulty evaluation, the student believed the instructor was creating a learning space where the instructor belittled and talked down to students. Additionally, comments reflected inequality in demonstrating respect. For instance, one instructor noted that, “Students don’t want to call me by my title.” Participants often commented on students’ perceived power imbalance within the larger classroom as well as perceived inconsistency amongst certain students (demonstrated as a distribution); on the other hand, power imbalances were also noted by instructors who received student indifference or disrespect.

Excessive

This subcategory emphasized perceived superabundance of workload and requirements ($n = 19$, 6.9%). Instructors received comments from students who perceived that the instructors were asking them to do too much work, usually grounded in relation to other classes or what the students believed was an appropriate amount of work. For example, one participant received “ones about course content such as ‘too much writing.’” Other comments involved superlatives such as “the hardest class I’ve taken,” and “the most stressful class I’ve taken.” Often comments articulated that the course was, “too demanding,” “tests are too long,” “too critical,” and “too much work.” One participant even received negative feedback about *helping* students too much: “negative as to too much time spent informing students of student success opportunities and strategies available to students.”

(Un)fairness

This subcategory reflected perception of unreasonable enforcement of course policies ($n = 13$, 4.7%). Instructors reported that students provided vague statements surrounding the policies and procedures inherent within the course structure, often and simply stating “unfair.” Other instructors noted more specifics (although still quite imprecise or general). For instance, “Complaints about changing the syllabus (we didn’t)” and “Complaints about not knowing when things were due (they received a schedule)” and “they don’t like attendance policy.” In some cases, participants believed they were accused of being biased against specific students, usually by “playing favorites,” grading students differently than others, or being “biased against me [the student] and my perspective.” Participants noted that students identified what they found to be unfair, such as an exam or attendance policy, but did not specify what was unfair or how to fix this perceived imbalance or unfairness.

Unmet Learning Expectations

Instructors reported receiving comments that illustrated students’ disappointment or dissatisfaction with the course outcome ($n = 13$, 4.7%). This subcategory is clearly exemplified by the statement, “I didn’t learn anything,” or “This class was the worst I’ve ever taken.” These types of comments suggest that instructors believed students had specific, but unstated expectations for the class—what they should learn, what they should get out of the class—that were not met by the instructor. Some student comments made sure to attack the instructor for their lack of learning, “I learned nothing from this instructor.” Such a critique places the onus of learning on the instructor, not the student, so when learning does not happen, it is not the student’s fault. This can extend beyond one class, as one participant was told, “I learned nothing and he made me hate statistics.”

Personal

This second thematic category was exemplified as perceived attack(s) toward an instructor based on discriminatory stereotypes ($n = 48, 17.3\%$). This category often identified course evaluation comments that did not focus on the course content. Personal critiques included attacks—general ad hominem, political and/or religious, race and/or ethnicity, and gender-based.

General Ad Hominem

This subcategory focused on perceived attacks directed at the instructor designed to hurt ($n = 30, 10.8\%$). Comments in this category were opportunities for students to name call professors, such as “jerk,” “rude,” have a “nasty attitude,” are “mean,” and a “bitch.” One participant said she received “Very personal ones, like ‘emotionally unstable,’ ‘teaches a lot about low self-esteem because that describes herself.’” Some instructors suggested these comments used ad hominem statements to attack an instructor’s credibility, such as “she acts like she knows more than she does.” Another participant explained how students made sure to focus on “how stupid and worthless I am.” Instructors believed these comments were meant to degrade instructors on a personal level and did not provide feedback for reflection of the course or improve the learning environment.

Political/Religious Attacks

This subcategory focused on perceived attacks based on political and/or religious differences ($n = 8, 2.9\%$). In this category, instructors believed they were blamed for exposing too much of their personal views about politics and religion. For example, one participant noted, “During the election, one student decided ‘I was unprofessional because I had a political sticker on my (personally owned) laptop cover that I use in class.’” Instructors were “accus[ed] of being liberal” or progressive, which was seen as problematic for students. Other comments centered around religion or faith. One participant noted that students commented on the lack of faith: “I have also been criticized for not incorporating my faith in the classroom (I teach at a faith-based institution).” In this case, the participant did not share *enough* with her students, and that was perceived as a criticism.

Race/Ethnicity Attacks

This subcategory focused on perceived attacks based on race and/or ethnicity differences ($n = 6, 2.2\%$). Instructors interpreted comments related to either their race or ethnic differences from students as negative comments. One participant explained, “Comments about the fact that I ‘like Hispanic and Black students better’ and other things related to my identity as a Latina and to my course content centering [on] women of color voices.” In this case, the instructor thought students used the instructor’s race against her as a reason for not privileging certain voices. Likewise, another participant was told she “should be happy to be experiencing less racism in California than Texas.” Other participants in this subcategory reported what they believed to be ethnicity attacks, such as “I sound like I just rolled off a boat” and “foreign accent, sometimes difficult to understand.” Additionally, participants received comments that they interpreted as students’ belief they were anti-American. For instance, “Lots of anti-American stuff (I’m an American in Canada and this was during Bush II).” All these comments focused on differences as problematic instead of as a learning opportunity and enriching the classroom environment.

Gender Attacks

This subcategory focused on perceived attacks based on non-confirmatory gender roles ($n = 4$, 1.4%). In this case, instructors thought students used instructors' gender as a reason for their displeasure with the course, usually because women professors did not perform the expected gender roles. This was illustrated by this participant: "That I'm too tough/strict, that I'm mean (I interpret this as backlash to a woman who doesn't read as warm/supportive/effusive)." Instructors also noted derogatory language specifically directed at women, such as "Screaming bitch." Additionally, another participant directly noted that she perceived her negative evaluation in comparison to her men colleagues, who would not receive the same negative evaluations: "A tendency to criticism [sic] the fact that I am not sensitive or feeling or caring enough (criticism I have noticed is not stated of my male colleagues who do less advising and mentoring for students than I do)." Instructors interpreted these perceived gendered attacks as a way to undermine women instructors.

Performance

This thematic category focuses on physical characteristics related to instructors' performance of the instructor role, including delivery and appearance ($n = 33$, 11.9%). Two subcategories comprised this theme: delivery of material and physical appearance/sexual comments.

Delivery of Material

This subcategory focused on the perceived criticism of instructor vocal delivery and presence in the classroom ($n = 21$, 7.6%). Instructor comments in this subcategory included issues related to speed, cadence, and sense of humor in speaking; one participant reported that a student said they did not like the instructor's laugh. One participant indicated that "Some of the negative comments are that my voice 'sounds annoying' when I use the microphone." Other participants noted students' comments about "dribbling" or "babbling on" in class, indicating that professors who do this "don't know what they are saying." Other instructors reported students commenting on perceived confidence exhibited by the instructor. One participant noted, "A lot of people call me timid or shy (which isn't necessarily a bad thing but I don't feel like [it] is an accurate representation)." This subcategory focused less on the content and more on the performance as ascribed by perceived students' expectations for what the instructor should sound like.

Physical Appearance/Sexual Comments

This subcategory focused on comments instructors believed were meant to sexualize or demean ($n = 12$, 4.3%); this was illustrated by this participant's comment, "That the course would be better if I took my top off and other sexual things." These comments highlighted professors' appearance, weight, and attire and objectified the instructor. Another participant said, "When I was younger most of evaluations were centered on my appearance (I was called 'hot' and students left me their phone numbers)." Participants who provided these comments believed that students may have meant to be flattering, but actually served to devalue the instructors' abilities and classroom credibility.

Discussion

Course evaluations can be effective tools for assessing instructors' classroom ability, yet they can also be used by students to criticize faculty. In this study, faculty reported perceived negative qualitative comments, focused on their professional (in)ability, personal attacks, and problematic performance. Comments ranging from job incompetence to sexual appraisals were all interpreted by faculty as negative comments, even if students may not have meant them as negative. The perceived negative comments communicated to faculty that students were comfortable writing hurtful and offensive comments on course evaluations, whether or not they relate to course outcomes. Instructors believed students thought these criticisms negatively impacted their learning environment, and in some cases, continued interest in the topic.

The most prominent thematic category centered on perceived professional critiques. The premiere subcategory emphasized job incompetence and how instructors do not satisfy their perceived job requirements. Instructors' interpretations of course comments showcased the multifaceted expectations surrounding job (in)competence. Instructors needed to simultaneously fulfill transmission of content, *and* also be inspiring and entertaining. This teaching expectation includes a performative element as an important element of education quality, which students consider to be the most significant source of student satisfaction (Ginns et al., 2007). Instructors conflated these issues in their qualitative comments surmising that students thought the instructor and/or course was boring or uninspiring. This calls into question a tension between satisfaction and effectiveness. Are students intertwining inspiring and boring, and if so, how does that influence the instructor role and educational climate? How does the operationalization of education innately become inclusive of performance?

Students can easily communicate about their experiences in a course, including audibility or clarity of the instructor, legibility or articulation of ideas, notes, or assignments, availability of the instructor, or opportunities for outside resources and support (Becker et al., 2012); however, these instances do not substantiate the ability to evaluate course pedagogy or content outside or beyond their experience and expertise (Hornstein, 2017). Should instructors conform to perspectives that are deemed normative or popular by students? Both the perception of job incompetence and the second subcategory, toughness, show how students believe their instructors exceed some normalized or standardized level of rigor students expected. Students may have disliked having to expend effort (Braga et al., 2014) so perceived toughness (i.e., "too hard") was framed as a negative and as evidence of job incompetence. A disconnect emerges between students and instructors: faculty may view being tough as a badge of honor, whereas students may view toughness as detrimental to learning. Faculty, then, have the added job of showing students that rigor is beneficial to learning.

Criticisms related to professional performance, including being boring or unorganized, were mostly directed toward content. Although perceived as negative, these types of comments can be addressed by instructors, even if they are vague. Perceived personal attacks, especially those focused on gender and race, cannot be addressed because they are directly connected to an instructor's identity. The fact that the second highest category was related to perceived personal attacks is telling. Why do students feel it is appropriate to comment on an instructor's race or gender? What is gained from telling a faculty member she or he is a jerk or a screaming bitch? This calls into question what students are trying to accomplish by including these comments in their evaluations.

If, as the participants in Kite et al.'s (2015) study stated, students take course evaluations seriously, then do they truly believe these things about instructors? A focus on personal attacks highlights the presence of engulfment in course evaluations; students are focusing on extraneous features rather than on content (Cooper, 1981; Feeley, 2002). Even if students believe they are being honest in their evaluation, their use of personal attacks may speak to the argument that they see course evaluations as satisfaction surveys. They are not concerned with what they learned, but rather with whether they liked the instructor.

The perceived personal attacks also underscored how faculty must be careful about how and what they disclose about their personal lives or identities, especially related to politics and religion. Faculty shared comments where students condemned faculty for both disclosing too much (related to politics) and not enough (related to religion). Past research showed that students appreciate when faculty disclose personal information (e.g., Downs et al., 1988); it makes them more personable and approachable, which can be direct counters to professional and performance critiques. However, as the instructors' comments show, there is a fine line for this disclosure. Part of the problem may be that students do not perceive these self-disclosures as related to course content, and thus believe the disclosure is inappropriate (Downs et al., 1988). For students, too much disclosure that differs from their own beliefs could be perceived as creating a negative communication climate (Kearney et al., 1991).

Finally, the perceived negative personal and performance comments call attention to the role of microaggressions in course evaluations. Although a smaller number were explicit in their perceived gender and race bias, many of the comments were implicit (e.g., annoying voice or weird laugh). The perceived explicit sexist and racist comments are certainly problematic—saying a class would be better if an instructor took her top off is never appropriate—but the implicit comments are more insidious because they attempt to erode instructors' credibility without directly saying it is because of bias. For example, the participant whose students refused to call her “Dr.” engaged in microaggression (Boysen, 2012). It is possible students do not realize their comments are coded as unintentional bias (Dovidio et al., 2002); however, these perceived implicit comments show how students are negatively influenced by this halo effect. In this case, microaggressions become a type of engulfing, where students may be unable to look past the vocal delivery or clothing of an instructor and evaluate their ability to teach.

Applying the Typology

The findings demand we consider the substance of the qualitative course evaluations. Instructors continue to face and consider the perceived negative course evaluations comments; however, these struggles, difficulties, and personal attacks are not constrained to the instructor. These negative course evaluations are retained and shared, or even continually seen as instructors are considered for retention, promotion, tenure, and/or merit. Instructors can use this typology to analyze their own course evaluation comments, using the categories to frame a narrative explanation of why comments about appearance, race, gender, and other personal definition characteristics are problematic and should not be included in their annual evaluations and promotion and tenure applications. Departmental and college administrators, such as department/unit heads and deans, can also use the typology to review negative comments with faculty, focusing on negative comments

specific to improving their teaching effectiveness. Heads and deans can be effective support systems (LeFebvre et al., 2020) for faculty attempting to make sense of negative and hurtful course evaluations, and the typology can be used to help frame those conversations.

Finally, institutions serious about changing the ways course evaluations can be used can use this typology to identify categories of comments that could be censored from evaluations. Administrators need to weigh the ability to determine the substance of negative course evaluations. Particularly, negative courses, such as personal or performance categories, whereby the comments are not grounded in means to benefit the instructional or pedagogical practices, rather only indicated to hurt, harm, or inflict pain. Recently, we were contacted by a representative from a mid-sized university interested in using this typology to create a tool to remove inappropriate, hurtful, and discriminatory comments from their evaluations. This study adds to the contemporary and contentious conversation of evaluations. Other task forces working toward creating these kinds of tools can also use the typology to craft workable metrics for determining what comments can help faculty improve their teaching. If qualitative course evaluation comments accompany quantitative course evaluations, educators must first begin with understanding the types of comments, and then how to re-evaluate their functionality, purpose, and outcomes.

Limitations and Future Directions

Several limitations exist with this study. First, the participants were disproportionately Caucasian women. Many of the perceived negative comments were explicitly or implicitly gendered in their criticism; however, there was a noticeable lack of race-centered comments. A more balanced gender breakdown, especially with the inclusion of more men, BIPOC faculty, and transgendered participants, could result in different comments. Tied to this, the researchers did not ask about the location of instructors' colleges or universities. The researchers relied on two primarily United States (U.S.)-based professional organizations to recruit; however, it is possible that participants hailed from institutions outside the U.S. Furthermore, this study asked instructors to provide their negative course evaluations. Many responses did include direct quotations from course evaluations; however, other participants only relayed the comments in their own words, which means the instructors' recall may misrepresent comments. Additionally, instructors were not asked about the recency or repetition of these comments and how the impact of repeated exposure to perceived negative comments may influence course evaluations or recall.

Based on the findings, there are a number of future directions for researchers interested in examining negative course evaluations. Our previous research explored how instructors make sense of these negative comments (Carmack & LeFebvre, 2019) and the support messages that help them deal with hurtful comments (LeFebvre et al., 2020). However, there is still more to explore. How do these negative comments impact how instructors approach future teaching and students? Finally, this study focuses on one voice in the classroom: *the instructor*. Students are the ones writing comments. Do students understand what happens when they write these comments? Do they see course evaluations as learning opportunities or do they see them akin to customer service reviews? Although faculty members perceive comments as negative, how do students perceive them? More research is needed from students' perspective to paint a more holistic picture of the course evaluation experience.

Course evaluations offer the opportunity to reflect on and critique course outcomes, instructors, and instruction. All too often, qualitative evaluation comments from students include negative, and subsequently, hurtful commentary that remains salient in instructors' memories. The negative qualitative course evaluations sometimes focus on usable content, but also include personal and performance characteristics beyond instructors' control and instructional purview. As such, agencies and agents in evaluation processes must become more mindful of how these negative comments and cumulative impacts can influence instructors and education. Hornstein (2017) cautioned faculty and administrators to not take course evaluations too seriously, noting that they are "measures of popularity and liking" rather than measures of ability (p. 4). Although students have the ability to express that they "honestly didn't learn shit," instructors do not have the same luxury to say the same about their course evaluations. Instead, they must find ways to reflect on and find ways to see merit beyond the negativity.

References

- Abrami, P. C. (1989). How should we use student ratings to evaluate teaching? *Research in Higher Education*, 30(2), 221–227. <https://www.jstor.org/stable/40195895>
- Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F., Mohanty, G., & Bray, M. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching*, 52(4), 134–141. <https://doi.org/10.3200/CTCH.52.4.134-141>
- Allen, M. (1996). Research productivity and positive teaching evaluations: Examining the relationship using meta-analysis. *Journal of the Association of Communication Administration*, 2(1), 77–95.
- Baker, P., & Copp, M. (1997). Gender matters most: The interaction of gender expectations, feminist course content, and pregnancy in student course evaluations. *Teaching Sociology*, 25(1), 29–43. <https://doi.org/10.2307/1319109>
- Becker, W. E., Bosshardt, W., & Watts, M. (2012). How departments of economics evaluate teaching. *Journal of Economic Education*, 43(3), 325–333. <https://www.jstor.org/stable/117133>
- Boring, A., Ottoboni, K., & Start, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research* [electronic source], 1–11. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1>
- Boysen, G. A. (2012). Teacher and student perceptions of microaggressions in college classrooms. *College Teaching*, 60(3), 122–129. <https://doi.org/10.1080/87567555.2012.654831>
- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41(1), 71–88. <https://doi.org/10.1016/j.econedurev.2014.04.002>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Buehler, E. M., Crowley, J. L., Peterson, A. M., & High, A. C. (2019). Broadcasting for help: A typology of support-seeking strategies on Facebook. *New Media & Society*, 21(11–12), 2566–2588. <https://doi.org/10.1177/1461444819853821>

- Caiata-Zufferey, M., & Schulz, P. J. (2012). Physicians' communicative strategies in interacting with internet-informed patients: Results for a qualitative study. *Health Communication, 27*(8), 738–749. <https://doi.org/10.1080/10410236.2011.636478>
- Carmack, H. J., & LeFebvre, L. E. (2019). "Walking on eggshells": Traversing the emotional and meaning making processes surrounding hurtful course evaluations. *Communication Education, 68*(3), 350–370. <https://doi.org/10.1080/03634523.2019.1608366>
- Clevenger, T., & Todd-Macillas, W. R. (1981). Student evaluations of communication courses. *Communication Quarterly, 30*(1), 48–55. <https://doi.org/10.1080/01463378209369428>
- Collier, D., LaPorte, J., & Seawright, J. (2012). Putting typologies to work: Concept formation, measurement, and analytic rigor. *Political Research Quarterly, 65*(1), 217–232. <https://doi.org/10.1177/1065912912437162>
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*(2), 218–244. <https://doi.org/10.1037/0033-2909.90.2.218>
- Dovidio, J. F., Gaertner, S. L., Kawakami, K., & Hodson, G. (2002). Why can't we just get along?: Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology, 8*(2), 88–102. <https://doi.org/10.1037/1099-9809.8.2.88>
- Downs, V. C., Javidi, M. M., & Nussbaum, J. F. (1988). An analysis of teachers' verbal communication within the classroom: Use of humor, self-disclosure, and narratives. *Communication Education, 37*(2), 127–141. <https://doi.org/10.1080/03634528809378710>
- El-Alayli, A., Hansen-Brown, A. A., & Ceynar, M. (2018). Dancing backwards in high heels: Female professors experience more work demands and special favor requests, particularly from academically entitled students. *Sex Roles, 79*(3–4), 136–150. <https://doi.org/10.1007/s11199-017-0872-6>
- Falkoff, M. (2018). Why we must stop relying on student ratings for teaching. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/why-we-must-stop-relying-on-student-ratings-of-teaching/>
- Feeley, T. H. (2002). Evidence of halo effects in student evaluations of communication instruction. *Communication Education, 51*(3), 225–236. <https://doi.org/10.1080/03634520216519>
- Flaherty, C. (2015, June 10). Flawed evaluations. *Inside Higher Ed*. <https://www.insidehighered.com/news/2015/06/10/aaup-committee-survey-data-raise-questions-effectiveness-student-teaching>
- Flaherty, C. (2018, May 22). Teaching eval shake-up. *Inside Higher Ed*. <https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story>
- Gannon, K. (2018). In defense (sort of) student evaluations of teaching. *The Chronicle of Higher Education*. <https://www.chronicle.com>
- Gee, N. (2015). A study of student completion strategies in a Likert-type course evaluation survey. *Journal of Further and Higher Education, 41*(3), 340–350. <https://doi.org/10.1080/0309877X.2015.1100717>
- Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education, 32*(5), 603–615. <https://doi.org/10.1080/03075070701573773>

- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1). [online]. <https://doi.org/10.1080/2331186X.2017.1304016>
- Kearney, P., Plax, T. G., Hays, E. R., & Ivey, M. J. (1991). College teacher misbehaviors: What students don't like about what teachers say and do. *Communication Quarterly*, 39(4), 309–324. <https://doi.org/10.1080/01463379109369808>
- Kite, M. E., Subedi, P. C., & Bryant-Less, K. B. (2015). Students' perceptions of the teaching evaluation process. *Teaching of Psychology*, 42(4), 307–314. <https://doi.org/10.1177/0098628315603062>
- Kluge, S. (2000). Empirically grounded construction of types and typologies in qualitative social research. *Forum: Qualitative Social Research*, 1(1) [online]. <http://www.qualitative-research.net/index.php/fqs/article/view/1124>
- Koermer, C. D., & Petelle, J. L. (1991). Expectancy violation and student rating of instruction. *Communication Quarterly*, 39(4), 341–350. <https://doi.org/10.1080/01463379109369810>
- Kornell, N., & Hausman, H. (2016). Do the best teachers get the best ratings? *Frontiers in Psychology*, 7 [electronic source]. <https://doi.org/10.3389/fpsyg.2016.00570>
- Kozey, R. S., & Feeley, T. H. (2009). Comparing current and former student evaluations of course and instructor quality. *Communication Research Reports*, 26(2), 158–166. <https://doi.org/10.1080/08824090902861630>
- LeFebvre, L. E., Carmack, H. J., & Pederson, J. R. (2020). “It's only one negative comment”: Women instructors' perceptions of (un)hurtful support messages following hurtful course evaluations. *Communication Education*, 69(1), 19–47. <https://doi.org/10.1080/03634523.2019.1672879>
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707–754. <https://doi.org/10.1037/0022-0663.76.5.707>
- Marsh, H. W. (1987). Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253–388. [https://doi.org/10.1016/0883-0355\(87\)90001-2](https://doi.org/10.1016/0883-0355(87)90001-2)
- Miller, A., & Pearson, J. (2013). Can I talk to you: The effects of instructor position, nationality, and teaching style on students' perceived willingness to communicate and on teacher evaluations. *Communication Quarterly*, 61(1), 18–34. <https://doi.org/10.1080/01463373.2012.719059>
- Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648–652. <https://doi.org/10.1017/S104909651800001X>
- Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, 16(2), 129–150. <https://doi.org/10.1080/03075079112331382944>
- Rice, R. E., Stewart, L. P., & Hujber, M. (2000). Extending the domain of instructional effectiveness assessment in student evaluations of communication courses. *Communication Education*, 49(3), 253–266. <https://doi.org/10.1080/03634520009379214>
- Salmon, C. T., Smith, S. W., Lee, B., & Miller, V. D. (2005). Exploring assorted beliefs, doubts, and legends regarding quantitative teaching evaluations. *Journalism & Mass Communication Educator*, 60(3), 257–271. <https://doi.org/10.1177/107769580506000305>

- Smith, S. W., Yoo, J. H., Farr, A. C., Salmon, C. T., & Miller, V. D. (2007). The influence of student sex and instructor sex on student ratings of instructors: Results from a College of Communication. *Women's Studies in Communication*, 30(1), 64–77. <https://doi.org/10.1080/07491409.2007.10162505>
- Tricker, T., Rangelcroft, M., & Long, P. (2005). Bridging the gap: An alternative for course evaluations. *Open Learning*, 20(2), 185–192. <https://doi.org/10.1080/02680510500094249>
- Wright, S. L., & Jenkins-Guarnieri, M. A. (2012). Student evaluations of teaching: Combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education*, 37(6), 683–699. <https://doi.org/10.1080/02602938.2011.563279>