



## CHAPTER 4

---

# Equity and Efficacy in Teaching Effectiveness Assessment (TEA)

*Brad Erickson and Wei Ming Dariotis*

## INTRODUCTION

### *Why Change TEA? Why Now?*

Since the onset of the COVID-19 pandemic, educators across the world have radically changed both how and what they teach. Shifting to emergency remote teaching modalities required more than grafting in-person teaching techniques into online environments; teachers had to completely

---

We would like to thank past and current members of the SFSU TEA Task Force: Roberta D'Alois, Arlene Daus-Magbual, Catriona Esquibel, Carrie Holschuh, Carleen Mandolfo, Zahira Merchant, Juliana Van Olphen, David Olsher, Gilberto Ramirez, Gitanjali Shahani, and Jonathan Stillman; SFSU Academic Senate Chair Teddy Albiniak and Academic Senate executive committee members Gilda Bloom, Marie Drennan, Nancy Gerber, and Jackson Wilson. Special thanks to Maggie Beers, SFSU's AVP of Teaching and Learning.

---

B. Erickson (✉) • W. M. Dariotis  
San Francisco State University, San Francisco, CA, USA  
e-mail: [darototis@sfsu.edu](mailto:darototis@sfsu.edu)

redesign activities to support student achievement of learning outcomes within new modalities and evolving high-stress contexts. At the same time, the inequities laid bare during the pandemic called for a shift to redress the *educational debt* owed to students subject to generations of cumulative educational disinvestment and oppression (Ladson-Billings, 2005); related calls to *decolonize* and redress racism within academic disciplines also have affected how and what instructors teach (Harrison, 1991). We also recognize that there has been a transformation of the ways we teach through changes to teaching and learning objectives; new knowledge about and developmental support for pedagogy; the integration of justice, equity, diversity, and inclusion (JEDI) principles in instruction and evaluation; and the radical shift in teaching conditions since the onset of the COVID-19 pandemic (Pokhre & Chhetri, 2021; Holme, 2020; Safir & Dugan, 2021; Thomas, 2020). Many universities made serious investments in faculty development and pedagogical innovation for both online and anti-racist pedagogies. And yet our assessment practices remain unchanged.<sup>1</sup>

Even before the pandemic, a groundswell of discontent prompted a change in teaching effectiveness assessment (TEA) practices. Though largely triggered by equity concerns about bias<sup>2</sup> in student evaluations and the workload of faculty peer review of teaching, a fundamental issue is teaching effectiveness assessment is not effective in achieving either of its primary goals: (1) supporting the development of more effective teachers and thus increasing student success; and (2) evaluating teaching effectiveness as part of the employment assessment process. If we do not assess the

<sup>1</sup>A study of nearly 1100 SET questions across a variety of US institutions of higher learning found “after 30 years, course evaluation questions still adhere to Marsh’s original model of nine categories. Second, we found a vast majority of the questions focus on instructors’ performance, using a predictable pattern of truth statements (i.e., ‘The instructor graded me fairly ...’) that may cause students to hyper-focus on their teacher rather than to evaluate their own learning ...; language emphasizing instructor performance makes teachers responsible for every aspect of a student’s success, a tacit assumption that conflicts with contemporary pedagogy” (Ray et al., 2018, n.p.).

<sup>2</sup>The extensive literature on this bias need not be reiterated here; however, it is worth mentioning this bias was part of the original design of SETs, first created in 1923, by psychologist Max Freyd: “Early on [...] SETs not only highlighted teacher personality and appearance but also were constructed to measure teachers’ subjective traits, not student learning. ... Students were thus put into a position where they were evaluating factors that neither they nor the instructor had sole control over and they could do little to affect” (Ray et al., 2018, n.p.).

achievement of primary goals, we will either not achieve them, or we will fail to support those who seek to achieve them. For example, at SF State, the mission is social justice through education,<sup>3</sup> but none of our teaching effectiveness assessment practices assess the achievement of that goal.

### *The Changing Context of Higher Education*

We write from the assumption universities should support faculty in the ongoing development of their pedagogy in order to improve student outcomes. This assumes the possibility of change in the ways we teach and the cultivation of spaces in which faculty feel safe to become learners themselves and have the freedom and support to continually adapt to new information, changing conditions, and student voices. We believe setting this goal for teaching effectiveness assessment is the most high impact practice in which any group of educators could engage, and the one most likely to support student success.

While the evaluation of teaching effectiveness is the focus of this chapter, we note that structural conditions in higher education constrict the possibilities of liberatory change. The neoliberal turn in higher education features disinvestment, the erosion of autonomy, and the reduction of the lives of students, faculty, and staff to data (Martell, 2021). This data is not, as many assume, *pure*, as in free from bias. The questions that frame data collection, the intentions of those who collect and use it, and the systems by which it is collected and interpreted may replicate and conceal bias (Benjamin, 2019). Those who rely on such data must be prepared to mitigate these biases; however, in campus climates in which dynamics such as stereotype threat (Collins, 2020; Steele & Aronson, 1995) or carceral antiblackness (Shange, 2019) are unacknowledged or superficially understood, we should expect bias to permeate the learning, teaching, and assessment environment.

Disinvestment in higher education followed the increasing percentage of students who are people of color and working class, the same students who demanded curricula that center their history<sup>4</sup> and instructors who

<sup>3</sup>“With the unwavering commitment to social justice that is central to the work of the University, SF State prepares its students to become productive, ethical, active citizens with a global perspective” (<https://sfsu.edu/mission/mission.html>).

<sup>4</sup>For example, the 1966 Ten Point Program of the Black Panther Party demanded “education that teaches us our true history and our role in the present-day society. We believe in an educational system that will give to our people a knowledge of self” (Newton, 2009, p. 4).

shared their experience.<sup>5</sup> The increasing share of faculty who are people of color, working class, and women would seem ideally positioned to address the needs of those students *except* this increase has corresponded to the casualization of faculty labor and a decrease of public higher education support by the state. Thus, women and BIPOC (Black, Indigenous, People of Color) faculty are disproportionately relegated to the inferior second-tier both of public higher education and within the systems into which they are allowed (Griffin, 2020; US Department of Education, 2020). The period of these demographic and structural changes in higher education corresponded to the emergence of mandatory student evaluations of teaching (SET),<sup>6</sup> which can be traced to the student protest movements of the 1960s (Gelber, 2020, p. 47) and may be seen as an administrative attempt to appease student demands for accountability without genuinely functioning to achieve the real goals of student protest: self-determination and sovereignty, education toward liberation, and hiring more diverse faculty (Epstein & Stringer, 2020).

Predominately white and male tenure-line faculty<sup>7</sup> teach fewer courses, receive more compensation for non-teaching activities, and enjoy greater support for professional development (Kezar, 2017; Thirolf & Woods, 2018). Tenure-line faculty are structurally better able to stay active in their disciplines, have a voice in shared governance including curricular design, and continue to learn to be better educators. Contingent faculty teach more courses and do not typically receive compensation to stay active in their fields or continue learning. In these ways, the liberatory potential of the demographic change in higher education has been stymied by the rise

<sup>5</sup>In 1968 the Black Student Union and the Third World Liberation Front at San Francisco State demanded fifty dedicated faculty for a School of Ethnic Studies and Black Studies Department with autonomy over hiring, retention, and curriculum (Epstein & Stringer, 2020).

<sup>6</sup>“Student Evaluations of Teaching (SETs) were independently developed in the 1920s by the educational psychologist Herman H. Remmers at Purdue University (e.g., Remmers & Brandenburg, 1927) and the learning psychologist Edwin R. Guthrie (e.g., Guthrie, 1953) at the University of Washington. Remmers and Guthrie wanted to provide university teachers with information about how their teaching was perceived by students and thus help them to make improvements, where necessary. They intended to limit access to these course evaluations to course teachers. Even though Guthrie warned in 1953 that ‘it would be a serious misuse of this information to accept it as ultimate measure of merit’ (p. 221), SETs soon became valued sources of information for university administrators, who used them as a basis for decisions about merit increases and promotion” (Stroebe, 2020).

<sup>7</sup>Faculty at the full professor rank are 80% white and 67% male. And 59% of all tenure-line faculty are men (US Department of Education, 2020).

of the two-tier faculty labor system (cf. Berry & Worthen, 2021, p. 84). Because of the failure to compensate contingent faculty for service labor, on many campuses, peer observations of contingent faculty are exclusively conducted by tenure-line faculty, such that educators under vastly different labor conditions evaluate one another without considering those differences in the evaluation process.

While academic freedom is beset by an increasing number of cases in which faculty are harassed, targeted, disciplined, or dismissed (Missé, 2021), the academic freedom of tenure-line faculty is ostensibly protected through the job security offered by tenure, affording them the freedom to teach, conduct research, and publish in one's discipline without fear of retaliation or intimidation (AAUP, 1970). This principle is effectively non-existent for contingent faculty because academic freedom is predicated on tenure (Berry & Worthen, 2021, pp. 99, 105). The literature shows educators who innovate in the classroom may receive negative student evaluations, especially in response to pedagogies that emphasize student agency.<sup>8</sup> Further, faculty know this and thus are influenced by the chilling effect of SETs within the neoliberal university in which students are recast as *consumers* and instructors as *service providers*:

We cannot compel educational consumers to attend classes; we cannot make them uncomfortable with their privilege or the state of the environment. We are not supposed to challenge their abilities or to insist on the integrity of academic disciplines. We are creating a space where it is difficult, if not impossible, to be the teachers we want to be. For students, consumerism in higher education creates a type of pseudo-agency where market power stands in as a proxy for real critical consciousness and community-building. (Hoben et al., 2020, p. 167)

<sup>8</sup>“Trends that emphasize student agency and decentering instructor’s authority raise questions about the validity of SETs, particularly those based on generic forms such as the SEEQ or IDEA that do not account for classrooms modeled on distributed agency. ... students may not immediately appreciate de-centered models of teaching in which an instructor acts more as a facilitator or workshop leader. First-year college students in particular may react negatively to instructors who do not fit with their perceptions of college professors as ‘sages on a stage.’ Likewise, open-ended writing prompts may appear to many students to lack organization or clarity. Such reactions reflect common beliefs that students may have about instructor agency—that an effective instructor is authoritative, controlling, and/or objectively unimpassioned about their subject” (Ray et al., 2018, n.p.).

Due to this model of students as consumers, existing SET structures actually discourage innovation and therefore undermine the possibilities of using TEA to support faculty growth. Because student evaluations are often the only data determining whether to rehire contingent faculty, the structural vulnerability of contingent faculty presents a formidable obstacle to pursuing pedagogical innovation (Erickson, 2021).

Managers have met the problem of decades of public disinvestment in higher education by raising tuition and increasing reliance on lower-paid, disposable contingent instructors, who are now the majority of faculty and who teach the largest proportion of courses and students (Berry & Worthen, 2021). In this context, it does not make sense to create policy using tenure-line faculty as the normative model of the educator. In fact, we must assume conditions of contingency as the baseline and lecturer faculty as the exemplary figures of teaching in higher education, particularly due to the fact the majority of contingent faculty have teaching as their sole assignment; teaching is normatively the only measure by which lecturer faculty are retained or rehired (Berry & Worthen, 2021, pp. 120–4; Erickson, 2021).

We might imagine a future Museum of Neoliberalism in which student evaluations of teaching are displayed as exemplary artifacts, like thumb-screws in a museum of torture. The docent might explain this instrument was once used to reduce the exploration, creativity, and dialogic exchange of a learning community to an abstract numerical ranking and managers far removed from the classroom created elaborate comparative spreadsheets, which they subjected to arcane, infinitesimal comparisons, like the reading of tea leaves, to craft justifications for denials of promotion, pay raises, and retention. The docent might also point out these abstract rankings had arguably concealed and amplified social biases based on race, gender, age, accent, or national origin.

What the future docent and we ourselves might miss is what this simultaneously crude and sophisticated technology didn't do. What these rankings and the majority of student comments haven't done is provide instructors with constructive feedback about how to improve their teaching.

### *Actionable Data, Bias, and Statistical Meaninglessness*

We argue for the radical transformation of the use of student feedback in the evaluation of teaching effectiveness based on three arguments

supported by data. First, student evaluations of teaching (SETs) contain little actionable information to improve teaching outcomes<sup>9</sup> and student achievement of learning outcomes; second, current policies provide little guidance on how to appropriately interpret and apply SET quantitative ratings and comments for employment purposes (particularly from an anti-bias perspective); and third, “When results are summarized and only mean or median ratings are included in a dossier, negative scores and comments are inadvertently awarded extra weight in a review” (Linse, 2017, p. 103), thus amplifying the harm of biases (whether implicit or explicit). Even if these limitations are addressed, SETs can be harmful to faculty because of the widespread lack of confidence in SETs and especially because concerns about their application in employment decisions undermine their use for teaching improvement.<sup>10</sup>

While some scholars of faculty evaluation propose methods for extracting useable insight from SETs while minimizing bias (Kreitzer & Sweet-Cushman, 2021; Linse, 2017), we note the application of these methods may require considerable additional labor, reducing the likelihood institutions will implement them. Further, despite prior claims of a high correlation between positive student evaluations and student learning, recent studies found low or even zero correlation, meaning students do not learn better from instructors who receive positive scores (Uttl et al., 2017). Wherever one falls on the debate about harm caused by SETs to women and BIPOC (Lazos, 2012) faculty, what is true across the board is they are rarely used in a way that supports the improvement of teaching and student learning. While Linse argues (2017) that studies on the negative impact of student evaluations are either flawed or taken out of context by higher education publications in a form of sensationalist journalism, she also argues:

Student ratings are “broad brush” instruments used to gather information from a group of students, not all of whom will agree. They are not precision

<sup>9</sup>One instructor reported, “For several years I consistently was dinged on the question ‘instructor is open to a variety of points of view.’ I was really confused by this as I feel that I’m very open to hearing what students have to say. It wasn’t until I had a peer observe my class that they commented that I sometimes cut students off, or would react to their ‘erroneous’ declarations with humorous disdain. It took someone [observing] my sometimes naturally interruptive style [for me] to make headway on this and I’m very proud of the fact that I have. But is that something I should not be hired over? And without explanation, what was I to do with that information?” (Anonymous, personal communication; November 3, 2021).

<sup>10</sup>“Inappropriate use of student ratings breeds mistrust, fosters inequities and inconsistencies, and ultimately demoralizes the faculty” (Linse, 2017, p. 103).

tools that produce a measurement that can then be compared to a known standard. Unfortunately, some faculty evaluators over-interpret small differences as indicative of a problem, a decrease in quality, or an indication one faculty member is materially better than another. (2017, p. 100)

We would like to emphasize this point because of the impact, in practice, of focusing on small differences in the results—which cannot be correlated to a measurable improvement of student learning. Over-interpretation of small variations in ratings can lead to big employment decisions. For example, at San Francisco State University (SFSU), the student rating system ranges from 1 to 5, with 1 being the best. However, many departmental retention, tenure, and promotion (RTP) criteria indicate to faculty and their supervisors any instructor receiving above 2.0 has “failed.” The majority of SFSU departmental RTP criteria documents include language similar to this: “Generally[,] scores of below 1.5 on the evaluation questions indicate excellent teaching; Scores between 1.5 and 2.0 are good; Scores of 2.0 or higher suggest a need for improvement.” In contrast, none of this is transparent to the students who are giving the ratings. For them, a five-point scale is more familiar as the A–F, in which a C is a passing grade. This would be equivalent to a “3” on the 1–5 rating for faculty, which is far above the static “2” most departments cite as the cut off for acceptable performance.

Further, at SFSU, many RTP policies require comparison to the departmental, programmatic, or college *mean*, creating an absurd system in which many faculty are guaranteed to “fail” purely because of a policy that falsely equates an arbitrary data point to effective teaching. According to Linse, “Unit means are not an appropriate cutoff or standard of comparison because there will always be some faculty members who are, by definition, ‘below the mean.’ This is particularly problematic in units with many excellent teachers” (2017, p. 102). Few RTP criteria at SFSU acknowledge a high degree of excellence within the department complicates a reliance on means; and even for these, a reliance on the mean may be substituted with the inflexible score of “2.”

While some RTP policies suggest the number that is best for the candidate should be used when there is a discrepancy between the college, departmental, or program mean and the fixed ratings number, most are muddled. There also is an extreme variance between departments. For example, one states, “Excellence in teaching will be gauged in reference to the College-wide average and should be better than the College-wide



average for the semester under review. Quantitative scores over 2.25 indicate serious concerns,” while another states, “SETE averages of 1.6 and better are deemed appropriate for tenure consideration.” However, there is no set rating number that would make sense, because the mean changes every semester. Likewise, reliance on the mean would also be inadvisable, because a requirement all faculty ratings be better than the mean relegates a significant proportion to categorical, undeserved failure.

Additionally, Linse argues poor ratings are often due to so many variables it is important to “not over-interpret ... relatively small differences in average ratings” (2017, p. 100). Linse presents myriad factors that impact ratings, and suggests potential remedies, all of which center on giving faculty resources, support, and time to improve. While these may be provided in the case of tenured/tenure-track faculty, lecturer faculty may be more vulnerable to the over-interpretations of ratings because there is less investment in their teaching development. Linse’s analysis shows that student ratings distributions are typically negatively skewed, giving more weight to students with biased outlier views:

In skewed distributions, means are sensitive to (influenced by) outlier ratings; in student ratings, these outliers are almost always low scores ... Student ratings instruments ... are best at capturing the modal perceptions of respondents, but they are not the best instruments for capturing rare views, i.e., the views of students represented by the tail of the distribution. While students with outlier views are not unimportant, they should not be given more weight than the views of most students. *This is particularly crucial when evaluating the ratings of non-majority [sic] faculty because we often see students with biased views represented in the tails of the distribution.* (pp. 101–102; *emphasis added*)

An argument administrators might make is bias impacts only a small number of faculty (and thus is not a concern), ignores the likelihood these outliers might be faculty who least resemble the traditional model of a professor, and those most impacted by imposter syndrome, stereotype threat, micro- and macro-aggressions, and who already swim against an underlying tide of bias and exclusion (Hune, 2020, p. 9; Muhs et al., 2012).

We strongly recommend no quantitative ratings of any kind be used in any part of the TEA process. However, if a system of teaching effectiveness assessment *must* use student ratings, they should be developed in consultation with statisticians and applied for specific purposes of supporting

faculty development and, if applied to employment decisions, with protective buffers built into both policy and practice. Administrators and department chairs in the position of assessing the rehiring of lecturer faculty, for example, must be trained<sup>11</sup> to understand how to interpret student ratings. If resources cannot be dedicated to developing instructor and administrator skills in interpreting student ratings, ratings *must not* be used in employment decisions.

Even more concerning than the lack of clarity or misinterpretation of ratings is the fact there is no apparent correlation between student ratings and student learning (Lawrence, 2018; Uttl, et al, 2017; Flaherty, 2016). In other words, there is no evidence these demonstrably harmful quantitative ratings offer any valid assessment of teaching effectiveness. In a 2017 “Meta-analysis of faculty’s teaching effectiveness,” Uttl, White, and Gonzalez argue, “The best evidence—the meta-analyses of SET/learning correlations when prior learning/ability are taken into account—indicates the SET/learning correlation is zero.” They conclude “simple scatterplots as well as more sophisticated meta-analyses methods indicate students do not learn more from professors who receive higher SET ratings.” Given one of the primary arguments for conducting student evaluations of teaching is they encourage student success via teacher effectiveness, this meta-analysis strongly suggests student evaluations fail to meet this purpose. As critics of the metrics-obsessed era of primary and secondary education remind us, “what is measurable is not the same as what is valuable” (Safir & Dugan, 2021, p. 12).

<sup>11</sup>The training suggested by “A Guide for Making Valid Interpretations of Student Evaluation of Teaching (SET) Results” is extensive and labor and resource intensive:

Relevant stakeholders should receive training on both basic survey research principles and psychometric concepts such as validity and reliability. Survey research training should focus on key concepts, such as sample size, MOE, and confidence levels, and how each of these factors interacts in the context of SETs. Training on basic psychometric properties such as validity and reliability should focus on the notion that validity not only addresses the accuracy of a set of scores but also the appropriate interpretation and use of scores. (Royal, 2017)

The modern educational data system itself has been implicated as a harmful form of *scientific colonialism*,<sup>12</sup> particularly in imposing standard models of comparison and evaluation criteria that give inadequate weight to the cultural perspectives and lived experience of the people subjected to assessment (Hall, 1992; McDougal III, 2014; Safir & Dugan, 2021). Even if these forms of measuring could be decoupled from their colonial effects, UC Berkeley Professor of Statistics Philip Stark and Richard Freishtat, Vice President of Curriculum at UC Berkeley Executive Education, expose the rating system of student evaluations of teaching as a house of cards predicated on multiple errors of basic statistical science. They conclude, “The common practice of relying on averages of student teaching evaluation scores as the primary measure of teaching effectiveness for promotion and tenure decisions should be abandoned for substantive and statistical reasons” (2014). They debunk the apparent objectivity of ratings and their use in employment decisions:

Personnel reviews routinely compare instructors’ average scores to departmental averages. Such comparisons make no sense, as a matter of statistics. They presume the difference between 3 and 4 means the same thing as the difference between 6 and 7. They presume the difference between 3 and 4 means the same thing to different students. They presume 5 means the same thing to different students and to students in different courses. They presume a 3 “balances” a 7 to make two 5s. For teaching evaluations, there is no reason any of those things should be true [6]. SET scores are ordinal categorical variables: The ratings fall in categories that have a natural order, from worst (1) to best (7). But the numbers are labels, not values. We could replace the numbers with descriptions and no information would be lost: The ratings might as well be “not at all effective,” ... “extremely effective.” It does not make sense to average labels. Relying on averages equates two ratings of 5 with ratings of 3 and 7, since both sets average to 5. (Stark & Freishtat, 2014, p. 2)

<sup>12</sup> Smith describes this phenomenon: “In Western epistemology, understanding is viewed as being akin to measuring. As the ways we try to understand the world are reduced to issues of measuring, the focus of understanding becomes more concerned with procedural problems” (as cited in Safir & Dugan, 2021, p. 14).

In light of the statistical meaninglessness of such ratings, their lack of correlation to student learning, and their inherent biases, we argue there is no way to recuperate quantitative ratings for any legitimate purpose.

*A Modest Proposal: TEA for Transformation Versus TEA  
for Status Quo*

Having argued the case against SETs as currently designed and used, and having outlined the challenges facing the assessment of teaching effectiveness, we propose the following set of practices to support improved teaching and learning outcomes.<sup>13</sup>

First, *identify* all members of the campus community with a stake in the outcome and *define* the scope of their interest. Faculty can legitimately expect assessment processes to be anti-biased, be transparent, and provide actionable feedback accompanied by institutional support for implementation. Administrators, department chairs, and members of RTP committees have a valid need for assessment data on which to make employment recommendations and decisions with confidence. Staff members dedicated to the educational mission may have a stake related to their work with faculty and with students. And students expect to have their feedback contribute to faculty and curricular development, to be informed about how their

<sup>13</sup>The alternative we propose is inspired in part by the suggestions in Stark & Freishtar's (2014) analysis: "In 2013, the University of California, Berkeley, Department of Statistics adopted as standard practice a more holistic assessment of teaching. Every candidate is asked to produce a teaching portfolio for personnel reviews, consisting of a teaching statement, syllabi, notes, websites, assignments, exams, videos, statements on mentoring, or any other materials the candidate feels are relevant. The chair and promotion committee read and comment on the portfolio in the review. At least before every 'milestone' review (mid-career, tenure, full, step VI), a faculty member attends at least one of the candidate's lectures and comments on it in writing. These observations complement the portfolio and student comments. Distributions of SET scores are reported, along with response rates. Averages of scores are not reported. Classroom observation took the reviewer about four hours, including the observation time itself. The process included conversations between the candidate and the observer, the opportunity for the candidate to respond to the written comments, and a provision for a 'no-fault do-over' at the candidate's sole discretion. The candidates and the reviewer reported that the process was valuable and interesting. Based on this experience, the Dean of the Division now recommends peer observation prior to milestone reviews. Observing more than one class session and more than one course would be better. Adding informal classroom observation and discussion between reviews would be better. Periodic surveys of former students, advisees, and teaching assistants would bring another, complementary source of information about teaching" (p. 4).

perspectives will be used, to have access to clearly defined mechanisms through which to seek redress for harms experienced in the classroom, and also to celebrate instructors who positively impact their learning and success.

Second, *engage* all stakeholders to determine the objectives to be assessed; then, *align* assessment questions and practices, including how the assessments will be used, toward the desired goals. Assessment questions should be achievable and assessable, and faculty development to achieve these objectives must be supported equitably for all instructors by the institution. Questions related to instructor effectiveness also must focus only on those things over which an instructor has control.<sup>14</sup> Assessment practices must be developed through systems of shared governance and must be transparent to all stakeholders, participants, and users. If there is a campus-wide commitment to principles such as equity, social justice, or anti-racism, these objectives must be explicitly integrated in teaching and learning objectives in every program.

Third, *decenter*<sup>15</sup> summative, end of semester, instruments such as SETs and *redesign* evaluation as an ongoing, growth-oriented process throughout the professional career of individual instructors and within the context of supportive teaching communities. Rather than one high-stakes instrument riddled with defects, the evaluation of teaching should include formative student feedback such as midterm evaluations, focus groups, open class discussion, and formative self and peer evaluations, such as through the self-peer observation tool (SPOT) process described below.

Fourth, *eliminate* any quantitative rating system, such as Likert scales, from self, peer, and student perspective gathering instruments. Include longitudinal evidence, such as surveys of students a year after they have

<sup>14</sup> “Our analysis reveals that instructors are often placed as grammatical subjects in questions despite their lack of agency over some areas such as participation and student devaluation of alternative teaching practices. Such question formation skews students toward evaluating instructor factors that are, at best, uninformed, and, at worst, biased. We recommend that SET questions align with student learning and/or engagement rather than teacher performance, rewriting questions to put students in subject positions and cue them on best practices as a better reflection of rhetorical theorizations of classroom agency” (Ray et al., 2018, n.p.).

<sup>15</sup> “While students are in a good position to evaluate some aspects of teaching, there is compelling evidence that student evaluations are only tenuously connected to overall teaching effectiveness. They offer only a single perspective on a very complex and multifaceted teaching and learning process that no single source of evidence can reasonably evaluate” (Stark & Freishtat, 2014).

completed a course, or student success in graduate school or career placement.<sup>16</sup> And, include analysis of other institutional factors that impact student experiences and student success in a particular course, such as how a program chooses to schedule the course, what aligned tutorial services are available, course enrollment caps, instructional aids or Graduate Teaching Assistants (GTAs), the course learning modality, the support available for appropriate faculty professional development, and other factors.<sup>17</sup>

Fifth, completely separate extraordinary employment decisions, such as failure to retain or promote faculty, from any mechanism designed to support instructor teaching effectiveness development. Any employment decision processes also must be supported by faculty and administrator development courses to learn best practices for gathering and applying any form of teaching effectiveness assessment for the purposes of making employment decisions.

Sixth, *transform* campus climate by creating systems to prevent and respond to bias. *Build* from justice, equity, diversity, and inclusion principles instead of adding them on to an ostensibly neutral model. *Center* the voices of the most disenfranchised students and faculty at all stages of the process (cf. Safir & Dugan, 2021, p. 52). This effort should feature proactive education about systems of bias and oppression including white supremacy, patriarchy, capitalism, and colonialism with attention to specific forms such as white privilege, anti-Black and anti-Asian violence, settler colonialism/gentrification, and the intersections between race, class, gender, sexuality, and other attributes.

### *Self and Peer Observation*

We offer the following draft models for formative self, peer, and student observation and reflection. We also provide a model for soliciting stakeholder observations to support extraordinary employment decisions.

<sup>16</sup>As Boyle and Schmierbach propose, “Midterm evaluations, focus groups, open discussion with the class, and even experimenting with different approaches at random within *or between courses* can all provide valuable data” (2021, *emphasis added*).

<sup>17</sup>In a collaborative autoethnographic analysis of how course evaluation questionnaires at Canadian universities “shape and restrict our teaching identities as well as the identities of our students,” the authors note that “only the teachers are assessed even though classrooms are a microcosm of the institutional forces, personalities, and cultures that intersect within these spaces. Within this model, it is easy for student satisfaction to take precedence over pedagogic goals” (Hoben et al., 2020, p. 161).

### *Self-Reflection*

Self-reflection is not codified within most institutional practices of TEA, but has the potential to be the most truly transformative. Bali and Caines argue for “dialogue and reflection with others” in order to achieve “transformative learning, learning that will create deep and lasting change in our practice because it is based on reflection on how our beliefs and values influence our practice, and the connections we make with others in the process” (2018, p. 20). Self-reflection is a meta-cognitive process that allows us to consider what we do well and what we think we do not do well, and thus to consider our relationship to new skills, such as learning new approaches to pedagogy (Haukås, 2018, p. 12). Self-reflection also allows for instructors as stakeholders to have a meaningful voice in their own development through TEA.

The sample self-peer observation tool (SPOT) described below can be used for self-reflection, perhaps with additional questions about changes the instructor would like to make, based on evidences such as student perspectives surveys, students responses to specific pedagogical strategies, or successful student achievement of learning outcomes through specific assignments or assessment activities. We have combined self-reflection with peer observation to enhance alignment between these practices and also to support the development of teaching and learning communities.

### *Peer Observation*

Many peer observation tools and practices are built on the same inherently biased framework as SETs; thus, despite a relative lack of research on peer observations, it is possible to infer from studies on bias in hiring, tenure, and other practices in which faculty evaluate one another, peer observations may be particularly harmful to BIPOC, women, and other marginalized faculty (Starck, et al, 2020). In a study on implicit and explicit bias among K-12 teachers, researchers found that “[f]aculty can also act from implicit bias in their evaluations of each other” (Gleason & Sanger, 2017, p. 14; *emphasis in original*). Thus, peer observation tools, policies, and practices must be designed, practiced, and analyzed from an explicit anti-bias stance.

To illustrate this impact, we provide the following example from a dissertation on men of color in the California community college system, written by a Black man in a tenure-track position in such an institution at

the time of the incident he describes. This example shows how the ascendance of white faculty over BIPOC faculty in rank can contribute to an accumulation of harmful bias. Eventually, the incident so disturbed the author that he separated from the institution prior to the tenure and promotion process. Would he have been harmed by this biased peer observation in terms of being denied tenure or promotion? Possibly. Was he harmed by it in other ways? Definitely. Dr. Collins' persistence within academia despite the negative impact of this peer observation by his then-department chair is a mark of his resilience, rather than of the negligible impact of biased peer observation.

Collins frames his narrative with an analysis of stereotype threat in educational contexts. Within this framework, he initially questions his own academic persistence in the face of racism, eventually becoming a careful practitioner of student-centered pedagogy: "Relying on Hammond's (2013) Culturally Responsive Pedagogy to ensure critical thinking and writing while integrating the cultural knowledge and background of my students, I carved a space in my classroom that celebrated authenticity, dialogue, and vulnerability (Ponjuán & Hernández, 2016)." However, the narrative below shows his liberation pedagogy conflicted with existing stereotype threats, and the institutional practices of peer observation in teaching effectiveness assessment externalized this conflict.

When I was hired, I was told I was hired because I was a "successful Black man" and I was expected to work with the Black student population, however, my methods for promoting Black authenticity, identity, and resourcefulness were criticized by both the dean and the department chair, both who happened [*sic*] to be white females. One area of critique was my style of classroom management. During one of my classroom [peer] evaluations, a Black male student walked in late. I simply said, "Hello Jay, thanks for being here," and continued lecturing. The department chair was upset and in my first tenure review meeting revisited the incident and told me the better way to handle the student would be to shame him in front of the entire class. She admonished me for welcoming him into the classroom without calling him out in front of the class for being late. The department chair contended that embarrassing him in front of his peers would make him come to class on time in the future. Her suggestion of how I should have handled the tardy student reminded me of my own student experience in community college. The memory of when I was locked out of the classroom for being late resurfaced. The memory of the time I was yelled at in front of the entire classroom because my research paper did not meet the teacher's expectations



hauntingly returned. The vicious institutional (micro)aggression was, and still is a problem in community college.

Some suggestions for reducing bias in peer observations include developing an awareness of such bias including bias related to instructor identity, student biases, sub-field biases, confirmation bias, and teaching approach bias (Troisi, 2021) within the pool of faculty and administrators who conduct and review such observations, as well as carefully designing peer observation tools to refocus on observations of equity and excellence in student-teacher interactions, to relate to pedagogical standards and innovations in the field, and to require specific evidence to support observations.

We strongly recommend peer observations be conducted with pre-observation meetings and post-observation meetings, as well as reviews of additional materials, including course syllabi, online course management systems, assignments, student learning assessment rubrics, and student work. In order to mitigate power differences and biases, and to better support faculty development, we also recommend self-reflections and peer observations be conducted in tandem, preferably with both parties conducting both a self-reflection and a peer observation, when possible. Peer observers may also want to reflect on how acting as observers/mentors impacts their own professional development as teachers. Doing this work should be both recognized and compensated as an important part of the labor of developing and maintaining equity and excellence in teaching and learning.

In fall 2021, San Francisco State University piloted a new self-peer observation tool developed by the Center for Equity and Excellence in Teaching and Learning (CEETL) with stakeholder input facilitated by the Academic Senate Teaching Effectiveness Assessment Task Force. The SPOT identifies five teaching areas that have been shown to support student success, especially for BIPOC and first-generation students, as verified by an extensive literature review conducted within SFSU's CEETL and sponsored by the California State University Quality Learning and Teaching Initiative. For each of these five teaching areas, the SPOT provides direct links to resources within the CEETL Online Teaching Lab (OTL) and the Justice, Equity, Diversity and Inclusion (JEDI) Institute,

among other offerings.<sup>18</sup> Faculty who used the SPOT as an optional formative assessment component in their Spring 2021 Faculty Teaching Squares (which is not part of their formal teaching evaluation process) generally reported positive experiences; reports from the Fall SPOT Pilot are forthcoming. The SPOT functions as two sides (self and peer) of a triangle of self, peer, and student perspectives in the teaching effectiveness assessment process. The purpose is to support teaching effectiveness development within an anti-oppressive framework. This framework seeks to support—rather than manage—faculty labor.

For each of the five areas addressed in the SPOT, we provide below a rationale (“Why”), suggested supports, and suggested assessment practices. Policies for the implementation of instructor self-reflection and peer observation should also include rationales, inventories of existing and needed institutional supports, and holistic assessment practices (including mandatory mentoring, e.g., pre- and post-observation meetings). These policies must also provide guidelines for how such self-reflections are to be used; we recommend divorcing them entirely from employment decisions. Faculty may wish to quote from their self-reflections or peer observations in their teaching narratives, but must not be required to do so.

### *Course Organization*

- Why: Courses should be organized in ways to support students in building self-efficacy and confidence in their ability to succeed.
- Support: For example, course organization can be supported by providing a syllabus template or online tool, online course management templates, and development to use these as part of faculty onboarding, and as an ongoing process of deepening faculty abilities to respond to student and environmental contexts and development in the field of instructional design.
- Development Assessment: Peer review and student experiences of course organization elements and learning environments should pro-

<sup>18</sup>These include Online Teaching Lab (<https://ceetl.sfsu.edu/online-teaching-lab>); Justice, Equity, Diversity and Inclusion PIE Institute (<https://ceetl.sfsu.edu/jedi-pie-institute-course>); Asian and AAPI Solidarity Statement and Teaching Resources (<https://ceetl.sfsu.edu/asian-and-aapi-solidarity-statement-and-teaching-resources>); and Anti-Racist Pedagogies (<https://ceetl.sfsu.edu/anti-racist-pedagogies>). Additional resources addressing other specific pedagogical approaches are under development in collaboration with campus partners.

vide constructive feedback during formative assessment.<sup>19</sup> Assessment of course design should include samples from extant course design elements, peer reviews, student experiences, and instructor self-reflections. Formative questions might include “how did the instructor communicate where to start, course outcomes, and other important course information, such as course materials, frequency of instructor communication, community agreements, and grading policies?” and “how did the syllabus provide an overview of the semester schedule and was the course divided into manageable pieces?”

### *Context and Purpose*

- Why: Articulating personal and collective purposes helps imbue student learning outcomes with meaning, motivating them to succeed.
- Support: Faculty should be supported to understand and communicate their personal, social, community, and local contexts and positionalities as well as to support students to develop and communicate their own contexts and positionalities. These might include land acknowledgments, pronoun declarations, preferred forms of address, and other markers of identity and vulnerability in order to build a learning community based in trust and mutual respect.
- Development Assessment: Analyze sections of peer observations, self-reflections, student experience surveys, and instructor and student context and positionality statements in formative assessment processes. Formative assessment questions might include: “how did the instructor acknowledge social conditions shaping student experience<sup>20</sup> without singling out individuals?” and “how did this course foster a culture of knowledge development that is co-constructed

<sup>19</sup>Instructors should compare their course organization at different stages and reflect on changes they have made and how these changes were indicated by peer and student comments and by professional development experiences. They may also benefit from surveying students on alternate course organization structures.

<sup>20</sup>Student academic success is often radically shaped by their material conditions and their emotional and physical wellbeing outside of the classroom environment. As Erickson documents in a study of Oakland’s lowest performing middle schools, realities outside of educational institutions can have catastrophic impact on student success, especially for working-class BIPOC students and members of their families subject to multiple forms of trauma—racial profiling, food insecurity, violence, housing insecurity, incarceration, deportation—that may interrupt or permanently derail educational goals (2014).

through students' lived experiences and particular contexts, and toward their personal goals?"

### *Student and Community Engagement*

- Why: Students who feel excluded, marginalized, or invisible have difficulty developing a sense of efficacy and benefit not only from feeling included but by gaining practice in successfully acting on the world as individuals and as members of a learning community.
- Support: Provide multiple and ongoing development opportunities for active, engaged pedagogy based on a shared learning community, and access to resources that support community service learning.
- Development Assessment: Analyze sections of peer observations, self-reflections, student experience surveys, and student learning artifacts in formative assessment processes. Formative assessment questions might include, "how did course activities include opportunities to make a difference in the world such as project-based learning, building a resource for the campus (or other) community, and sharing information with peers?" and "how did the instructor engage students through active and group learning?"

### *Teacher Presence*

- Why: The development of an inclusive teaching presence that communicates expectations of equity and excellence begins with faculty expertise in the identities and social realities of BIPOC, LGBTIQAA+, differently abled, neurodiverse, and other students whose identities have not been normativized in academic student learning assessment.
- Support: Provide faculty development courses and consultations that support faculty to: (1) identify and assess personal goals for intersectional anti-racist pedagogy; (2) examine and demonstrate knowledge of historical and contemporary institutional and individual racism and white supremacy in education practice; (3) assess current assignments, assessments, and teaching practices through a critical race and

intersectional perspective; and (4) design strategies for inclusive and equitable engagement.<sup>21</sup>

- Development Assessment: Peer observations, self-reflections, student experience surveys, and teaching philosophy or pedagogy statements should be discussed formatively. Formative questions for self, peer, and student might include “how did the instructor position themselves within hierarchies of oppression such as gender, race, and class?” and “how did course materials center the knowledge and accomplishments of members of diverse communities?”

### *Student Learning Assessment*

- Why: Ensure assessment strategies are aligned with student learning outcomes and reduce barriers to student achievement. Anti-biased assessment strategies should include how to create an ecology of trust to cultivate a trust-based learning community.
- Support: Provide faculty development courses and consultations that support faculty to: (1) align learning activities with stated outcomes; (2) use student-friendly language to communicate expected outcomes, grading policy, and transparent grading practices; (3) provide regular feedback to students across a variety of modalities; (4) provide multiple opportunities for students to demonstrate their learning; (5) provide opportunities for reflection and metacognition one or more times throughout the course; (6) actively incorporate strategies that promote justice, equity, diversity, and inclusion pertaining to assessment and feedback on student learning.
- Development Assessment: Analyze formative surveys of student learning; track student achievement by associating student learning assessment activities with particular outcomes; review students’ summative reflection on their achievement of learning outcomes. Questions might include “how did assessment activities (essays, quizzes, tests, etc.) provide opportunities for students with diverse learning styles to succeed?”

<sup>21</sup>These supports are those of SFSU’s Justice, Equity, Diversity and Inclusion Pedagogies for Inclusive Excellence (PIE) Institute, <https://ceetl.sfsu.edu/jedi-pie-institute-course>.

### *Student Perspectives*

Student perspectives have a legitimate, useful place in teaching effectiveness assessment—especially for formative feedback that can improve a course immediately.

#### *Sample Student Perspective Survey*

The following questions are built around the same five categories as the SPOT, with the addition of a section on learning modality<sup>22</sup> and context, and a global reflection on all the other areas.

#### *Course Design*

- How did the syllabus and course materials provide the information you and your classmates needed to be successful in this course? Please mention specific elements of the syllabus, online course management system, and other course materials.
- How did course activities and assignments help you see connections between what you learned and your future goals?

#### *Inclusion and Belonging*

- What about this course helped you experience a sense of community and connection with your classmates?
- How did your instructor help you and your classmates feel welcomed into and valuable to the class?
- Does everyone in the class know how to pronounce your name?
- How did the instructor acknowledge social realities (such as white supremacy or patriarchy) shaping your life experience without making you feel singled out as an individual?

<sup>22</sup>As the title of one article argues, there is a significant “Difference Between Emergency Remote Teaching and Online Learning.” This article suggests that “[c]olleges and universities working to maintain instruction during the COVID-19 pandemic should understand those differences when evaluating this emergency remote teaching.”

### *Teacher Presence*

- How did your instructor motivate you and your classmates to work hard and to believe you could succeed?
- In what ways did your instructor make adjustments to instruction based on your and your classmates' learning needs and feedback?

### *Engagement*

- How did the course materials and assignments help you achieve the learning outcomes?
- How did the instructor engage and motivate you and your classmates to learn during discussions and learning activities?
- What did you and your classmates do to support your learning in this course?
- Did any aspect of this course help you feel you could make a difference to something you care about?

### *Assessment*

- How did the feedback you received from your instructors and classmates help you improve your performance in this course?
- How did the processes used to determine grades support your learning in this course? In what ways were these processes clear and equitable and how can they be improved to be more so?

### *Modality and Context*

- What were the learning modalities of this course? (online, in-person, HyFlex, etc.)
- How did the learning modalities of your other courses impact your learning experiences in this course?
- What was the context of this course for you personally (e.g., did you experience a major life change?), locally (e.g., environmental or social factors), nationally, or internationally (e.g., significant social factors) and how did any of these contexts impact your learning experiences in this course?

### *Global*

- What factors most impacted your successful completion of this course? (Select all that apply)
  - Course Design
  - Inclusion and Belonging
  - Teacher Presence
  - Engagement
  - Assessment
  - Modality
  - Other
  
- Please comment on those factors you selected in terms of their impact on your learning in this course.

Do student perspectives also have a place in employment decision processes? Perhaps so, but we would argue this is for extraordinary cases and should be solicited outside of student perspectives used to inform faculty development. These two conflicting purposes should not be mixed; doing so erodes faith the solicited information will be used supportively and thus can impact the candor of respondents, as well as how instructors perceive, receive, and use the information provided. It is for these reasons we recommend an extraordinary commendations and concerns process, outlined below.

### *Extraordinary Commendations and Concerns*

*The biggest challenge we have faced in reimagining TEA processes is how to reconcile the creation of a zone of faculty autonomy that supports growth with the use of TEA data in employment decisions, which forecloses that autonomy.* We have concluded TEA for faculty development must be completely separated from TEA for employment purposes. TEA, for reasons presented above, must not be used at all in decisions to retain, promote, or separate. We assume the majority of faculty are performing within an acceptable range and can improve their teaching if they are positively drawn to development opportunities. The current model puts all faculty on the chopping block (with lecturer faculty closest to the blade). We propose taking all faculty off the chopping block except in cases of extraordinary cause.



**CONCERN**

Please name an individual or office through which you have experienced incidents or patterns of concern

Their department or program

Course(s) taught and semester(s) (if faculty)

What happened; why is it cause for concern?

What changes would you recommend?  
What outcome would you like to see?

Your name and contact information (required)

**Fig. 4.1**

The instrument we propose is a Commendations and Concerns Comment Box, separate from other TEA processes. This extraordinary process could reside on department websites, online course management system homepages, graduation/separation surveys, as links from formative or summative instruments, or within bias-incident reporting structures. Submitting a commendation would trigger a process leading to recognition and awards, while submitting a concern would initiate fact-finding to determine if intervention is necessary and to recommend a process of redress where warranted.

The commendations box (Fig. 4.1) encourages an environment of celebration: in the best cases of transformative teaching that supports student success, this method would support and reward excellence, while encouraging achievement through a reward nomination system to recognize and incentivize teaching excellence. Whether or not commenders should be anonymous will probably engender some debate; in our opinion,

**COMMENDATION**

Please name an individual or office through which you have experienced extraordinary service, learning or inspiration

Their department or program

Course(s) taught and semester(s) (if faculty)

What was extraordinary; what positive impact was made?

Would you like to nominate this person or office for an award for teaching or service? Why?

Your name and contact information (required)

Fig. 4.2

commendors should be encouraged to go on record and participate in celebrating those being commended

The concern box (Fig. 4.2) provides accountability: how can ineffective faculty or those who commit harm be supported to learn to create an effective, inclusive learning environment, or, if they refuse to shift, be held accountable and ultimately removed? Whether or not complaints should be anonymous will probably engender serious debate; in our opinion, the complainant should be encouraged to go on record and be available for further engagement toward resolution. Such complaints should be directed to the HR Title IX office, with a Faculty Rights panel of the faculty union notified as watchdog, perhaps including student voices in the process as well. Title IX staff can clarify the facts of the incident to determine next steps. Consider the following typical incidents and the differing responses they might prompt, either a determination that no harm was done or a recommendation for mediation and support:

1. A faculty member teaches critical racial analysis; a student feels uncomfortable and reports the instructor.
2. An instructor uses the word “negro” in historical context; a student hears it as the n-word and complains.
3. A faculty member presents Palestinian perspectives on the Israeli occupation and a student denounces the instructor as an anti-Semite.
4. Other students misgender a trans student and the instructor doesn’t intervene.
5. An instructor asks a third-generation Asian American student to describe their own immigrant experience.
6. All the course materials are written by men; women students complain they have been erased from the discipline.

In these examples we see the potential for de-escalation, reconciliation, growth, and development rather than immediate escalation to formal grievance with potential employment consequences. Universities could benefit from developing the capacity for practices of *restorative justice*, a facilitated process in which those who have done harm can take responsibility for their actions and grow from the experience and those who have been harmed can experience being heard, receive redress, and heal (cf. Karp & Schachter, 2018). There are multiple potential benefits: a sense that justice was done and agency for the person harmed; the opportunity to develop strong equity practices; an opportunity for redemption for persons who commit harm; and, for administrators, a reduction in the number of formal grievances filed at their university. Although this process may not always be appropriate or possible—both parties must be willing participants—creating the capacity to practice restorative justice can contribute to the real structural transformation of the institution so equity is not relegated to superficial declarations, siloed programs, or token spokespeople (cf. Dugan, 2021).

### *Hard Choices and Obstacles*

We have shown prevalent TEA practices work at cross purposes to positive outcomes for students, faculty, and administrators alike. We have presented proposals for the reorientation of TEA practices to support pedagogy development that each campus community might pursue in its own way.

Despite widespread frustration with SETs and eagerness for change among faculty, we foresee numerous traps preventing campuses from

putting down swords of TEA torment and picking up plowshares of TEA transformation. We predict the following half-measures likely to occur but unlikely to produce improved outcomes for any campus stakeholders:

1. Campuses will rename SETs, *Student Opinions of Teaching* or *Student Voice Surveys*, while continuing to misuse student data in the same unproductive ways.
2. Campuses will update SETs with new questions, which might produce better data but which will still be misused as employment management tools, undermining their value as faculty development resources.
3. Campuses will improve their summative instruments such as SETs and peer observations but fail to develop formative processes and support essential to faculty growth.
4. Campuses will bow to white fragility and reduce commitments to equity to toothless diversity discourse<sup>23</sup> that contains rather than liberates students and faculty of marginalized identities.
5. Administrators will insist on Likert scales, arbitrary numeric means, and comparative norms despite lack of legitimacy and potential for harm, thus maintaining faculty and student distrust of TEA processes.
6. Campuses will design transformational TEA processes around tenure-line faculty, marginalizing the majority contingent faculty and thus excluding the majority of students from the benefits.
7. Campuses will focus on isolated incidents of bias instead of attending to systemic bias-producing processes.

Campuses that surpass these obstacles will create conditions to produce tangible benefits for all stakeholders:

- Students can become active agents in their learning processes, witness responsive faculty, and feel empowered to hold faculty accountable in instances of harm and to publicly recognize exceptional educators.

<sup>23</sup>Ahmed documents the ways *diversity* serves as a managerial discourse and marketing term that conceals systemic inequalities within universities and thus serves as a containment strategy that makes challenges to racism appear impertinent (2012, pp. 52–3).

- Faculty can experience security in which to innovate and be open to continual learning toward excellence and equity in teaching.
- Administrators may see fewer complaints escalate into formal grievances, fewer disengaged students drop out, and more students graduating in less time.

In the end, we are left with an existential question about the purpose of higher education today. Do we exist to equitably serve and co-liberate the human potential of an already heterogeneous learning community? Or is our purpose to reduce learning processes to quantifiable data to serve labor management? It will not and cannot be both.

## REFERENCES

- Ahmed, S. (2012). *On being included: Racism and diversity in university life*. Duke University Press.
- American Association of University Professors. (1970). *1940 Statement of principles on academic freedom and tenure*.
- Bali, M., & Caines, A. (2018). A call for promoting ownership, equity, and agency in faculty development via connected learning. *International Journal of Educational Technology in Higher Education*, 15, 46. <https://doi.org/10.1186/s41239-018-0128-8>
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.
- Berry, J., & Worthen, H. (2021). *Power despite precarity: Strategies of the contingent faculty movement in higher education*. Pluto Press.
- Boyle, M., & Schmierbach, M. (2021). *Measurement in the classroom: Using student evaluations to explain research concepts and improve your teaching*. Routledge.com. <https://www.routledge.com/blog/article/using-student-evaluations-to-explain-research-concepts>
- Collins, C. (2020). *Structural racism disguised as (micro)aggressions. A narrative inquiry into the lived experience of men of color*. Doctoral dissertation. San Francisco State University.
- Dugan, J. (2021). No shortcut: Avoiding equity traps and tropes. In S. Safir & J. Dugan (Eds.), *Street data: A next generation model for equity, pedagogy, and school transformation* (pp. 25–45). Corwin.
- Epstein, K. K., & Stringer, B. (2020). *Changing academia forever: Black student leaders analyze the movement they led*. Myers Education Press.
- Erickson, B. (2014, August 21). Blue on Black violence and original crime. *Anthropoliteia: Critical Perspectives on Police, Security, Crime, Law and Punishment*.

- <https://anthropoliteia.net/2014/08/21/blue-on-black-violence-and-original-crime-a-view-from-oakland-california/>
- Erickson, B. (2021). Abolish the lecturer: A manifesto for faculty equity. In K. Roth & Z. S. Ritter (Eds.), *Whiteness, power and resisting change in higher education* (pp. 215–228). Palgrave Macmillan.
- Flaherty, C. (2016). Zero correlation between evaluations and learning. *Inside Higher Ed.* <https://www.insidehighered.com/news/2016/09/21/new-study-could-be-another-nail-coffin-validity-student-evaluationsteaching>
- Gelber, S. (2020). *Grading the college: A history of evaluating teaching and learning*. Johns Hopkins University Press.
- Guthrie, E. R. (1953). The evaluation of teaching. *The American Journal of Nursing*, 53(2), 220–221. <https://doi.org/10.2307/345992>
- Gleason, N. W., & Sanger, C. S. (2017). *Guidelines for peer observation of teaching: A sourcebook for international liberal arts learning*. Centre for Teaching and Learning, Yale-NUS College.
- Griffin, K. (2020). Institutional barriers, strategies, and benefits to increasing the representation of women and men of color in the professoriate: Looking beyond the pipeline. In L. W. Perna (Ed.), *Higher education: Handbook of theory and research* (pp. 277–349). Springer. <https://diversity.ucdavis.edu/sites/g/files/dgvnsk731/files/inline-files/InstitutionalBarriersStrategies2020.pdf>
- Hall, S. (1992). The west and the rest: Discourse and power. In S. Hall & B. Gieben (Eds.), *The formations of modernity* (pp. 276–320). Polity and Open University.
- Harrison, F. (Ed.). (1991). *Decolonizing anthropology: Moving further toward an anthropology for liberation*. Association of Black Anthropologists/American Anthropological Association.
- Hune, S. (2020) in Valverde, C.K.L & Dariotis, W.M. (Eds.) *Fight the Tower: Asian American Women Scholars’ Resistance and Renewal in the Academy* (pp. 1–28). Rutgers University Press
- Hammond, Z. (2013). *Culturally Responsive Teaching and The Brain: Promoting Authentic Engagement and Rigor Among Culturally and Linguistically Diverse Students*. Corwin Press.
- Lazos, S. (2012). Are Student Teaching Evaluations Holding Back Women and Minorities?: The Perils of “Doing” Gender and Race in the Classroom. In Gutiérrez y Muhs, G., Niemann, Y., Gonzalez, G, & Harris, A., (Eds.) *Presumed incompetent: The intersections of race and class for women in academia* (pp. 164–185). University Press of Colorado, Utah State University Press. <https://doi-org.jpallnet.sfsu.edu/10.2307/j.ctt4cgr3k.1>
- Linse, A. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees *Studies in Educational Evaluation* 54, 94–106.

- Haukås, Å. (2018). Metacognition in language learning and teaching an overview. In Å. Haukås, C. Bjørke, & M. Dypedahl (Eds.), *Metacognition in language learning and teaching*. Routledge.
- Hoben, J.L., Badenhorst, C., & Pickett, S. (2020). Student evaluations and the performance of university teaching: Teaching to the test. *LEARNing Landscapes*, 13: 161–172. <https://files.eric.ed.gov/fulltext/EJ1261356.pdf>
- Holme, T. (2020). Introduction to the journal of chemical education special issue on insights gained while teaching chemistry in the time of COVID-19. *Journal of Chemical Education*, 97(9), 2375–2377. <https://doi.org/10.1021/acs.jchemed.0c01087>
- Karp, D., & Schachter, M. (2018). Restorative justice in universities. In T. Gavrielides (Ed.), *Routledge international handbook of restorative justice* (pp. 247–263). Routledge.
- Kezar, A. (2017). *Adjunct faculty voices: Cultivating professional development and community at the front lines of higher education*. Stylus Publishing.
- 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*. <https://doi.org/10.1007/s10805-021-09400-w>
- Ladson-Billings, G. (2005). From the Achievement Gap to the Education Debt: Understanding Achievement in U.S.Schools. *Educational Researcher* 35(7), 3–12.
- Lawrence, J. (2018, May). Student evaluations of teaching are not valid. *American Association of University Professors*. <https://www.aaup.org/article/student-evaluations-teaching-are-not-valid>
- Martell, J. (2021). Resisting the neoliberal university with a general strike. In K. Roth & Z. S. Ritter (Eds.), *Whiteness, power and resisting change in higher education* (pp. 193–214). Palgrave Macmillan.
- McDougal III, S. (2014). Research methods in Africana studies (1st ed.) *Black studies and critical thinking* (Book 64). Peter Lang, International Academic Publishers.
- Missé, B. (2021). A matter of academic freedom. In K. Roth & Z. S. Ritter (Eds.), *Whiteness, power and resisting change in higher education* (pp. 129–157). Palgrave Macmillan.
- Muhs, G. G. y., Niemann, Y. F., González, C. G., & Harris, A. P. (2012). *Presumed incompetent: The intersections of race and class for women in academia*. Utah State University Press.
- Newton, H. (2009). *To die for the people*. City Lights Books.
- Ponjuán, L. & Hernández, S. (2016). Untapped potential: Improving Latino males' academic success in community colleges. *Journal of Applied Research in the Community College* 23(2), 1–20

- Pokhre, S., & Chhetri, R. (2021). A literature review on impact of COVID-19 pandemic on teaching and learning. *Higher Education for the Future*, 44(2), 133–141. <https://doi.org/10.1177/2347631120983481>
- Ray, B., Babb, J., & Wooten, C. A. (2018). Rethinking SETs: Retuning student evaluations of teaching for student agency. *Composition Studies*, 46(1), 34–56, 192–194.
- Remmers, H. H., & Brandenburg, G. C. (1927). Experimental data on the Purdue rating scale for instructors. *Educational Administration and Supervision*, 13(6), 399–406.
- Royal, K. (2017). A guide for making valid interpretations of student evaluation of teaching (SET) results. *Journal of Veterinary Medical Education*, 44(2), 316–322. <https://doi.org/10.3138/jvme.1215-201R>. Published Online: August 03, 2016.
- Safir, S., & Dugan, J. (2021). *Street data: A next generation model for equity, pedagogy, and school transformation*. Corwin.
- Shange, S. (2019). *Progressive dystopia: Abolition, antiblackness, and schooling in San Francisco*. Duke University Press.
- Stark, J. G., Riddle, T., Sinclair, S., & Warikoo, N. (2020). Teachers are people too: Examining the racial bias of teachers compared to other American adults. *Educational Researcher*, 49(4), 273–284. <https://doi.org/10.3102/0013189X20912758>
- Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811. <https://doi.org/10.1037/0022-3514.69.5.797.Pdf>
- Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, 42(4), 276–294. <https://doi.org/10.1080/01973533.2020.1756817>
- Thirolf, K. Q., & Woods, R. S. (2018, March 25). Contingent faculty at community colleges: The too-often overlooked and under-engaged faculty majority. *Wiley Online*. <https://doi.org/10.1002/ir.20244>.
- Thomas, M. (2020, December 3). Virtual teaching in the time of COVID-19: Rethinking our WEIRD pedagogical commitments to teacher education. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2020.595574>
- Troisi, J. (2021). *Summary version: Potential biases in peer reviews of teaching center for teaching and learning*. <https://blogs.rollins.edu/facultyevaluation/wp-content/uploads/2021/03/Implicit-Bias.pdf>



- US Department of Education, National Center for Education Statistics. (2020). *Characteristics of postsecondary faculty*. <https://nces.ed.gov/programs/coe/indicator/csc>
- Uttl, B., White, C., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42. <https://doi.org/10.1016/j.stueduc.2016.08.007>