Published in final edited form as:

J Dent Educ. 2017 November ; 81(11): 1362–1372. doi:10.21815/JDE.017.093.

Should Student Evaluation of Teaching Play a Significant Role in the Formal Assessment of Dental Faculty? Two Viewpoints

Dr. Susan Rowan, DDS,

Clinical Associate Professor and Clinical Dean, Department of Restorative Dentistry, University of Illinois at Chicago College of Dentistry

Dr. Elmer J. Newness, DDS, MBA,

Clinical Assistant Professor, Department of Oral Health and Integrated Care, University of Detroit Mercy School of Dentistry

Dr. Sotirios Tetradis, DDS, PhD,

Professor and Chair, Section of Oral and Maxillofacial Radiology, University of California, Los Angeles School of Dentistry

Dr. Joanne L. Prasad, DDS,

Assistant Professor, Department of Oral Biology, University of Pittsburgh School of Dental Medicine

Dr. Ching-Chang Ko, DDS, MS, PhD, and

Distinguished Professor and Vice Chair, Department of Orthodontics, University of North Carolina at Chapel Hill School of Dentistry

Dr. Arlene Sanchez, DMD, MHSA

Professor and Assistant Dean for Academic Affairs, Department of Restorative Dentistry, University of Puerto Rico School of Dental Medicine

Abstract

Student evaluation of teaching (SET) is often used in the assessment of faculty members' job performance and promotion and tenure decisions, but debate over this use of student evaluations has centered on the validity, reliability, and application of the data in assessing teaching performance. Additionally, the fear of student criticism has the potential of influencing course content delivery and testing measures. This Point/Counterpoint article reviews the potential utility of and controversy surrounding the use of SETs in the formal assessment of dental school faculty. Viewpoint 1 supports the view that SETs are reliable and should be included in those formal assessments. Proponents of this opinion contend that SETs serve to measure a school's effectiveness in support of its core mission, are valid measures based on feedback from the recipients of educational delivery, and provide formative feedback to improve faculty accountability to the institution. Viewpoint 2 argues that SETs should not be used for promotion and tenure decisions, asserting that higher SET ratings do not correlate with improved student learning. The advocates of this viewpoint contend that faculty members may be influenced to

focus on student satisfaction rather than pedagogy, resulting in grade inflation. They also argue that SETs are prone to gender and racial biases and that SET results are frequently misinterpreted by administrators. Low response rates and monotonic response patterns are other factors that compromise the reliability of SETs.

Keywords

dental education; dental faculty; staff development; teaching; faculty development; student evaluation of teaching; SET; teaching effectiveness; promotion and tenure

Dental schools must maintain high teaching standards to ensure that learning takes place and that the knowledge gained will assist students in achieving competence. Effective teaching can contribute to the successful delivery of information and support learning. However, assessment of the quality of teaching poses some challenges. Student evaluation of teaching (SET) is one way to measure teaching effectiveness and to assess faculty performance.

The use of SETs in higher education to assess faculty performance has been researched extensively, with varying conclusions. For this reason, the validity of SETs in promotion and tenure decisions is a matter of ongoing debate. This Point/Counterpoint article debates the utility and fairness of using SETs in assessment of dental school faculty performance and for administrative decisions related to promotion and advancement. Viewpoint 1 presents research supporting the validity of SETs, assuming an appropriately constructed, implemented, and interpreted SET tool is used. Their value in improving teaching performance as well as assessing faculty accountability is discussed. Viewpoint 2 presents research that raises considerable doubt about the validity and reliability of SETs, including poor response rates, biases, and misinterpretation of the data. This viewpoint also discusses the potential conflict of interest in SETs that could cause faculty members to compromise learning in order to please students.

Viewpoint 1: Formal Faculty Assessment Should Include Student Evaluation of Teaching

SETs have been used in many ways, including instructional improvement, program and course evaluations, and promotion and tenure decisions.^{1–8} Properly constructed SETs are reliable and should be considered in the formal assessment of faculty members. Proponents of this opinion contend that SETs serve to measure a school's effectiveness in support of its core mission, are valid measures based on feedback from the recipients of educational delivery, and provide formative feedback to improve faculty accountability to the institution. We argue that the evidence supports the use of SETs in the formal assessment of dental faculty members' performance and, by extension, in a school's making of personnel decisions (Table 1).

Value of Assessing Teaching Effectiveness

According to Clayson's and Kelly's studies, accountability for outcomes of learning and effective delivery of information is an important administrative responsibility and core

mission of any academic institution.^{3,4} A strategy for assessment of curriculum content, course structure, and individual teaching effectiveness should be in place to meet an institution's academic obligations. The studies by Clayson and Kelly found that SET scores can be used as a quantitative measure of teaching effectiveness. In order to assess faculty members' teaching performance and assess the effectiveness of their teaching, student input in the form of SETs is essential and should be used in concert with other criteria in promotion and advancement decisions. Peer evaluations, trained observer assessments, self-evaluation, teaching portfolios, learner outcomes on standardized examinations, and SETs may all be used in the global assessment of teaching performance.

Kelly reported that use of SET scores to evaluate faculty members in the promotion and tenure process varied across universities.⁴ Although promotion and tenure dossiers typically consist of many components, including course outlines, summary of accomplishments, scholarly activities, and statements of teaching philosophies, Arah et al.'s study found that multi-source assessments increased validity in evaluating teaching effectiveness.⁷ Davidovitch found that, in many cases, universities use the average scoring of SETs for all courses to evaluate teaching effectiveness rather than individual course evaluations.⁸ In this way, outlier evaluations, both high and low, are neutralized. In a study of 570 medical faculty members, Arah et al. found a clear differentiation between high and low faculty performance, as well as a high internal consistency between faculty self-evaluation and their SET assessments.⁷ This study suggests that administrators should use the average data set to identify exceptional teachers, as well as those who consistently receive poor student evaluations as part of the assessment measure of overall academic performance.

Teaching effectiveness has long been said by dental school deans to be essential to evaluating faculty members. In a study by Jones and Preusz published in 1989, the deans of U.S. and Canadian dental schools were asked to rank the importance of teaching in comparison to publications, another commonly used measure of performance.⁵ According to the responding deans, classroom, clinical, and laboratory teaching ranked higher in order of importance than number of publications, which ranked fourth. In addition, SETs appear to be widely used as measures of teaching effectiveness in dental education. In 2008, Jahangiri et al. reported that 81% of the 39 U.S. dental schools responding to their survey were using SETs as a method of evaluating teaching performance.⁶ Jahangiri et al. argued that student evaluations/ratings of faculty members should be an essential component of measuring teaching performance, especially because a large number of students typically complete surveys of teaching effectiveness, making the information more reliable.

Assessment of teaching effectiveness is so important to U.S. dental schools that it is mandated by the Commission on Dental Accreditation (CODA). As part of compliance with the Accreditation Standards for Dental Education Programs, Standard 3–4 states a “defined evaluation process must exist that ensures objective measurement of the performance of each faculty member in teaching, patient care, scholarship, and service.”⁹ Although not specifically mandated by CODA, SETs can be a valuable mechanism to gather the required data measuring objective faculty performance in conjunction with other evaluation tools.

Students Can Evaluate Teaching Effectiveness

Students, as the recipients of instruction, are capable of evaluating teaching effectiveness based on their perceptions and experiences and thus can provide meaningful formative and summative feedback to instructors and administration. Al Kuwaiti, in his study of health science students' evaluations of courses and instructors at the University of Dammam, argued that, provided an adequate response rate and class size, students are the most qualified individuals to evaluate whether the learning experience was productive or worthwhile.¹⁰ In a survey of 378 first-, second-, and third-year PharmD students by Surratt and Desselle, respondents reported that they recognized the necessity and importance of evaluating their instructors.¹¹ Students in that study also reported that they appreciated the opportunity to voice their opinions about improving teaching quality.

As Al Kuwaiti emphasized, students are the individuals with the greatest exposure to a teacher's performance and are most affected by its quality, so their input is essential to evaluation of teaching quality and value.¹⁰ Moreover, in efforts to use evidence-based decision making, it would seem reasonable that administrators consider students as consumers and benefactors of the teaching enterprise to provide useful data sets. Murray et al.'s study of SET effectiveness in a sample of 46 psychology instructors, published in 1990, concluded that student raters of individual faculty members were consistent across items and time periods and were only minimally impacted by confounding factors such as grade scale or class size.¹² Furthermore, he found that their evaluations were consistent with those of such other raters as observers, alumni, and peer evaluators. We agree that learners are the most appropriate raters of teaching effectiveness, assuming valid methods of data-gathering are used as cautioned by Davidovitch.⁸

SETs Provide Formative Feedback to Teachers

Students' assessments of teaching quality can be used for formative as well as summative feedback. Kornell and Hausman found that awareness of SET inclusion as part of formal faculty assessment made faculty members more concerned about student input and, by extension, more eager to address student needs, improve teaching delivery, and enhance learning effectiveness.¹³ In Arah et al.'s study of medical school faculty members, the respondents reported that they often reviewed their student evaluations with their program directors and that the personal feedback provided in SET data increased their self-awareness of teaching quality.⁷ Overall, those respondents reported that feedback from student evaluations improved their effectiveness.

In Beran and Rokosh's study, 88% of the 357 instructors supported the use of student ratings of teaching, and 54–57% considered SETs to be the most useful tool for improving general teaching quality, refining overall instructions, and improving lecture quality.¹ Without this motivation, faculty members may be less inclined to devote adequate attention to the critical role of teaching. SETs can thus help to provide essential formative feedback to and ensure accountability of faculty members.

SETs Are Valid and Reliable Measures

SETs can be a valid and reliable mechanism to evaluate teaching effectiveness if constructed and delivered appropriately. Marsh's extensive research on SETs, published in 1984, provides an overview of multiple studies suggesting that students can provide appropriate, although imperfect, assessment of good teaching, especially as it pertains to certain attributes, such as organization of material and creation of a learning environment.¹⁴ It is necessary to acknowledge that other studies do not uniformly support that conclusion. As Marsh and Roche acknowledged in 1997, the strong dependence many institutions place on SETs in faculty assessment is most likely due to the lack of other indicators that could more reliably and objectively evaluate teaching effectiveness.¹⁵ If there is agreement that teaching performance is important in faculty assessment, SETs appear, at the present time, to be the most appropriate instrument to measure outcome of this performance. Thus, it is critical that the SET instrument be appropriately constructed and effectively delivered.

Teaching is a complex and multidimensional task. Effective assessment of teaching requires the utilization of scientifically sound, but also practically feasible, measurement instruments.¹⁰ Psychometricians and survey methodologists recommend the use of properly worded questionnaires and avoiding confusing words that might have multiple meanings or yield various interpretations.¹¹ When such instruments are used, student input is reliable, stable, and generalizable.^{6,15} The use of other accepted survey methods, such as Students' Evaluation of Educational Quality (SEEQ), is important to ensuring the soundness of the assessment. The SEEQ instrument is comprised of nine dimensions of teaching and has been studied extensively. Research over a 13-year period, with responses of almost one million students at various levels of education, both undergraduate and graduate, found SEEQ to be both valid and reliable.¹⁵ In that study, the instrument's reliability was demonstrated based on 10–15 or more responses and retrospective data of former students. Much research has been performed over the years to test the validity of the SEEQ design since its creation in the late 1970s by Herbert Marsh. In a more recent study, 1,264 students in eight Greek universities participated in SETs using the SEEQ questionnaire tool, and the results confirmed the validity of the tool relative to the multidimensionality of assessing teaching effectiveness.¹⁶

Other factors must be considered to ensure effectiveness and validity of SETs. A small number of evaluations decrease reliability and stability. Medium to large numbers of responses, however, correlate strongly with teaching effectiveness, greatly affect validity of data, and heighten accuracy of conclusions.¹⁰ There is a better likelihood that students will take the SETs seriously if there is a strong indication by administration that their input matters. Observed changes in teaching approach between a midterm SET and the end of course demonstrate to the students that the tool is useful and has the power to initiate improvement.⁴ Thus, ensuring robust communication to the students relative to the value of their input can positively affect the reliability of the data. Otani et al.'s study of 1,410 university students identified four areas as most influential in perceived overall effectiveness of teaching, which should be considered in the construction of survey tools.¹⁷ Those four areas were clarity of the instructor's explanations of the course subject; creating a positive and risk-free learning environment; effective use of class time; and stimulating course

materials. One of the least influential factors was about the instructor rating work in fair and appropriate ways. That study indicated that SETs should be constructed so that consideration is given to most influential items related to teaching effectiveness.

An additional way to increase the validity of SET scores is giving more weight to very good students' evaluations, which Braga et al. found to be a more reliable measure than using all students' evaluations.² Also, Braga et al. urged that surveys be conducted later in academic training, thereby allowing students the time to recognize the real value of teaching effectiveness as it related to their subsequent learning. Grillo et al.'s survey of dental and dental hygiene students at the University of Michigan found that, for the 329 respondents, SET format should be limited in length, hosted during class time, and allow for short answer responses to be most effective.¹⁸ In general, these studies found that, if efforts were made to appropriately construct, administer, and evaluate SET, teaching performance can be accurately reflected.

Overall SET Effectiveness

Because SETs have been found to be valuable in assessing teaching effectiveness and are an integral component of the overall evaluation of faculty performance, we conclude that SETs should be used during personnel decisions, along with other factors such as scholarship, research, authorship, faculty service, mentoring, growth, and self-evaluation. However, when SETs are used as part of an evaluation of overall faculty performance, it is critical to appreciate the strengths and weaknesses of SET instruments in order to maximize their efficacy and minimize their flaws. SETs should only be used in assessing teaching effectiveness and not to evaluate course goals, objectives, and content.¹⁰ SETs should not be the only tool to assess teaching effectiveness, but should be combined with other methods of evaluations, including peer assessment and self-assessment.^{2,6} The weight that SETs' input carries in career decisions may be great and justifies the need to enhance their usefulness.^{10,11,15} An administrative commitment to provide the appropriate human and monetary resources to support all aspects of faculty assessment is essential.^{6,17} The construction, delivery, collection, and evaluation of responses using up-to-date evidence are required to create reliable outcomes.¹⁰ Ensuring adequate student participation and educating students about the importance of the evaluation process are also critical. Clearly, although extensive research has been conducted on the topic of SETs since the 1920s, more progress is needed in constructing and validating effective and reliable SET tools since their outcomes have the potential to significantly impact faculty careers.

Viewpoint 2: Student Evaluation of Teaching Should Not Be Part of Formal Faculty Assessment

According to CODA, evaluations of teaching must be included as a component of course evaluations, and dental schools are required to review faculty teaching performance using "objective measurement."⁹ Even though use of SETs is not mandated by CODA, they are frequently used as the primary measure of teaching effectiveness in the formal assessment of a faculty member.⁶ An underlying assumption of this practice is that SETs are objective and valid measures of teaching performance. Yet, there is compelling evidence that this premise

is seriously flawed. Numerous studies have shown that SETs correlate poorly with teaching effectiveness; create a conflict of interest; are biased; are inherently unreliable; and are judged subjectively. Given these concerns, we argue that SETs should not be used in formal faculty reviews or for determining promotion or tenure (Table 2).

SETs May Not Reflect Teaching Effectiveness

Studies have found that SETs correlate poorly with teaching effectiveness.^{19,20} One of these studies is a recently published meta-analysis by Uttl et al. that provides convincing evidence that there is no significant correlation between SET ratings and student learning.¹⁹ The authors reviewed three meta-analyses and reviewed 97 studies evaluating the relationship between SETs and objective measures of student learning and found no correlation between higher SET ratings and improved learning. Most of the studies included in the meta-analysis involved U.S. institutions of higher learning, although some Canadian and European studies were also included. Various disciplines at the undergraduate and graduate level were represented in their analysis. A few studies of SETs in health sciences disciplines were also included although none were from dental education. Stark and Freishtat reviewed the statistical evidence linking SETs to teaching effectiveness and also concluded that SETs did not reflect teaching effectiveness.²⁰ In a third study, high SET ratings and student performance in a downstream related course were in fact negatively correlated.²

“Consumerism” in higher education is one framework that may explain the lack of relationship between SET ratings and student learning. In this framework, students behave like consumers of a commodity. Studies have shown that teachers who are more “likable” receive disproportionately higher SET scores.^{21,22} In one study, students verbalized common reasons for giving high ratings to professors.²² These reasons revolved around non-pedagogical characteristics, such as the judgment that a professor was “charismatic,” “animated,” or “entertaining.” “Boring” was a common reason to rate teaching effectiveness as “low.” As the author explains, in these situations the students’ level of enjoyment “becomes conflated with teaching quality.”

SETs Pose Conflicts of Interest for Faculty

The perceived “control purposes” of SETs in determining promotion or tenure are associated with increased stress among faculty members.²³ If SETs are linked to promotion and tenure, one cannot deny that SETs create a significant conflict of interest for faculty. Professors may be unduly influenced, consciously or unconsciously, to emphasize entertainment at the expense of pedagogy or to inflate grades to please the class, since students with higher grades are more likely to give high SET ratings.²⁴ Not surprisingly, high SET ratings do in fact correlate with narrower grade curves.^{2,25} SETs may thus compromise the very thing that they are intended to assess: teaching effectiveness and student learning.

Another affront to faculty members comes in the form of cruel remarks by students, spawned by the anonymous format of SETs.^{26,27} Inappropriate remarks are demoralizing and may sway instructors to modify their behavior to become more “likable,” possibly at the expense of student learning. Finally, the career advancement of a truly effective teacher who

emphasizes pedagogy over entertainment could be negatively impacted. This is another compelling reason to discontinue using SETs in the formal assessment of faculty.

SETs Are Biased Against Female Faculty

SETs were found to be significantly biased against female faculty members in several studies.^{27–34} While gender inequalities in SETs have been described since the mid-1970s, surprisingly little has changed with regards to gender bias.³⁰ Nevertheless, SETs continue to be used today as objective measures of teaching performance and to determine faculty promotion and tenure.

In one randomized control experiment, the mere perception of being taught by a female instructor in an online course resulted in lower SET ratings.³¹ All course sections were identical, except for the listed instructor's name (male vs. female). Students rated the male identities significantly higher than the female identities, regardless of the actual gender of the person teaching the course. Furthermore, the gender bias affected "objective" criteria, such as how quickly assignments were graded, even when no actual differences existed.

Similar findings were observed in another study examining the influence of age and gender biases on SETs for online courses.³² In that study, first-year dental students at two dental schools were randomly assigned to watch one of four videos. Each video was narrated by one of four instructors (young male, young female, older male, older female). The content of the video was identical except for the voice characteristics of the narrator. At both institutions, the older female voice was rated significantly lower for almost all items on the questionnaire, including more objective qualities such as "being organized."

Other research corroborates a gender bias in SETs. In one study of academic physicians, female faculty members received significantly lower SET scores than male faculty members in all core clinical rotations.³³ Another study found a gender bias against females in multi-instructor courses that were co-taught by faculty groups that were gender- and ethnically diverse.³⁴

SETs also provide a venue for students to anonymously express their gender biases in more obvious ways. Comments made as part of evaluations of female faculty are frequently unprofessional and/or of a personal nature, emphasizing such things as dress style and physical appearance. Female faculty of color often receive both racist and sexist comments.²⁷

Divergence from stereotypical gender-role expectations can translate into lower SET ratings.³⁵ A female faculty member who does not exhibit gender-expected traits or behaviors, such as appearing more caring, is more likely to get lower SET ratings. The issue of gender bias is further complicated by the fact that the student's gender has also been found to impact SETs.^{36,37} In those studies, male faculty SETs were unaffected by student gender, whereas female faculty members received their highest ratings from female students and their lowest ratings from male students. In another study, female instructors were given significantly lower SET ratings than their male counterparts if they assigned a low grade to a student.²⁹

Dental programs seek to attract and retain a gender-diverse and racially rich group of faculty members, but administrators and supervisors are unlikely to consider gender, age, or racial biases when evaluating faculty for tenure or promotion. Hidden biases in teaching evaluations are not only unfair, but result in unequal opportunities for the very faculty members that dental programs are trying to attract in efforts to diversify the faculty pool.

SET Results Are Unreliable

Even if students did not act as consumers and hidden biases did not exist, SETs would still fail the test of an “objective” measure of performance due to other factors that undermine their validity or interpretability. One assumption underlying the validity of SET results would be that the students who are completing the evaluations understand the meaning of all items on the evaluation. Yet this assumption was called into question in one qualitative study in which medical students voiced their thoughts while completing their course evaluations.³⁸ Some educational terms, such as “independent learning,” were misinterpreted by a significant number of students, in sometimes divergent ways. The study also highlighted other problems with SETs, such as frequent use of the second highest rating as a default, for reasons not necessarily related to a truthful and favorable impression of a teacher (e.g., uncertainty about the meaning of an item).

“Straight-lining” behaviors, also called monotonic response patterns (MRPs), may also invalidate the results of SETs. MRPs occur when a student provides identical responses on all evaluation items. In some instances, students use MRPs by identifying the column that reflects their “general level of enjoyment” in the course.²² In a study of dental and medical students in a multi-instructor course, MRPs were noted in 37.5% of SETs.³⁹ The dental cohort tended to use MRPs when they had less accurate recall of an instructor. Additionally, MRP ratings tended to be situated closer to the extreme ends of the scale, thus potentially skewing the SET means. In another study, researchers inserted a fictitious instructor into multi-instructor preclinical medical courses.⁴⁰ Students were given the choice to rate each faculty member. Approximately two-thirds of students chose to rate the “phantom professor,” and about half did so even when provided with a photo. Mindless behavior or poor recall on the part of students is therefore a factor to consider for multi-instructor courses.

Another significant problem undermining the validity of SETs seen at many institutions is a low response rate.²⁴ This problem has become more prevalent with the transition to online SETs. This issue cannot be understated as it compromises the data.^{10,24,41} As response rates decline, the potential for error increases, and the data should be called into question. A low response rate increases the possibility that the opinions of nonrespondents differ from those of respondents (nonresponse bias). A concern for a “double-bias” also exists, since students with higher grades are more likely to respond and more likely to give positive ratings.²⁴ Class size should also be considered. A minimum response rate of 74–91% is deemed necessary to obtain stable evaluations in classrooms with less than 60 students.¹⁰

SETs Are Prone to Subjective Misinterpretation

Finally, misinterpretation of SET results by administrators is a compelling reason to discontinue using SETs in formal faculty reviews. In a recent study, perception of teaching quality was influenced by small, statistically insignificant variations in SET averages (means).⁴² The variations in SET means deviated from the departmental comparison mean by a mere 0.03 to 0.28 points. But regardless of the size of the difference between the means, participants rated an instructor less favorably if his or her SET mean fell below the comparison mean. Participants still misinterpreted the SET results in the presence of 95% confidence intervals overlapping the comparison mean or when specifically informed that the means were “not significantly different.” These findings suggest that subjective judgment exists regardless of the statistical evidence. Heuristic thinking (“higher is better”) is likely to have skewed the participants’ view of the results. Regardless, the potential for subjective judgement exists, which undermines a fair process.

In conclusion, SETs are anything but objective measures of teaching performance and should not be relied upon in the formal assessment of faculty members. Furthermore, the inclusion of SETs as a component of course evaluations should not be mandated for the purposes of accreditation.

Response by Drs. Rowan, Newness, and Tretadis to Viewpoint 2—Viewpoint 2 raises several important concerns that, if addressed, would improve the effectiveness and objectivity of SETs. However, we disagree with its main arguments. Students are key stakeholders in the education process, and their voice needs to be considered in the operational decision making of administrations, including decisions about faculty retention and promotion. The point is made that SETs do not reflect teaching effectiveness. We agree that instruments to assess teaching effectiveness can be designed in ways that are prone to consumerism and do not capture the ability of teachers to create a learning environment and provide lifelong learning. However, use of properly constructed instruments, such as SEEQs, provide reliable, valid assessment of teaching effectiveness.¹⁵ In an attempt to remove confusion and ambiguity, these instruments should be continuously assessed, restructured, and improved following current evidence-based knowledge.⁷

A second point is that SETs create a conflict of interest for faculty in a way that could compromise teaching effectiveness in favor of pleasing the students. Although this might be true for some individuals, an even greater danger of lack of accountability could be introduced if SETs are excluded from faculty assessment.¹³ This is particularly true in the current academic environment of dental schools, in which the faculty have to meet a multitude of demands with frequently competing priorities. If faculty members are not held accountable for poor teaching performance, teaching quality and effectiveness would likely decrease.

The point that SETs could be biased against female faculty raises a very important concern. The problem unfortunately is not exclusive to SETs, but is a small part of a larger fabric that reflects conscious and unconscious biases for age, race, religion, sexual orientation, or disability for a variety of assessments, evaluations, expectations, or beliefs. It is not clear if the reported SET gender bias is any greater than the gender bias for evaluation of

credentials, professional competence, or salary reimbursement. Raising awareness, increasing conscious counterbalances, or even creating statistical formulas that quantitatively adjust for such biases and discrepancies could provide remedies to the problem.

SET unreliability due to students' not understanding the meaning of all items they evaluate, inclusion of monotonic responses, poor recall, or low response rate are not reasons to eliminate SETs. Such problems can be alleviated by clarification of SET instruments, calibration of students, avoidance of monotonic or confusing questions, explanation to students of SET importance in faculty evaluation, and mandatory participation of students in SETs.⁷ We strongly believe that dental students would positively respond to the challenge of providing valuable SETs, rather than being excluded from or marginalized in the process.

A final point is made that SETs are prone to subjective misinterpretation by administrators. This point does not necessarily reflect negatively on the SETs but rather on their poor utilization and misunderstanding. Heuristic thinking probably is not limited to SETs but expands to other aspects of assessing a faculty dossier. Decision makers at all levels need to use scientific methodology to reach objective and unbiased outcomes at all levels.

Current SET instruments are not perfect. Their strengths and limitations need to be understood and taken into consideration. However, we disagree with a "throwing the baby out with the bathwater" approach, and we advocate a continuous re-evaluation and re-adjustment of SETs to improve their validity and performance. There is strong evidence to support that, when used appropriately, SETs constitute an important component of a faculty member's performance and provide valuable assistance to inform personnel decisions.

Response by Drs. Prasad, Ko, and Sanchez to Viewpoint 1—We completely agree with our colleagues that dental schools must maintain high educational standards and must promote teaching effectiveness. We also agree that teaching quality should be appropriately ascertained, as mandated by CODA. However, we object to the practice of using SETs as quantitative measures of teaching "quality." As we reviewed in Viewpoint 2, there is much evidence to show that SETs are not objective measures of teaching effectiveness. Indeed, multiple studies have shown a lack of correlation (or even an inverse correlation) between SET ratings and student learning.^{2,19,20} As consumers of education, students are more likely to rate instructors based on factors that have little to do with teaching effectiveness, such as their level of entertainment in the course. Numerous studies have also demonstrated that subjective biases (gender, racial, age) significantly impact SET ratings, even for "objective" categories, such as speed of grading³¹ or organization.³² It is disturbing that the mere perception of being taught by a female instructor could result in lower SET ratings. We cannot ignore these biases when we are aware that they exist, since it is not possible for administrators to account for such biases adequately when considering SETs in the formal review of a faculty member. In addition, administrators are influenced by heuristic thinking biases in evaluation of SET results, which also introduces a large degree of subjectivity.³⁸ SETs are unreliable for a host of other reasons, such as student misinterpretation of survey questions, low response rates, or inappropriate response behaviors (monotonic responses). As a discipline that prides itself for promoting evidence-based practice, why would we

disregard a large volume of evidence indicating that SETs are poor, biased, and unreliable measures of teaching effectiveness?

The authors of Viewpoint 1 also claim that students are the most qualified persons to evaluate teaching effectiveness since they have the greatest exposure to teaching performance and, as consumers of education, can provide useful data on teaching effectiveness. However, in this manner, teaching yields positive evaluations only to the extent that the professor meets the student-as-consumer's expectations and values. This evaluation is not predictive of learning or teaching effectiveness, as expressed in Viewpoint 1, but rather a measure of customer satisfaction. Faculty member evaluations are partially based on factors outside their control, such as student dissatisfaction for receiving low grades. Student evaluations tend to be higher when a course is easier or when teachers are more likeable or entertaining.^{21,22} SET implementation can actually come at the expense of learning, as faculty members may raise grades, employ grading leniency, and socialize with students outside of class to minimize receiving negative evaluations.²⁶ It has been found that teachers who make large contributions to effective learning tend to receive lower ratings, perhaps by creating "desirable difficulties" in learning.¹³ The student-as-consumer model also highlights the students' lack of expertise in recognizing effective teaching, so that evaluations become more a measure of popularity rather than serving as effective real indication of pedagogical quality.

Viewpoint 1 also contends that "awareness of SET inclusion as part of formal faculty assessment makes individual faculty members more concerned about student input and, by extension, more eager to address student needs, improve teaching delivery, and enhance learning effectiveness." This point loses sight of a more important concern created by the use of SETs for career advancement purposes: SETs create a conflict of interest for faculty members. A professor could be influenced to focus on entertainment or inflate grades in order to achieve greater student satisfaction and receive higher ratings. This, in turn, can actually compromise teaching effectiveness and student learning, as likeability may be emphasized at the expense of pedagogy.²⁴ SETs may provide an instructor with some useful feedback that he or she may use to improve a course. However, the results of SETs should not be used for determination of tenure or promotion since the control purposes of SETs can be detrimental.

SETs are heavily relied upon in faculty assessment and advancement in higher education. In Viewpoint 1, they are described as imperfect but valid and reliable if optimally constructed, using an accepted survey method such as the SEEQ. However, the same issues arise with SEEQ surveys as with any other SET. The authors of Viewpoint 1 do mention that a low response rate is a problem that decreases the reliability and validity of the survey results. These surveys are also susceptible to monotonic response patterns and gender biases, and such surveys are certainly not immune to an administrator's potentially erroneous subjective interpretation of the results. The authors of Viewpoint 1 also suggest that "surveys could be provided later in academic training, thereby allowing students the time to recognize the real value of teaching effectiveness as it related to their subsequent learning." This practice would be highly impractical in a dental school setting. Given various challenges and flaws

associated with SETs, we continue to contend that they should not be used in the formal assessment of faculty.

References

1. Beran T, Rokosh J. Instructors' perspectives on the utility of student ratings of instruction. *Instructional Sci.* 2009; 37:171–84.
2. Braga M, Paccagnella M, Pellizzari M. Evaluating students' evaluations of professors. *Econ Educ Rev.* 2014; 41:71–88.
3. Clayton D. Student evaluations of teaching: are they related to what students learn? A meta-analysis and review of the literature. *J Marketing Educ.* 2009; 31(1):16–30.
4. Kelly, M. Student evaluations of teaching effectiveness: considerations for Ontario universities. Toronto: Council of Ontario Universities; 2012. COU academic colleagues discussion paper, no. 866
5. Jones JE, Preusz GC. Evaluating dental faculty performance: perceptions of dental school deans. *Psychol Rep.* 1989; 64(2):647–51. [PubMed: 2710900]
6. Jahangiri L, Mucciolo T, Choi M, et al. Assessment of teaching effectiveness in U.S dental schools and the value of triangulation. *J Dent Educ.* 2008; 72(6):707–18. [PubMed: 18519601]
7. Arah O, Hoekstra J, Bos A, et al. New tools for systematic evaluation of teaching qualities of medical faculty: results of an ongoing multi-center survey. *PLoS One.* 2011; 6(10):1–10.
8. Davidovitch N. Student surveys and their applications in promoting academic quality in higher education. *J Coll Teach Learn.* 2011; 8(6)
9. Commission on Dental Accreditation. Accreditation standards for dental education programs. Chicago: American Dental Association; 2016.
10. Al Kuwaiti A. Health science students' evaluation of courses and instructors: the effect of response rate and class size interaction. *Int J Health Sci.* 2015; 9(1):52–60.
11. Surratt C, Desselle S. Pharmacy students' perceptions of a teaching evaluation process. *Am J Pharm Educ.* 2007; 71(1):6. [PubMed: 17429506]
12. Murray H, Rushton P, Paunonen S. Teacher personality traits and student instructional ratings in six types of university courses. *J Educ Psychol.* 1990; 82(2):250–61.
13. Kornell N, Hausman H. Do the best teachers get the best ratings? *Frontiers Psychol.* 2016; 7(570): 1–8.
14. Marsh H. Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases, and utility. *J Educ Psychol.* 1984; 76(5):707–54.
15. Marsh H, Roche L. Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. *Am Psychol.* 1997; 52(11):1187–97.
16. Grammatikopoulos V, Linardakis M, Gregoriadis A, et al. Assessing the students' evaluations of educational quality (SEEQ) questionnaire in Greek higher education. *Int J Higher Educ Res.* 2015; 70:395.
17. Otani K, Kim BJ, Cho J. Student evaluation of teaching (SET) in higher education: how to use SET more effectively and efficiently in public affairs education. *J Public Affairs Educ.* 2012; 18(3): 531–44.
18. Grillo A, Murdoch-Kinch C, Ramaswamy V, et al. Student evaluations of teaching: dental and dental hygiene students' and faculty members' perspectives. *J Dent Educ.* 2016; 8(4):439–50.
19. Uttl, B., White, CA., Gonzalez, DW. Meta-analysis of faculty's teaching effectiveness: student evaluation of teaching ratings and student learning are not related. 2016. At: www.sciencedirect.com/science/article/pii/S0191491X16300323. Accessed 26 Jan. 2017
20. Stark, PB., Freishtat, R. An evaluation of course evaluations. Burlington, MA: Science Open Research; 2014.
21. Delucchi M. Don't worry, be happy: instructor likability, student perceptions of learning, and teacher ratings in upper-level sociology courses. *Teach Sociol.* 2000; 28(3):220–31.
22. Titus JJ. Student ratings in a consumerist academy: leveraging pedagogical control and authority. *Sociol Perspect.* 2008; 51:397–422.

23. Elstad E, Lejonberg E, Christophersen KA. Teaching evaluation as a contested practice: teacher resistance to teaching evaluation schemes in Norway. *Educ Inquiry*. 2015; 6(4)
24. Adams MJ, Umbach PD. Nonresponse and online student evaluations of teaching: understanding the influence of salience, fatigue, and academic environments. *Res Higher Educ*. 2012; 53(5):576–92.
25. Stroebe W. Why good teaching evaluations may reward bad teaching: on grade inflation and other unintended consequences of student evaluations. *Perspect Psychol Sci*. 2016; 11(6):800–16. [PubMed: 27899725]
26. Lindahl MW, Unger ML. Cruelty in student teaching evaluations. *Coll Teach*. 2010; 58(3):71–6.
27. Deo ME. A better tenure battle: fighting bias in teaching evaluations. *Columbia J Gender Law*. 2015; 31(1):7–43.
28. Boring, A., Ottoboni, K., Stark, PB. Student evaluations of teaching (mostly) do not measure teaching effectiveness. Burlington, MA: Science Open Research; 2016.
29. Sinclair L, Kunda Z. Motivated stereotyping of women: she's fine if she praised me but incompetent if she criticized me. *Pers Soc Psychol Bull*. 2000; 26(11):1329–42.
30. Ferber MA, Huber JA. Sex of student and instructor: a study of student bias. *Am J Sociol*. 1975; 80:949–63.
31. MacNell L, Driscoll A, Hunt AN. What's in a name: exposing gender bias in student ratings of teaching. *Innov Higher Educ*. 2015; 40:291.
32. Doubleday AF, Lee LM. Dissecting the voice: health professions students' perceptions of instructor age and gender in an online environment and the impact on evaluations for faculty. *Anat Sci Educ*. 2016:31.
33. Morgan HK, Purkiss JA, Porter AC, et al. Student evaluation of faculty physicians: gender differences in teaching evaluations. *J Women's Health*. 2016; 25(5):453–6.
34. Wagner, N., Rieger, M., Voorvelt, KJ. Gender, ethnicity, and teaching evaluations: evidence from mixed teaching teams. Rotterdam: International Institute of Social Studies, Erasmus University; 2016. ISS working paper series, no. 617
35. Anderson K, Miller ED. Gender and student evaluations of teaching. *Polit Sci Politics*. 1997; 30:216–9.
36. Centra JA, Gaubatz NB. Is there gender bias in student evaluations of teaching? *J Higher Educ*. 2000; 70:17–33.
37. Basow SA. Student evaluations of college professors: when gender matters. *J Educ Psychol*. 1995; 87(4):656–65.
38. Billings-Gagliardi S, Barrett SV, Mazor KM. Interpreting course evaluation results: insights from think aloud interviews with medical students. *Med Educ*. 2004; 38:1061–70. [PubMed: 15461651]
39. Stratton TD, Witzke DB, Freund MJ, et al. Validating dental and medical students' evaluations of faculty teaching in an integrated, multi-instructor course. *J Dent Educ*. 2005; 69(6):663–70. [PubMed: 15947212]
40. Uijtdehaage S, O'Neal C. A curious case of the phantom professor: mindless teaching evaluation by medical students. *Med Educ*. 2015; 49(9):928–32. [PubMed: 26296409]
41. Nowell C, Gale LR, Handley B. Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assess Eval Higher Educ*. 2010; 35(4):463–75.
42. Boysen GA. Uses and misuses of student evaluations of teaching: the interpretation of differences in teaching evaluation means irrespective of statistical information. *Teach Psychol*. 2015; 42(2): 109–18.

Table 1

Key points in support of viewpoint 1: formal faculty assessment should include student evaluation of teaching (SET)

Key Points

-
- Ensuring teaching effectiveness sustains the core mission of a university and is critical to its success. SETs can be a valid measure of teaching effectiveness.
 - Students are the recipients of teaching and are the most appropriate individuals to rate teaching effectiveness.
 - SETs provide formative feedback for faculty members and can improve their performance and accountability to higher administration.
 - With proper construction and utilization, SETs can provide valid, reliable, and useful data for both faculty and administration.
 - Overall, evaluation of teaching quality and accountability for teaching performance are critical to a university's success, and properly constructed, administered, and reviewed SETs should be used in the formal assessment of faculty.
-

Table 2

Key points in support of viewpoint 2: student evaluation of teaching (SET) should not be part of formal faculty assessment

Key Points

-
- High SET ratings do not correlate with improved student learning, and some studies even show a negative correlation between the two measures.
 - When SETs are used for career advancement purposes, a conflict of interest results that can unduly influence teaching behavior. A relationship has been found between high SET ratings and grade inflation and dispersion.
 - SETs are biased against women, and the mere perception of being taught by a female instructor is associated with lower SET ratings. Some research also shows racial and age biases.
 - SETs are inherently unreliable for a host of reasons such as low response error, monotonic response patterns, misinterpretation of rated items, and class size effect.
 - SET results are frequently misinterpreted and are prone to subjective judgment regardless of the statistical evidence. As a result, a fair process is seriously undermined.
 - Overall, SETs are far from being objective measures of performance and should not be used in the formal assessment of faculty.
-