Student Course and Teacher Evaluations

A Report by the Course and Teacher Evaluation Working Group of the Scholarship of Teaching and Learning Task Force

April 2020

ASPPH ASSOCIATION OF SCHOOLS & PROGRAMS OF PUBLIC HEALTH

Association of Schools and Programs of Public Health

The Association of Schools and Programs of Public Health (ASPPH) is the voice of accredited academic public health, representing schools and programs accredited by the Council on Education for Public Health (CEPH). ASPPH is a 501c3 membership organization headquartered in Washington, DC.

The Association's mission is to strengthen the capacity of its members by advancing leadership, excellence, and collaboration for academic public health. The ASPPH's Strategic Framework 2020 calls for the Association to create educational tools and resources for faculty to advance quality and innovation in academic public health.

For more information on ASPPH, see www.aspph.org

Suggested citation:

Association of Schools and Programs of Public Health. Student Course and Teacher Evaluations: A Report by the Course and Teacher Evaluation Working Group of the Scholarship of Teaching and Learning Task Force. Washington, DC; 2020 Apr. 47 p.

Table of Contents

Foreword

The Association of Schools and Programs of Public Health (ASPPH), under the jurisdiction of its Education Advisory Committee, launched the Scholarship of Teaching and Learning (SoTL) Task Force in 2018 to continue enhancing quality and innovation in education for public health. The aim was to create a locus and mechanism for the ASPPH-accredited academic public health community to lead in discovery and reflection on evidence-based methods to research, practice, document, and publish on scholarly teaching and learning.

ASPPH subsequently formed four working groups (WGs): Conceptual Framing; Institutional Models, Programs, and Policies; Teaching; and Course and Teacher Evaluation. Each working group was tasked in January 2019 to produce deliverables in 2020. Working group members were nominated by deans and program directors of ASPPH-member schools and programs of public health. Populated with exceptionally talented, experienced, and devoted educators, the working groups collaborated to support ASPPH's mission to "strengthen the capacity of members by advancing leadership, excellence, and collaboration for academic public health."

This *Student Course and Teacher Evaluations* report and the full complement of forthcoming task force products targets ASPPH-member faculty in *both* baccalaureate and graduate schools and programs. I envision this work will support the value proposition for an important aspect of our members' core business, assuring strong teaching to produce optimal student learning outcomes.

On behalf of ASPPH, I would like to commend the Course and Teacher Evaluation WG chair, Dr. Ali Weinstein (George Mason) and her working group for their dedicated research and recommendations in the following pages. The field of academic public health stands to benefit from their seminal work to encourage methods and practices for faculty and leaders' consideration to improve student course and teacher evaluation in meaningful ways.

This effort could not have been accomplished without Dr. Shan Mohammed (Northeastern), chair of the ASPPH Education Advisory Committee and chair of the Scholarship of Teaching and Learning Task Force, for his intellectual leadership, experience, and thoughtful guidance, as well as ensuring connections across working group products. In addition, ASPPH's director of education, Ms. Elizabeth Weist, deserves thanks for her expert insight, encouragement, and logistical support for the task force throughout each of the working groups' processes.

ASPPH looks forward to assisting the membership in advancing the scholarship of teaching and learning, beginning with this first, exciting task force product.

Laura Magaña, PhD, MS
President and CEO, Association of Schools and Programs of Public Health (ASPPH)
March 2020

Executive Summary

This report is the culmination of research by the Course and Teacher Evaluation Working Group of ASPPH's Scholarship of Teaching and Learning Task Force. The working group met from January 2019 through March 2020 to fulfill its charge to report and analyze novel, promising, and effective evidence-based course and teacher evaluation methods and metrics.

As the literature yielded a paucity of research on student course and teacher evaluations (SCTEs) in academic public health and a lack of valid, evidence-based methods and metrics across all fields, the working group focused on identifying the methods and tools most commonly in use. The report also provides a context for current promising practices in the field of student course and teaching evaluations and includes recommendations and potential areas for future research and faculty development.

The report is divided into five sections:

- Section 1: The State of Student Course and Teacher Evaluations
- Section 2: Evaluation of Student Learning
- Section 3: Factors that Affect Student Course and Teacher Evaluations
- Section 4: Use of Student Course and Teacher Evaluations and Feedback Loops
- Section 5: Characteristics of Currently Used Tools

Key recommendations from the report and the relevant section in which each item resides follow:

Approaches to Student Course and Teacher Evaluations

- Recognize that SCTEs are students' opinions and perceptions about their experiences in courses but are not measures of teaching effectiveness or student learning. *(Section 3)*
- Take a broader view of student learning by incorporating future classroom and professional achievement into current assessments. *(Section 2)*
- Assess teaching effectiveness separately from the student experience in other ways, such as peer evaluation or direct observation of improvements in student work across the semester. *(Section 5)*
- Although units may be mandated to use certain tools, faculty need not limit their evaluations to these prescribed instruments and are encouraged to seek to supplement the evaluations with questions of their choosing. *(Section 5)*
- Draw on the many sources of data already collected by schools and programs of public health as part of accreditation (e.g., learning outcomes as reflected by attainment of foundational competencies, exit and alumni ratings, employer ratings) with intentional reflection on how these data demonstrate our ability to prepare future public health professionals. *(Section 1)*

Student Course and Teacher Evaluation Methods and Implementation

- Conduct a mix of learning assessments that include both student and instructor points of view, recognizing that no single tool can adequately assess the complex multivariable nature of teaching and learning. *(Section 2)*
- Measure learning at different time points throughout the course and examine the correlation of learning measures across time. *(Section 2)*
- Use theory-based SCTE instrument construction with psychometric testing. *(Section 4)*
- Adopt more comprehensive use of qualitative SCTE responses with natural language processing approaches such as sentiment analysis and factor analysis. *(Section 4)*
- Develop new SCTEs that minimize bias in responses attributed to instructors, students, and course characteristics by directly asking about the student experience. *(Sections 3 and 5)*

Advocacy for Improving Student Course and Teacher Evaluations

- Engage institutional leadership and resources to promote development, use, and feedback of appropriate and meaningful SCTE tools. *(Sections 3 and 4)*

Introduction

Approach and Assumptions

The Course and Teacher Evaluation Working Group quickly determined that within the area of student course and teacher evaluations, there exist few existing evidence-based methods and metrics. The working group determined that while the literature is largely focused on undergraduate education, undergraduate approaches and tools are similar to other levels and, thus, are equally applicable to graduate education. For example, in most institutions the same end-of-course SCTEs are used for both undergraduate- and graduate-level courses.

Since the complexity of how course and teaching evaluations are used contributes to the lack of evidence-based methods and metrics, in creating this report, the authors have sought to vet existing tools. In doing so, the authors considered important audience and contextual issues while focusing on practical solutions to complex issues, both with undergraduate and graduate courses (i.e., use of learning outcome measures, peer review, assessment of student learning). In addition, the authors identified characteristics (teacher, student, course) that could affect student course and teacher evaluations, as well as explored evaluation feedback loops. At the conclusion of the report, the authors provide a summary of student course and teacher evaluations tools currently used by a small sample of Council on Education for Public Health (CEPH)-accredited schools and programs.

Each of the five sections in the report ends with its own conclusions, recommendations, and/or future research as appropriate for the content presented.

Terminology

Varied terminology is a key contributor to the complexity evident in the current discourse about evaluations of course content, course structure, and instructional effectiveness. For example, terms such as “student course evaluations” or “students’ evaluations of teaching” fail to distinguish between the individuals providing the ratings (i.e., students) and those reviewing, evaluating, and drawing conclusions based on the ratings (i.e., school administrators) (1). Similarly, “student ratings of instruction/teaching” are restrictive as many assessment instruments also evaluate course content and structure. While no consensus exists about the most appropriate terminology, the authors in this report use the term “student course and teacher evaluation,” or SCTE, and its plural SCTEs. This term is not without its drawbacks, but it recognizes that most evaluative instruments assess student feedback on course and teacher effectiveness.

Suggestions for Dissemination

Various school and program entities are suggested as potential targets for: disseminating this report; advancing conversations that reexamine student course and teacher evaluation methods, policies, and practices; and moving the field towards more effective assessment. Specific target groups follow:

- Academic affairs committees
- Curriculum committees
- Promotion and tenure committees
- Other departmental, program, school, and university-wide entities.

In addition, attendees at the ASPPH Sections' Retreat and ASPPH Leadership Retreat may wish to review this report, discuss its implications, and consider strategies to assist and guide member schools and programs in implementing changes that could improve the effectiveness of SCTEs for teaching excellence and optimal learning in public health.

Section 1: The State of Student Course and Teacher Evaluations

Erika (Ela) Austin[1], Nicole M. Gatto[2], and George Hack[3]
[1] University of Alabama at Birmingham, [2]Claremont Graduate University, [3]University of Florida

When the authors consider student course and teacher evaluations, thoughts immediately go to the end-of-semester ritual of students filling out paper evaluations while the instructor waits outside the classroom (or, increasingly, to students completing the same task online). Indeed, instruments administered to students are the most common and frequently used tool by universities to evaluate teaching and courses (2–5). Student evaluations offer an easy and efficient way to capture information about student perceptions, though examinations of the value of student-based data range widely from reliable, valid, and useful to unreliable, invalid, and useless (6–10). The utility of student evaluations is further diminished by typically low response rates (4). There are strategies that have been shown, anecdotally, to increase response rates, including providing incentives (both at institutional and at individual course levels) and articulating to students that the feedback is used to improve/modify the course, but there are no clear evidence-based practices for addressing low response rates.

Perhaps, more critically, a singular approach to evaluating course and teacher effectiveness is indicative of several common errors that both faculty and administrators in higher education typically make when attempting to assess our educational programs, including:

Challenges in Measuring Teacher Effectiveness and Student Learning

- There is debate both about what defines teacher effectiveness and whether it can be measured (11). Student learning likely is a far easier outcome to assess, but because most faculty and administrators in higher education are not adequately trained in pedagogy and educational assessment, it is often difficult to assess overall student learning reliably (11,12).

Asking Students to Evaluate Characteristics of Instructors Rather than Measurable Behaviors/Evidence-based Practices Known to Increase Student Learning

- This focus on teacher characteristics can be traced back to Marsh (2,13,14), who emphasized the multidimensionality of effective teaching and outlined eight characteristics of effective teachers (most of which the authors would recognize from the course evaluations in use today). A number of these items reflect student assessment of characteristics that instructors may possess (enthusiasm, breadth of knowledge) rather than observable behaviors linked to effective pedagogical practices.

Reinforcing Biases, Particularly When Used in Isolation

- Further complicating the reliance on student evaluations, which focus on teacher characteristics, are the known biases that consistently result in lower evaluations for certain groups (particularly women and minorities) and that reflect other contextual characteristics (type of course, subject, etc.) distinct from teaching effectiveness (4).

Overemphasizing Student Satisfaction with Courses or Teachers

- Typical evaluations reflect student satisfaction with the course, which is distinct from teaching effectiveness and is likely to be influenced by a range of factors such as student mood, attention, motivation, etc. (15–18). Students may rate characteristics that make a course appealing, which does not equate to course effectiveness (19). Many student-centered methods to evaluate teaching were found to be unrelated to effective pedagogy (20) or effective teaching performance (21).

Relying on Students as Evaluators, Rather than Trained Educational Professionals

- The IDEA Center, a leader in higher education evaluation, advocates using the term "student ratings" rather than evaluations, as evaluation connotes a judgment about worthwhile ratings requiring interpretation within context (12). Using the term "rating" rather than "evaluation" helps to distinguish between the people who provide the information (sources of data) and those who interpret it (evaluators). Such ratings are considered, ideally, in combination with other sources of information. The IDEA Center argues that viewing student ratings as one source of data rather than as evaluations puts them into the proper perspective (12).

"…[S]tudent course and teacher evaluations cannot and should not be used or interpreted in isolation; rather, multiple forms of evaluations should be used to evaluate the quality of educational programs."

Given these limitations, how should educators evaluate course and teacher effectiveness? The key takeaway from this examination of the literature is that student course and teacher evaluations cannot and should not be used or interpreted in isolation; rather, multiple forms of evaluations should be used to evaluate the quality of educational programs. This recommendation is vital, given the consequences not only for how well academe prepares students to enter the workforce but also because the use of potentially unreliable and invalid student evaluations of courses and teachers has direct implications for faculty retention and promotion decisions (21).

The overreliance on student evaluations is not due to a lack of alternatives. More than a dozen potential sources of evidence of teaching effectiveness have been reported in the literature: [1] student ratings, [2] peer observations, [3] peer review of course materials [4] external expert ratings, [5] self-ratings, [6] videos, [7] student interviews, [8] exit and alumni ratings, [9] employer ratings, [10] mentor advice, [11] administrator ratings, [12] teaching scholarship, [13] teaching awards, [14] learning outcome measures, and [15] teaching (course) portfolios (5,22). Scholars have proposed several comprehensive models of faculty evaluation (23–28) that include multiple sources of evidence, and apply greater weight to student and peer input and less weight to self-evaluation, alumni and administrators evaluation, and others' assessments. The conclusion is clear: multiple sources should be used to develop a comprehensive understanding of teacher performance (29).

Recommendations for Assessments Beyond Student Ratings

Faculty and administrators will rightly ask how to incorporate assessment methods that move accredited schools and programs beyond student ratings. Below, the authors detail two recommended approaches: learning outcomes measures and peer feedback.

Learning Outcomes Measurements

There are multiple options to evaluate learning outcomes, all with potential strengths and limitations. Options for assessing outcomes include instructor-made measures such as tests or projects, perceived learning measures reported by students, and professional licensure and certification tests.

Instructor-made measures are the easiest to administer and interpret for a single course, in part because they are usually already administered during the course to assess students. Nevertheless, their reliability may be low as it may not be feasible to administer them at two or more points to assess learning over time. Such assessments may apply to the content covered by the course, but not necessarily relate to the teaching effectiveness. As such, the results are not generalizable to department, school, or institutional levels. The research shows that perceived learning measures do not measure actual learning by students but, instead, represent self-reports by students regarding their level of learning in the course; however, these measures have been shown to correlate with student achievement (22).

Certification tests, such as the Certified in Public Health (CPH) exam, hold a great deal of potential for the practice-based disciplines in public health and should be considered more widely to assess the overall effectiveness of both courses and teachers.

Peer Feedback

Another underused but potentially powerful form of assessment for evaluating course and faculty instructional effectiveness is peer review, which we are calling peer feedback to highlight the observations and suggestions that a well-prepared colleague could provide to a teaching faculty member. While peer feedback requires significant planning and buy-in, the insight provided into both course and teacher effectiveness is invaluable. Recommendations for developing or enhancing peer feedback programs follow (30–39):

“While peer feedback requires significant planning and buy-in, the insight provided into both course and teacher effectiveness is invaluable.”

- Effective peer feedback programs should have *well-defined and established outcomes* that are agreed upon by administrators and faculty. These outcomes guide the activities of the program and align with the institutional culture. Programs are most effective when they are tailored to the academic setting where they will be implemented.
- Oversight of successful peer feedback programs works best when centralized within the college or university, with *clearly defined administrative support and resources*. Consistent training of dedicated observers is essential, including training for how to deliver informative feedback to instructors. Successful peer feedback programs develop liaison faculty within departments and smaller academic units to promote the advantages of the program and to communicate its value to faculty.
- Optimal results occur when *well-crafted observation forms* are used by reviewers that include variations for the different learning environments and teaching contexts. Feedback forms must be adaptable to varied pedagogies such as blended learning, team-based learning, emerging technologies, etc.
- *Formalized professional learning communities* can be instrumental in getting teachers to rethink teaching practices and visioning new possibilities in the classroom. The best programs include mechanisms that promote a community of practice that encourages faculty to share helpful practices, innovative ideas or techniques, and successful solutions to instructional challenges.
- There should be a *planned, structured pre-class meeting* where the peer observer consults with the instructor about the goals for the feedback session, the instructional strategies planned to achieve them, and class challenges related to student behaviors or any other logistical obstacles.
- There should be a *planned follow-up meeting* between the peer observer and the instructor where the conversation is structured to reflect on the session and address

the comments from the observer about the effectiveness of the session, student engagement, achieved outcomes, and potential strategies for improving the instruction.
- There should be a *minimum of one fully complete session observed* in a manner that is the least obtrusive to the students and, in cases where the course is taught in components (lecture/lab), multiple observations are used to capture the complete instructional unit.
- Feedback from peer observers should take on a *formative approach* rather than a summative assessment, so that the whole process becomes focused on faculty development and the improvement of instruction and curricular activity within the program.
- Prior to any meetings or observation of sessions, *peer observers should examine the course materials*, such as the course syllabus, assignments, assessments, student evaluations, and any feedback from previous peer reviews of teaching in order to better understand the teaching development of the instructor under observation.
- A *structured peer-observation rubric is recommended* that includes multiple criteria that address student engagement, management of class logistics, demeanor and mannerisms of the instructor, organization of the lesson, quality of the selected materials, the culture of the classroom, technology integration, and the overall instructor preparation.
- A *written narrative should be prepared* that summarizes the overall effectiveness of the instructor to facilitate the mastery of the stated learning objectives, including feedback on instructor behaviors and interactions with students that appeared to contribute to learning.
- At the post-observation meeting, the peer observer could *offer suggestions for faculty development opportunities* that could enhance the instructor's teaching skills.
- Once the peer feedback process is complete, it is recommended for the *instructor to prepare a self-reflection on the peer-observation experience* and note plans for growing and developing as an instructor.

Conclusions and Recommendations

There are many approaches available for evaluating courses and teachers in higher education, yet most institutions rely solely on student ratings captured using a standardized instrument at the conclusion of each semester. While this singular approach has the benefit of simplicity – faculty select a few key objectives and possibly put aside ten minutes of class time – it is ultimately of limited use in advancing high-quality teaching and learning. Student ratings offer insight into how courses and teachers are perceived, but the evaluation of quality in education requires more effort.

Two recommended approaches for assessment methods that move schools and programs beyond student ratings include using learning outcomes measurements and peer feedback. Peer feedback programs represent the opposite extreme in terms of the faculty time and commitment required, but the insight provided by this intensive process is also likely to be far more informed by high-impact, educational practices than typical student ratings.

As a first step toward more robust evaluation of courses and teachers, the authors recommend an approach in between these two extremes, possibly drawing on the many sources of data already collected by schools and programs of public health as part of accreditation (e.g., learning outcomes as reflected by attainment of foundational competencies, exit and alumni ratings, employer ratings). The task thus becomes not one of additional data collection but of more intentional reflection on data collected for other purposes, with an eye toward how this information speaks to the ability of courses and teachers to prepare effective public health professionals. Unlike end-of-semester student course ratings, this approach puts more weight on the perspective of "experts" (those already in the field of education and/or public health) rather than "novices" (students emerging into or new to the field).

Indeed, quality education in an applied field such as public health is directly linked to students' ability to enter the workforce confidently and well-prepared. Academic public health approaches to student course and teacher evaluations would benefit significantly from leaning into this linkage and allowing it to guide more effective approaches to improving the quality of education that benefits populations and the health of the public.

Section 2: Evaluation of Student Learning

Donetta Cothran[1] and Eyal Oren[2]
[1]Indiana University, [2]San Diego State University

In this section, the authors focus on whether learning has occurred in the classroom, as opposed to the role of different assessment tools. The authors have specified four classifications determined by how and when the instructor collects the data, as described by Nilson (40): [1] indirect, end-of-course measures; [2] indirect, pre- and post-test measures, [3] direct, end-of-course-only measures, and [4] direct, pre- and post-test measures to check student learning. The authors also explore how learning may be measured through future course performance and provide some recommendations based on the current evidence.

Indirect, End-of-Course Measures

An overview of the relationship between student course and teaching evaluations (SCTEs) and learning via indirect, end-of-course measures follows. Spooren, Brockx, and Mottelmans have characterized this approach as representing the convergent validity (i.e. the strength of relationship between measures) of SCTE instruments (10).

Indirect, end-of-course assessments typically tap *students' perceived learning*. These measures are represented by standard student reports of teaching or specialized instruments developed by research teams. To a large degree, how well students perform on these assessments are framed by students' intelligence or ability to learn, prior relevant knowledge, and motivation to learn. Students' ability to judge how much they learned is also dependent on their intelligence or ability to learn (41). Additional factors include students' familiarity, or lack thereof, with an instructor's pedagogy (42).

More than three decades ago, Cohen conducted the first meta-analysis reporting that SCTE ratings correlated with student learning with r=.43, a small-to-moderate correlation (43). However, since then, the approach and methodology has been refuted. The most recent meta-analysis by Clayson concluded that SCTE ratings are not related to student learning (44). Specifically, Clayson reported that the correlation between SCTEs and learning was only .33 when correlations reported in the primary studies were averaged, regardless of the sample size and only .13 when they were weighted by the sample size.

Uttl's more recent meta-analysis showed that studies typically included a very limited number of sections, with most employing 10 or fewer sections (41). Scatterplots of SCTEs/learning correlations as a function of study size, funnel plots, and funnel asymmetry tests indicated the presence of strong small-study size effects (41). Estimated learning correlated very weakly with SCTE ratings, accounting for up to 1% of variance in

> **“…[S]tudents do not seem to learn more from professors who receive higher SCTE ratings.”**

learning/achievement measures (41). In particular, the authors noted that students do not seem to learn more from professors who receive higher SCTE ratings.

In one of the few standardized and objective studies of student perceptions of teaching and how much knowledge the students gained, Ketcham, Nigro, and Roberto examined learning as measured by the Educational Testing Service Major Field Test in Business (45). For their sample at least, there was not a strong relationship between positive SCTEs and acquired knowledge. Other problems have been noted with perceived learning — for example, students in the Wabash study failed to perceive how much they had learned as measured by several standardized tests (46,47). Nevertheless, Beleche found a weak yet positive and statistically significant association between an objective measure of student learning and student course evaluation scores (48). The authors note that it may not be prudent to rely solely on course evaluations as a means of gauging student learning. Sections with high performance on post-tests appear, on average, to receive slightly higher course evaluations. Questions that reflect learning in the course, as opposed to characteristics of the instructor, appear to better reflect estimated post-test results.

A very recent study compared students’ self-reported perception of learning with their actual learning under controlled conditions in large-enrollment introductory college physics courses taught using: [1] active instruction (following best practices in the discipline) and [2] passive instruction (lectures by experienced and highly-rated instructors). Results suggested that when students experience the increased cognitive effort associated with active learning, they initially take that effort to signify poorer learning (42). Students are less familiar with active instruction and learning techniques and, without contextualization, may not clearly understand the benefits of these approaches towards their learning despite mounting evidence that supplementing or replacing lectures with active learning strategies and engaging students in the scientific process improves learning and knowledge retention (49).

Two more rigorous student tools are provided as examples: the Student Assessment of Learning Gains survey instrument and the Transparency in Learning and Teaching survey instrument.

Student Assessment of Learning Gains

The Student Assessment of Learning Gains (SALG) survey instrument (http://www.salgsite.org) asks students to assess their learning gains in a course and the degree to which specific course components helped that learning. The items address five different facets of learning: [1] general; [2] understanding concepts; [3] acquiring skills; [4] developing positive attitudes about the course or subject matter; and [5] integrating information.

In validity testing, student scores on the SALG correlated moderately but significantly (r =.41) with student scores on the final exam. In specific topic areas, the correlations between students' SALG scores and the corresponding subsection on the final exam ranged between zero and .49 (50–52).

Transparency in Learning and Teaching

The Transparency in Learning and Teaching (TLT) survey instrument (https://tilthighered.com) aims to measure the learning impact, as perceived by students, of transparent teaching methods. The TLT starts out with three general questions: [1] how well students understand course content; [2] how accurately their submitted work reflects this understanding; and [3] the extent to which course work and course activities benefited their learning. These and the other 25 perceived-learning items offer a choice of five Likert-type responses. The next nine questions ask students how much the course has helped them acquire or refine certain skills, such as writing effectively, learning how to learn, and applying concepts to practical problems or in new situations. Another nine questions begin with the stem, "As a result of taking this course." These questions also ask students to assess their judgment about opinions different from their own, ideas in general, the reliability of sources, and the like; their confidence in their ability to succeed academically or in this field; and the likelihood of their discussing course-acquired ideas outside class and asking future instructors about the learning benefits of course components. The final four perceived-learning items address miscellaneous issues, such as whether students understand what constitutes successful work in the course and how much the instructor valued them as students.

TLT investigators have examined the effect when teachers provide two transparently-designed, problem-based, take-home assignments (compared to the unrevised, business-as-usual, take-home assignments in the comparison group) on spring-term, first-year college students' learning experiences, especially regarding students' experiences. They specifically measured the amount of transparency students perceived in the course in the following areas indicated as predictors of success: academic confidence; sense of belonging; improved mastery of skills that employers value; direct assessment of students' work as indicated by scored student work samples; and, selected randomly, short-term retention rates. In courses where students perceived more transparency as a result of receiving the transparently-designed, problem-centered, take-home assignments, they experienced significantly greater learning benefits compared with their classmates who perceived less transparency around assignments in a course. Specifically, students who received more transparency reported gains in three areas that are important predictors of students' success: academic confidence, sense of belonging, and mastery of the skills that employers value most when hiring (53).

Indirect, Pre- and Post-test Measures

Knowledge surveys are classic examples of indirect, pre- and post-test measures and are useful because their results appear to correlate with student learning (54–56). Knowledge surveys also claim the scientific legitimacy of pre- and post-test measures because the instructor administers the same survey at the beginning and end of the course. A course score representing student learning can be calculated by finding the average difference in pre- and post-test confidence ratings across items.

Knowledge surveys have the same validity problem, however, as indirect, end-of-course-only measures. When students come into a course, most of them, especially non-science introductory-level students, are overly confident about what they know and can do. At the end of the course, however, they may underestimate what they know and can do and perform better on the final exam than their knowledge survey results would predict. Wirth and Perkins recorded this pattern in at least one course (54).

Stark-Wroblewski, Ahlering, and Brill directly examined eight different sections of a psychology course taught by different instructors and compared SCTEs and scores on a pre- and post-course learning survey (57). They suggest that SCTEs and learning measures are not directly related and that instructors should develop multiple ways of documenting teaching proficiency beyond SCTEs.

Direct, End-of-Course-Only Measures

The most common end-of-course and direct measure of student learning is class grades, either by student or instructor report. This is problematic as specific course grades do not correlate necessarily with learning outcomes (58–61). A study by Canfield was designed to test reliability and validity of instructors assigning course grades, as well as to demonstrate a relationship among course grades in General Education (GE) and related major courses (62). The study used two Multi-Trait Multi-Method (MTMM)-based procedures. Results indicated statistical significance in convergent validity but not in discriminant validity (i.e., course grades could be considered as useful and meaningful assessment methods external to the particular class and instructor); this finding infers that although methods of grading may vary, grades can measure student learning outcomes directly. The two MTMM studies showed support by convergent and discriminant validity for course grades. The findings address the perceived limitations of course grades based upon concerns of the validity and reliability of grades. The instructors noted, “if we continue to establish these levels of validity and reliability in class grading, we can assume a greater level of confidence in the use of course grades for discipline and program (e.g., GE or discipline) assessment.”

Given the role that competencies play in multiple accredited academic programs of many types, the relationship between SCTEs and essential learning outcomes (ELOs) has rarely been examined. A recent mixed-methods study explored experiences of students, faculty, and professionals as they engaged in a project focused on teaching and learning with ELOs. The researchers relied on the use of descriptive statistics and interpretive phenomenological analysis to capture faculty and professional staff participants' engagement in a collaborative study group designed to support pedagogy integrating ELOs into courses, assignments, and co-curricular activities and to gauge students' experienced competence as a result. Using a pre- and post-student survey of self-perceived ELO competence, researchers found a difference in student ratings of 0.25, ($p \leq 0.05$), suggesting an increase in students' perception of their ELO competence over the course of the semester (63).

In learning environments that encourage and assess student understanding (rather than recall), students achieving higher-quality learning outcomes are found to adopt a deeper approach to learning, such as a search for meaning, relevance, application, and relations to other experience (64,65).

Direct, Pre- and Post-test Measures

Direct, pre- and post-test measures attempt to account for students' pre-course knowledge when assessing the end-of-course knowledge. They entail an assessment at the beginning of the course, the pre-test, and the same or very similar assessment at the end, the post-test.

For example, in physics, Lee showed normalized gains on the Force Concept Inventory (FCI) as a valid, reliable, and objective measure of student learning (66). Normalized gain is the proportion of improvement on an instrument from pre- to post-instruction compared to the maximum possible improvement of the class. In this study, participating faculty (n=15) were given a list of multiple-choice survey instruments and asked to select an appropriate instrument to give to their students at the beginning and end of the semester so that an objective measure with a valid and reliable instrument could be used to make claims about student learning in the course. The research team used the matched student data when available (about 80% of the courses). Participating faculty also submitted the results of their student evaluations of instruction, which was unique to each instructor. Students' average rating of the quality of instruction was graphed against the FCI normalized gain for each class. The coefficient of determination for the data was just R^2=0.01544, suggesting there was no correlation between pre- and post-test evaluation results and end-of-course student evaluations.

A variation of the pre- and post-test measure approach compares student success in an earlier class to success in an upper-level course. A small number of investigations have examined student performance in future classes as an assessment of student learning. For example, Carrell and West suggest that a longer-term view of student learning might be

needed since the initial level of student learning in introductory classes is not always a good predictor of success in future classes (67). It is possible that instructors' short-term approaches to learning habits may improve their own evaluations as instructors but harm students' long-term achievement. Weinburg, Hashimoto, and Fleisher found that SCTEs are positively related to current grades in a class, but are not as well-linked to learning, in part because students are often not aware of how much they have learned in a class (68).

Conclusions and Recommendations

Findings on both subjective and objective measures are mixed. Some researchers report the SCTE-learning relationship is not significant (e.g., (57,69) while others suggest a small but positive relationship exists (e.g.; (44,70). It may be that both perspectives are correct, and the varied results reflect the complicated nature of the teaching and learning setting, SCTEs, and analyses used. For example, Clayson's meta-analysis suggests that the relationship between SCTEs and learning does exist, but it is situational and dependent on the teacher, academic discipline, level of instruction, and the level of objectivity of the assessment tool (44). Gailbraith, Merrill, & Kline offer similar support for this complicated relationship in their finding that the most effective teachers (as measured by a standardized test) fall within the middle percentiles of SCTE scores (71).

Recommendations for public health faculty, based on the current evidence, follow:

- Conduct a mix of learning assessments that include both student- and instructor points of view. Although various studies examined employ a wide variety of tools in a broad range of SCTEs, nearly all conclude with similar advice to educators: no single tool can adequately assess the complex multivariable nature of teaching and learning.
- Consider measuring learning at different time points throughout the course and examining the correlation of learning measures across time
- Take a longer-term perspective on student learning that allows for multi-semester and/or post-graduation evaluations of key concepts and skills.
- Provide students with clear instruction and an explanation of how active and transparent approaches may improve their learning.

"Take a longer-term perspective on student learning that allows for multi-semester and/or post-graduation evaluations of key concepts and skills."

Section 3: Factors that Affect Student Course and Teacher Evaluations

Ali A. Weinstein[1] and Marie Diener-West[2]
[1]George Mason University, [2]Johns Hopkins University

In this section, the authors briefly review the factors that have been shown to affect student course and teacher evaluations (SCTEs). As with the other sections, this approach is not intended to provide a comprehensive review of the literature, but rather highlight some of the important factors that have been shown to affect SCTEs with exemplar references. In addition, the authors highlight areas that deserve more attention in the literature and hypothesize that there are other factors that also systematically affect SCTEs. The review draws from prior, published papers utilizing experimental designs examining both undergraduate and graduate courses.

Numerous prior findings have identified generally weak or negative associations between SCTEs and instructor effectiveness, as measured by student performance or grades (41,72–76). Therefore, it is important to understand the factors affecting SCTEs that are most likely not related to instructor effectiveness. In fact, there is evidence that these factors (biases) can be so large that more effective teachers get lower SCTEs than less effective teachers (73). These factors are categorized into: [1] instructor characteristics; [2] student characteristics; [3] course characteristics; and [4] institutional characteristics.

Instructor Characteristics

Various characteristics of instructors may affect SCTEs. Many of these factors are protected characteristics under the law (e.g., gender, race, age) and others are quite variable across instructors. Sources of these biases include demographic characteristics: instructor gender, with female instructors rated lower (73,76–81); instructor attractiveness, with more attractive instructors rated more highly (82–84); instructor race, with minority instructors rated lower (78,85–87); and whether the instructor speaks with an accent versus native speaker, with instructors having an accent rated lower (88).

The association between instructor age and SCTEs is equivocal although younger instructors have often been rated more highly (77,79,89,90). However, the rank of the faculty instructor (i.e., assistant professor, associate professor) does not appear to have an effect on SCTEs (91). Other characteristics include personality and behavioral traits: instructor likeability, with more likeable or supportive instructors receiving higher evaluations (92,93), and instructor expressiveness, with more charismatic or enthusiastic instructors rated higher (82,93,94).

In addition, instructors with strong organizational skills (vis-à-vis lecture materials, stating objectives, providing overviews) are rated more highly (93). Teaching assistants (when present) may be rated more highly than the faculty instructor (91). However, it is unclear whether the faculty member's rating changes when there is a teaching assistant or is influenced by other aspects of the course or course organization.

The authors were not able to identify literature addressing other instructor characteristics, such as team-taught instruction that may be especially important in the field of public health.

Student Characteristics

Student characteristics also have been shown to affect SCTEs. These characteristics can be classified as two types: [1] perceptions and expectations; and [2] demographic and background characteristics. Within the area of student perceptions and expectations, the student's expected grade is related to SCTEs, with a positive correlation between expected grade and student ratings (2,73,89,95–98). There have been various reasons purported for this association, with two main hypotheses. One is the grade leniency hypothesis, which basically implies that instructors can buy better evaluations by giving higher grades; the second is the validity hypothesis, which asserts that effective instructors cause students to work harder, thereby students earn higher grades from effective instructors (90). The leniency hypothesis has received a great deal of attention in the literature (2,99,100), which shows that stricter grading may lead students to rate the instructor lower on components related to grading fairness, as well as in areas such as instructor attitude to students (99). Also, this source of bias may be enhanced when the SCTE is administered near a recent assessment (100).

Within the area of student demographic and background characteristics, gender has been shown as important. Female students tend to rate instructors more positively, regardless of the gender of the instructor (101). The impact of student age and ethnicity on student ratings has received far less empirical attention. There are not enough well-controlled studies to assess if there are systematic effects of student age and race on SCTEs (89,102–104). More advanced students (i.e., seniors compared to freshmen) rate courses more positively than newer students (105). The student's prior interest level in the topic (or whether the course is in the student's major or department) has been shown to be positively associated with SCTEs (89,92,95,106–108) and is by far the most well-researched topic in this area.

Course Characteristics

In addition to instructor and student characteristics, there are factors inherent in the course itself that can affect SCTEs. These include the subject of the course material (i.e., instructors who teach courses with mathematical content tend to get lower ratings) (109), the level of the course and whether the course is required (110), the course format (111), and the size of the

course (112). For example, instructors of general education courses receive lower ratings than instructors of elective courses (113). Humanities, art, and social sciences instructors typically receive higher ratings on student evaluations than instructors of math and science courses (104,114,115). Instructors of early morning classes receive lower ratings than those of courses taught later in the day (116).

Another important course characteristic is the method of delivery (online versus traditional in-person courses). There is some evidence that courses taught by the same instructor using the same course content are rated lower when delivered in the online modality (117) with the same result for hybrid courses receiving lower SCTEs, in spite of similar student outcomes, as compared to traditional courses (118).

Other factors that have received less attention in the literature but may be important (particularly in public health), are the association between SCTEs and community-based, service-learning, or practicum courses. In addition, the pedagogical approach used (i.e., flipped classrooms, active learning strategies) may also affect SCTEs. Although these innovative approaches may improve student learning, it cannot be assumed that there is a corresponding positive effect on SCTEs. Further research needs to be completed to understand the effects of the type of course and the pedagogical approach used in the course and the subsequent effect on SCTEs.

Institutional Characteristics

SCTE ratings can also be influenced by the characteristics of the institution with respect to educational resources and priorities, emphasis on technology in instruction, and overall value on teaching effectiveness and student performance. All these variables depend on the university, school, and/or program leadership's commitment to the educational mission. Poor or outdated physical condition of the classroom reflects a lack of emphasis or importance and is associated with lower SCTE ratings (119). Lack of technology-related professional development for instructors may result in lower or suboptimal usage of technological teaching aids (120) that can be negatively associated with instruction and SCTE ratings.

Departments of economics were surveyed on how they used SCTE ratings, peer evaluation, and other measures of teaching effectiveness and scholarship when evaluating instructor promotion and merit increases (121). Although most departments used SCTE ratings, the survey demonstrated that their value remained unclear as to how these findings were used by instructors to modify their courses or by administrators in evaluating instructors (121). It is possible that SCTE ratings are potentially biased if and when students observe that instructors do not incorporate their suggestions.

The type or amount of value messaging that institutions provide regarding SCTEs are associated with response rates. Factors associated with non-response have been attributed to courses outside of a student's major or interest area, lower grades earned in a course, and survey fatigue (122). Communicating the importance of SCTEs to both instructors and students is important for soliciting valuable feedback from all students (122). Investment by the school to incentivize the completion of SCTEs can decrease both non-response and bias in findings.

Conclusions and Recommendations

There are various factors, associated with the instructor, student, course, or institution, that can systematically affect the value of SCTEs. Therefore, it is important that these negative factors are recognized and minimized as much as possible. New research is beginning to demonstrate a path forward with SCTEs that may minimize potential biases. An interesting recent investigation demonstrated that gender bias in SCTEs could be mitigated by utilizing an instrument with explicit language intended to reduce the gender bias (123).

"New research… demonstrated that gender bias in SCTEs could be mitigated by utilizing an instrument with explicit language intended to reduce the gender bias (123)."

SCTEs have become an accepted component of the higher education sector. Future research should focus on ways to improve the current assessment tools used for SCTEs by acknowledging that SCTEs do not measure teaching effectiveness but, instead, measure student opinion (124). Opinions include perceptions that can be based on expertise, knowledge, and logical reasoning, but they can also be based on bias or prejudice (conscious or unconscious). Therefore, future research needs to create SCTEs that capture and value student opinion, minimize biases in response, and provide useful feedback to both the instructor and institution.

Key recommendations follow:

- Recognize that SCTEs represent students' opinions and perceptions about their experiences in courses but are not measures of teaching effectiveness.
- Develop new SCTEs that minimize biases in response attributed to instructor, student, and course characteristics.
- Deploy institutional leadership and resources to promote development, use, and feedback of appropriate and meaningful SCTE tools in accredited schools and programs of public health.

"Deploy institutional leadership and resources to promote….meaningful SCTE tools in accredited schools and programs of public health."

Section 4: Use of Student Course and Teacher Evaluations and Feedback Loops

Delia Lang[1] and Elizabeth Stuart[2]
[1]Emory University, [2]Johns Hopkins University

In this section, the authors provide an overview of the use of student course and teacher evaluations (SCTEs) by different constituents, including students, faculty, and program and school administrators, as well as accrediting bodies. The authors identify common challenges and limitations to the use of SCTEs that can result in problematic feedback loops and conclude with suggestions on how SCTE users might (and should or should not) use the results.

Use of Course and Teacher Evaluations

Students, faculty, school administrators, and accrediting agencies use feedback from SCTEs in various ways (125).

- *Students* often have access to SCTE ratings from previous academic terms, which may inform their course and/or faculty selection when options are available. While SCTEs typically provide both course and faculty ratings, students tend to make course selections primarily based on faculty ratings rather than ratings related to course content or structure (126).
- *Faculty* use SCTEs as feedback loops in the short-term for formative purposes to improve their own teaching, applying findings to specific existing courses or to identify needs for additional or substantially modified courses that could meet varied student audiences.
- *Directors of academic programs* may use SCTEs to understand student feedback trends, compare types of courses (e.g., across departments, online vs. onsite), and/or assess how different groups of students evaluate different courses (e.g., undergraduate vs. graduate students or those for whom the course is required vs. optional).
- *School administrators* often use SCTEs in ways similar to directors of academic programs, and sometimes rely on SCTEs as information loops for summative purposes to inform promotion, tenure, merit/salary adjustments, and contract renewals of faculty. To that end, and to the degree that faculty mobility across academic institutions throughout their careers may partially rest on a record of strong teaching performance, faculty have an interest in SCTEs in the long term as well.
- Lastly, *the Council on Education for Public Health (CEPH)*, the accrediting agency for schools and programs in public health, requires institutions to describe processes used to assess faculty instructional effectiveness including SCTEs. The criteria focus both on student satisfaction with instructional quality, which relates directly to formative

purposes of SCTEs, as well as the role these findings play in summative decisions relating to faculty advancement.

Challenges

In September 2019, the American Sociological Association (ASA) released a statement that was endorsed by 17 additional scholarly associations declaring aspects of the current practice of SCTEs as "problematic" (127). As noted by the ASA, and reviewed in several other publications, when used in isolation, SCTEs can produce inaccurate and invalid results at best and perpetrate biases against gender and ethnic minority faculty at worst (41,73,125). The authors briefly outline four key areas of concern in the use of SCTEs: [1] reporting and interpretation of scores; [2] validity; [3] bias; and [4] the student approach, outlined below.

Reporting and Interpretation of Scores

Significant flaws have been identified in reporting and interpreting SCTEs. Mean ratings across SCTE items are reported in aggregate form and are often interpreted in the absence of statistical distributions to contextualize the scores. Average ratings are frequently derived from SCTE items designed at the categorical level of measurement. Finally, low response rates are often overlooked and small differences in mean scores are overemphasized (124,128).

Validity

An underlying assumption supporting the use of SCTEs is that the resulting scores are valid indicators of teaching effectiveness. As noted in Section 2 of the report, Evaluation of Student Learning, mounting evidence does not support this assumption. Both observational and experimental research supports the conclusion that SCTEs are heavily influenced by factors unrelated to teaching effectiveness, including instructor personality, student characteristics, student grade expectations, and student motivation for learning (41,73,125). Other contextual factors impacting the validity of scores include academic discipline, class level (undergraduate v. graduate), class size, meeting times, and whether the course is required vs. elective (1,41,70,129–131). Section 3, Factors that Affect Student Course and Teacher Evaluations, discusses these and other factors that affect SCTEs.

Bias

As described in Section 3 of this report, Factors that Affect Student Course and Teacher Evaluations, SCTEs have consistently been found to show bias against a broad range of instructor characteristics, most notably gender and ethnicity. Women are often rated lower than males, especially in what is typically considered a male-dominated discipline, while instructors of color also received lower scores compared to their white colleagues, especially

when evaluated by white students. Consequently, reliance on SCTEs as a primary measure of teaching effectiveness is likely to present an additional barrier for women and minority faculty to promotion (70,76,132).

The Student Approach

Some evidence suggests that students are not always motivated to complete SCTEs thoughtfully, perhaps in part because requests to complete multiple evaluations seem burdensome at a busy time of the semester. Additionally, students may not know or may not trust that their feedback is truly valued or could inform change. They may also feel that even if changes were made, the improvements would not impact them directly (133–135).

> **"…[A]dopting holistic approaches to the evaluation of teaching effectiveness can provide more balanced feedback and create stronger teaching cultures."**

Despite these challenges, SCTEs represent an efficient approach to obtaining student feedback and therefore, continue to be used across academic institutions, including across schools and programs in public health. The emerging consensus, informed by a strong evidence base, suggests that adopting holistic approaches to the evaluation of teaching effectiveness can provide more balanced feedback and create stronger teaching cultures. Below, the authors summarize evidence-based and promising practice recommendations that must be implemented across three areas: [1] improving the utilization of existing SCTE processes; [2] improving the analysis and presentation of results; and [3] augmenting existing processes with holistic evaluation methods to assess teaching effectiveness.

Conclusions and Recommendations

Improving the Utilization of Existing SCTE Processes

- Shift cultural norms around evaluation to create a culture that is welcoming of ongoing, constructive evaluation. Inspiring an evaluative culture, where feedback is expected for continuous growth and improvement is important on multiple fronts:
  - Encourage students to take evaluations seriously, know how the evaluations are used, and are given tips for providing constructive feedback (examples provided in Bandy, 2019 (136)). As part of this approach, encourage high response rates. Some schools provide incentives for students to complete evaluations, such as allowing access to grades more quickly if evaluations are completed.

    - Give faculty tools for interpreting evaluation results and include guidance on methods to handle positive and negative feedback and means to use it for improvement. Some faculty elect not to read their reviews in part because of a perception of lack of constructive feedback. These instructors would benefit from help to develop strategies to interpret their evaluations and opportunities to connect with resources, such as centers for teaching and learning to put feedback suggestions into practice.
    - Ensure leadership knows how to interpret the results appropriately and to view individual evaluations within a broader context; one resource is Linse (1). Holistic evaluation methods are discussed further below.
- Clarify the target(s) for measurement by renaming SCTEs to validate the importance of student feedback related to course/instructor satisfaction, but not to imply that feedback is a true measure of teaching effectiveness. The instrument name and items should reflect this view, per the ASA statement (127).

Improving the Analysis and Presentation of SCTE Results

- Conduct descriptive data analyses to understand the correlations between course characteristics and evaluations within an institution.
- Observe and document evaluation across various types of course, considering the context for the courses. Key factors include, for example: online vs. onsite; foreign language speakers v. native English speakers; course difficulty; and whether the course is required or optional.
- Probe the extent to which SCTEs may contain evidence of implicit bias.
- Improve reporting on methodology, providing distributions, sample size, and response rates for each (124) to contextualize individual ratings.

Augmenting SCTEs with Holistic Evaluation Methods

- Explore the role of timing of the feedback, such as mid-term v. end-of-term evaluations. Mid-term evaluations can be particularly useful for mid-course "corrections" of instructional methods, and thus may be of more direct use to individual faculty teaching specific courses. Snooks, Neeley, and Revere showed that in a pilot study, students participating in mid-term feedback opportunities reported increases in their learning and satisfaction (137). Schools should make mid-term evaluation opportunities widely available to help facilitate mid-course evaluations.
- Use mixed methods evaluation approaches. End-of-term focus groups with students and/or instructors across courses, pre- and post-course student surveys and other structured procedures (e.g. Small Group Instructional Diagnosis) can provide more nuanced and constructive feedback regarding teaching effectiveness (138). These approaches necessitate additional class time and would need to be carefully planned.

- Adopt peer or expert feedback of teaching. Studies suggest that faculty do make changes to their courses based on constructive peer/expert review (139–141). Structures should be considered when using peers v. outside experts to observe instruction, including considerations for whether feedback is to be used for formative and/or summative purposes, and the delivery of constructive and potentially sensitive feedback. This approach is addressed more fully in Section 1, Student Course and Teacher Evaluations, of this report.
- Encourage instructor reflections, portfolios, and professional development. A teaching portfolio allows a faculty member to describe one's teaching role holistically, and to present one's philosophy and performance with respect to teaching, beyond just the SCTEs. Seldin, Miller, and Seldin (2010) provide models of teaching portfolios and concrete advice to use them (142).

"A teaching portfolio allows a faculty member to present one's philosophy and performance with respect to teaching, beyond just the SCTEs."

Future Research

There is still more work to be done to ensure that SCTEs are used in ways that could benefit students, faculty, and administrators. This research could include:

- Theory-based instrument construction with psychometric testing, and
- More comprehensive use of qualitative responses, such as through qualitative methods and natural language processing approaches, a promising example of which is sentiment analysis, as conducted by Newman, et al. (143). Sliusarenko et al. illustrate the use of factor analysis to help find comments that are outliers and to summarize the most important points of students' satisfaction or dissatisfaction (144).

Section 5: Characteristics of Currently Used Tools

Matt Hayat[1] and Laura Liang[2]
[1]Georgia State University, [2]Rutgers University

In order to assess the tools that are currently used by Council on Education for Public Health (CEPH)-accredited schools and programs, the authors disseminated an informal poll to ASPPH Academic Affairs Section representatives to assess the delivery format, evaluation system, and the questions used to assess their courses and the teachers. Respondents were invited to respond with samples of SCTE surveys used at their institutions. This section summarizes the results, highlighting the similarities and differences that currently exist in a small sample of responding CEPH-accredited schools and programs, termed units, and the various approaches to student course and teacher evaluations (SCTEs).

Data Source

Eighteen (18) CEPH-accredited units reported on the SCTE system in place in their institution.

Delivery Format

All 18 units use an online delivery format for SCTE.

Evaluation System

Across responding units, three reported using eXplorance Blue, two reported using EvaluationKit, two reported using CoursEval (through CampusLabs), and 11 reported using 11 different SCTE systems, such as in-house systems.

Standardization

All units reported standardized core questions asked across course and disciplines. In addition to the core question in their SCTEs, eight units reported some degree of instructor control with the option to select additional, customized questions. In a few cases, instructors could select additional questions from a closed SCTE of optional items. Five units reported an option for their instructors to add additional questions of their own wording.

Question Type, Scale, and Structure

Of the 18 units that provided information about their SCTE systems, 11 submitted an example of the SCTE survey used in their institution. The most common type of question included on the SCTE surveys were Likert-type questions using a 5-point scale. The predominant scale included “strongly disagree, disagree, neutral, agree and strongly agree” items. Most submitted SCTE surveys included three sections: questions about the course, questions about the instructor, and open-ended questions.

Most Common Areas for Questions about the Course

- Assessments/assignments (10/11 surveys included at least one question related to course assessment/assignments)
- Quality Overall (9/11 surveys included at least one question related to the overall quality of the course)
- Objectives (6/11 surveys included at least one question related to course objectives)
- Organization (5/11 surveys included at least one question related to the course as well-organized)
- Feedback (5/11 surveys included at least one question related to course feedback)
- Materials (5/11 surveys included at least one question related to course materials)

Most Common Areas for Questions about the Instructor

- Overall effectiveness/teaching ability (10/11 surveys included at least one question related to the overall effectiveness/teaching ability of the instructor)
- Promotion of critical thinking (7/11 surveys included at least one question related to whether the instructor promoted/encouraged critical thinking/thought)
- Clear communication (6/11 surveys included at least one question related to whether the instructor communicated clearly)
- Accessibility (6/11 surveys included at least one question related to whether the instructor was accessible outside of classroom time)
- Inclusive classroom (5/11 surveys included at least one question related to whether the instructor created an inclusive/respectful classroom)

Conclusions and Recommendations

The findings indicate areas that overlap in the use of SCTEs in CEPH-accredited schools and programs, specifically around the format of SCTEs and the categories of questions used. However, as was discussed in Section 3, Factors that Affect Student Course and Teacher Evaluations, the questions used are subject to various biases. More research is needed on ways to improve SCTEs to mitigate this bias.

One potential area to improve SCTEs is the restructuring of items included in typical SCTEs. Items pertaining to teaching effectiveness, course effectiveness, course organization, and course relevance are particularly susceptible to bias and therefore should be avoided (124). Unfortunately, many of the responses to the authors' survey included these types of biased

> **"SCTEs should focus on student reporting on their own experiences."**

questions. Instead, SCTEs should focus on student reporting on their own experiences, such as: was the workload greater than or less than that of other courses, whether the student has greater or less interest in the subject after taking the class, did the student learn new knowledge and skills from the course, etc. Such questions would ask directly about the student experience. In addition, as reported by some units already, schools and programs are urged to explore options for instructors to be able to add additional, customized questions to the institutionally prescribed SCTE.

“Schools and programs are urged to explore options for instructors to be able to add additional, customized questions to the institutionally prescribed SCTE.”

Teaching effectiveness should then also be assessed in other ways, such as peer observation or direct observation of the improvement of student work across the semester, areas that were covered in Sections 1: The State of Student Course and Teacher Evaluations and 2: Evaluation of Student Learning. These data. furthermore, could be incorporated into appropriate feedback loops as discussed in Section 4: Use of Student Course and Teacher Evaluations and Feedback Loops of this report.

While this research did not yield enough information to create a comprehensive matrix of evaluation methods and tools, Horigian’s research on peer review measures (both processes and tools) derived from ASPPH-member schools and programs and partner institutions is posted at https://populationhealthexchange.org/teph-peer-review/ (145). These data are suggested as a helpful reference and may serve as the foundation for future, deeper exploration of issues in peer observation of teaching effectiveness.

Gaps for Further Exploration

The working group identified issues beyond the scope of their charge that represent gaps for potential exploration by ASPPH, including:

- Evidence-based approaches to improve low SCTE response rates
- Methods for supplementing mandated evaluation systems with additional customized questions and evaluation tools
- Means of mitigating bias in SCTEs
- Improving SCTEs using theory-based instrument construction with psychometric testing and more comprehensive use of qualitative responses, among other, potential strategies
- Addressing the challenges of using a variety of different assessment methods
- Providing strategies on advocating for SCTE modifications at the institutional level
- Supporting faculty development, for example by conducting workshops for skill-building on comprehensive evaluation strategies/initiatives.

References

1. Linse AR. Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. Stud Educ Eval. 2017 Sep 1;54:94–106.

2. Marsh HW. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential baises, and utility. J Educ Psychol. 1984;76(5):707–54.

3. Onwuegbuzie A, Witcher A, Collins KM, Filer J, Wiedmaier C, Moore C. Students' Perceptions of Characteristics of Effective College Teachers: A Validity Study of a Teaching Evaluation Form Using a Mixed-Methods Analysis. Am Educ Res J. 2007 Jan 1;44.

4. Vargas-Madriz LF, Nocente N, Best-Bertwistle R, Forgie S. "Somebody has to teach the 'broccoli' course": Administrators Navigating Student Evaluations of Teaching (SET). Can J High Educ. 2019 Apr 21;49(1):85–103.

5. Berk RA. Survey of 12 Strategies to Measure Teaching Effectiveness. Int J Teach Learn High Educ. 2005;17(1):48–62.

6. Aleamoni LM. Student ratings of instruction. In: Millman J, editor. Handbook of teacher evaluation. Beverly Hills, CA: Sage; 1981. p. 110–45.

7. Grammatikopoulos V, Linardakis M, Gregoriadis A, Oikonomidis V. Assessing the students' evaluations of educational quality (SEEQ) questionnaire in Greek higher education. High Educ. 2014;70(3):395–408.

8. Khong TL. The validity and reliability of the student evaluation of teaching. Int J Innov Educ Res. 2014;2(9):57–63.

9. Drysdale MJ. Psychometric Properties Of Postsecondary Students' Course Evaluations [PhD Thesis]. Utah State University; 2010.

10. Spooren D P, Brockx, B, Mortelmans. On the Validity of Student Evaluation of Teaching: The State of the Art. Rev Educ Res. 2013;83(4):598–642.

11. Grant F. Cross National Comparisons Of Excellence In University Mathematics Instructors: An Analysis of Key Characteristics of Excellent Mathematics Instructors based on Teacher Evaluation Forms [PhD Thesis]. Columbia University; 2014.

12. Benton S, Cashin W. Student ratings of teaching: A summary of research and literature. IDEA Pap. 2011;50:1–22.

13. Marsh HW. Seeq: A Reliable, Valid, and Useful Instrument for Collecting Students' Evaluations of University Teaching. Br J Educ Psychol. 1982;52(1):77–95.

14. Marsh HW. Students' evaluations of University teaching: Research findings, methodological issues, and directions for future research. Int J Educ Res. 1987 Jan 1;11(3):253–388.

15. Zumbach J, Funke J. Influences of mood on academic course evaluations. Pract Assess Res Eval. 2014;19(4):1–12.

16. Chen Y, Hoshower LB. Student evaluation of teaching effectiveness: An assessment of student perception and motivation. Assess Eval High Educ. 2003;28(1):71–88.

17. Uijtdehaage S, O'Neal C. A curious case of the phantom professor: Mindless teaching evaluations by medical students. Med Educ. 2015;49(9):928–32.

18. Brown GD, Wood AM, Ogden RS, Maltby J. Do student evaluations ofuniversity reflect inaccurate beliefs or actual experience? A relative rank model. J Behav Decis Mak. 2015;28(1):14–26.

19. Chonko LB, Tanner JF, Davis R. What are they thinking? Students' expectations and self-assessments. J Educ Bus. 2002;77(5):271–281.

20. Myers CB, Myers SM, Stewart T, Nynas S. Institutional Policies on Assessment of Pedagogy and Faculty Classroom Practices: Evidence from 4-Year Colleges and Universities in the United States. High Educ Policy. 2015;28:315–32.

21. Lau T, Wines W. Observations on the Folly of Using Student Evaluations of College Teaching for Faculty Evaluation, Pay, and Retention Decisions and Its Implications for Academic Freedom. William Mary J Women Law. 2006;13:167–202.

22. Berk RA. Should Student Outcomes Be Used to Evaluate Teaching? J Fac Dev. 2014;28(2):87–96.

23. Arreola RA. Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system (2nd ed.). Bolton, MA: Anker; 2000.

24. Braskamp LA, Ory JC. Assessing faculty work. San Francisco: Jossey-Bass; 1994.

25. Centra JA. Reflective faculty evaluation:Enhancing teaching and determining faculty effectiveness. San Francisco: Jossey-Bass; 1999.

26. Keig LW, Waggoner MD. Collaborative peer review: The role of faculty in improving college teaching (ASHE/ERIC Higher Education Report, No. 2). Washington, DC: Association for the Study of Higher Education; 1994.

27. Romberg E. Description of peer evaluation within a comprehensive evaluation program in a dental school. Instr Eval. 1985;8(1):10–6.

28. Soderberg LO. A credible model: Evaluating classroom teaching in higher education. Instr Eval. 1986;8:13–27.

29. Sayavedra M. Teacher Evaluation. ORTESOL J. 2014;31:1–9.

30. Hyland KM, Dhaliwal G, Goldberg AN, Chen L, Land K, Wansley M. Peer review of teaching: Insights from a 10-year experience. Med Sci Educ. 2018;28(4):675–81.

31. Vescio V, Ross D, Adams A. A review of research on the impact of professional learning communities on teaching practice and student learning. Teach Teach Educ. 2008;24:80–91.

32. White K, Boehm E, Chester A. Predicting academics' willingness to participate in peer review of teaching: A quantitative investigation. High Educ Res Dev. 2014;33(2):372–85.

33. Barnard A, Nash R, McEvoy K, Shannon S, Waters C, Rochester S, et al. LeaD-In: A cultural change model for peer review of teaching in higher education. High Educ Res Dev. 2015;34(1):30–44.

34. Newman LR, Brodsky DD, Roberts DH, Pelletier SR, Johansson A, Jr VCM, et al. Developing expert-derived rating standards for the peer assessment of lectures. Acad Med. 2012;87(3):356–63.

35. Martin GA, Double JM. Developing Higher Education Teaching Skills Through Peer Observation and Collaborative Reflection. Innov Educ Train Int. 1998;35(2):161–70.

36. Carroll C, O'Loughlin D. Peer observation of teaching: Enhancing academic engagement for new participants. Innov Educ Teach Int. 2014;51:446–56.

37. Hammersley-Fletcher L, Orsmond P. Evaluating our peers: Is peer observation a meaningful process? Stud High Educ. 2004;29:489–503.

38. Hendry GD, Oliver GR. Seeing is believing: the benefits of peer observation. J Univ Teach Learn Pract. 2012;9(1).

39. Shortland S. Feedback within peer observation: Continuing professional development and unexpected consequences. Innov Educ Teach Int. 2010;47:295–304.

40. Nilson LB. Measuring Student Learning to Document Faculty Teaching Effectiveness. Improve Acad. 2013;32(1):287–300.

41. Uttl B, White CA, Gonzalez DW. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. Stud Educ Eval. 2017 Sep;54:22–42.

42. Deslauriers L, McCarty LS, Miller K, Callaghan K, Kestin G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc Natl Acad Sci. 2019 Sep 24;116(39):19251–7.

43. Cohen PA. Student Ratings of Instruction and Student Achievement: A Meta-Analysis of Multisection Validity Studies. Rev Educ Res. 1981;51(3):281–309.

44. Clayson DE. Student Evaluations of Teaching: Are They Related to What Students Learn?: A Meta-Analysis and Review of the Literature. J Mark Educ. 2009 Apr 1;31(1):16–30.

45. Ketcham DC, Nigro PJ, Roberto M. Student Perceptions and Learning Outcomes: Evidence from the Education Testing Service (ETS) Major Field Test in Business [Internet]. Rochester, NY: Social Science Research Network; 2017 Aug [cited 2019 Nov 19]. Report No.: ID 3025598. Available from: https://papers.ssrn.com/abstract=3025598

46. Bowman NA, Hill PL. Measuring how college affects students: Social desirability and other potential biases in college student self-reported gains. New Dir Institutional Res. 2011;2011(150):73–85.

47. Porter SR. Do college student surveys have any validity? Rev High Educ J Assoc Study High Educ. 2011;35(1):45–76.

48. Beleche T, Fairris D, Marks M. Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. Econ Educ Rev. 2012 Oct 1;31(5):709–19.

49. Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, DeHaan R, et al. Scientific Teaching. Science. 2004 Apr 23;304(5670):521–2.

50. Schraw G. Promoting general metacognitive awareness. Instr Sci. 1998 Mar 1;26(1):113–25.

51. Zimmerman BJ. Becoming a Self-Regulated Learner: An Overview. Theory Pract. 2002 May 1;41(2):64–70.

52. Zimmerman BJ. Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In: Schunk DH, Zimmerman BJ, editors. Self-regulated learning: From teaching to self-reflective practice. New York, NY, US: Guilford Publications; 1998. p. 1–19.

53. Winkelmes M, Bernacki M, Butler J, Zochowski M, Golanics J, Weavil K. A Teaching Intervention that Increases Underserved College Students' Success. Peer Rev. 2016 Jan 1;18.

54. Wirth KR, Perkins D. Knowledge Surveys [Internet]. National Association of Geoscience Teachers Workshops; 2008 [cited 2019 Nov 19]. Available from: https://serc.carleton.edu/NAGTWorkshops/assess/activities/wirth-perkins.html

55. Wirth KR, Perkins D. Knowledge Surveys: An Indispensable Course Design and Assessment Tool. Innov Scholarsh Teach Learn. 2005;1–12.

56. Nuhfer E, Knipp D. The Knowledge Survey: A Tool for All Reasons. Improve Acad. 2003;21(1):59–78.

57. Stark-Wroblewski K, Ahlering RF, Brill FM. Toward a More Comprehensive Approach to Evaluating Teaching Effectiveness: Supplementing Student Evaluations of Teaching with Pre-Post Learning Measures. Assess Eval High Educ. 2007 Aug;32(4):403–15.

58. Maki PL. Assessing for Learning: Building a Sustainable Commitment Across the Institution. Sterling, VA: Stylus Publishing; 2012. 356 p.

59. Penn JD. Assessing Complex General Education Student Learning Outcomes Number 149. 1 edition. San Francisco: Jossey-Bass; 2011. 142 p.

60. Rogers G. Do Grades Make the Grade for Program Assessment? [Internet]. Accreditation Board for Engineering and Technology; 2003. Available from: https://www.abet.org/wp-content/uploads/2015/04/do-grades-make-the-grade.pdf

61. Suskie L. Using assessment results to inform teaching practice and promote learning. In: Joughin G, editor. Assessment, Learning and Judgement in Higher Education [Internet]. Dordrecht, Netherlands: Springer; 2009 [cited 2019 Nov 19]. p. 1–20. Available from: https://www.springer.com/gp/book/9781402089046

62. Canfield ML, Kivisalu TM, Van Der Karr C, King C, Phillips CE. The Use of Course Grades in the Assessment of Student Learning Outcomes for General Education. SAGE Open. 2015 Oct 1;5(4):2158244015615921.

63. Cydis S, Galantino M, Hood C, Padden M, Richard M. Integrating and Assessing Essential Learning Outcomes: Fostering Faculty Development and Student Engagement. J Scholarsh Teach Learn. 2015 Jun 5;15.

64. Trigwell K. Evidence of the Impact of Scholarship of Teaching and Learning Purposes. Teach Learn Inq ISSOTL J. 2013;1(1):95–105.

65. Trigwell K, Ashwin P, Millan ES. Evoked prior learning experience and approach to learning as predictors of academic achievement. Br J Educ Psychol. 2013 Sep;83(Pt 3):363–78.

66. Lee LJ, Connolly ME, Dancy MH, Henderson CR, Christensen WM. A comparison of student evaluations of instruction vs. students' conceptual learning gains. Am J Phys. 2018 Jun 19;86(7):531–5.

67. Carrell SE, West JE. Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors [Internet]. National Bureau of Economic Research; 2008 Jun [cited 2019 Nov 19]. Report No.: 14081. Available from: http://www.nber.org/papers/w14081

68. Weinberg BA, Hashimoto M, Fleisher BM. Evaluating Teaching in Higher Education. J Econ Educ. 2009 Jul 1;40(3):227–61.

69. Mohanty G, Gretes J, Flowers C, Algozzine B, Spooner F. Multi-Method Evaluation of Instruction in Engineering Classes. J Pers Eval Educ. 2005 May 1;18(2):139–51.

70. Spooren P, Brockx B, Mortelmans D. On the Validity of Student Evaluation of Teaching: The State of the Art. Rev Educ Res. 2013 Oct 29;83:598–642.

71. Galbraith CS, Merrill GB, Kline DM. Are Student Evaluations of Teaching Effectiveness Valid for Measuring Student Learning Outcomes in Business Related Classes? A Neural Network and Bayesian Analyses. Res High Educ. 2012 May 1;53(3):353–74.

72. Carrell SE, West JE. Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. J Polit Econ. 2010 Jun 1;118(3):409–32.

73. Boring A, Ottoboni K, Stark P. Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. Sci Res [Internet]. 2016 Jan 7 [cited 2019 Nov 6]; Available from: https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

74. Braga M, Paccagnella M, Pellizzari M. Evaluating students' evaluations of professors. Econ Educ Rev. 2014;41(C):71–88.

75. Johnson VE. Grade Inflation: A Crisis in College Education [Internet]. New York: Springer-Verlag; 2003 [cited 2019 Nov 6]. Available from: https://www.springer.com/gp/book/9780387001258

76. MacNell L, Driscoll A, Hunt AN. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. Innov High Educ. 2015 Aug;40(4):291–303.

77. Arbuckle J, Williams BD. Students' Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations. Sex Roles. 2003 Nov 1;49(9):507–16.

78. Basow SA, Codos S, Martin JL. The effects of professors' race and gender on student evaluations and performance. Coll Stud J. 2013;47(2):352–63.

79. Bianchini S, Lissoni F, Pezzoni M. Instructor characteristics and students' evaluation of teaching effectiveness: evidence from an Italian engineering school. Eur J Eng Educ. 2013 Mar 1;38(1):38–57.

80. Boring A. Gender biases in student evaluations of teaching. J Public Econ. 2017 Jan 1;145:27–41.

81. Mengel F, Sauermann J, Zölitz U. Gender Bias in Teaching Evaluations. J Eur Econ Assoc. 2019 Apr 1;17(2):535–66.

82. Ambady N, Rosenthal R. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. J Pers Soc Psychol. 1993;64(3):431–41.

83. Wolbring T, Riordan P. How beauty works. Theoretical mechanisms and two empirical applications on students' evaluation of teaching. Soc Sci Res. 2016 May;57:253–72.

84. O'Reilly MT. Relationship of physical attractiveness to students' ratings of teaching effectiveness. J Dent Educ. 1987 Oct;51(10):600–2.

85. Bavishi A, Madera JM, Hebl MR. The effect of professor ethnicity and gender on student evaluations: Judged before met. J Divers High Educ. 2010;3(4):245–56.

86. Reid LD. The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.Com. J Divers High Educ. 2010;3(3):137–52.

87. Smith BP, Hawkins B. Examining Student Evaluations of Black College Faculty: Does Race Matter? J Negro Educ. 2011;80(2):149–62.

88. Subtirelu N. "She does have an accent but…": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. Lang Soc. 2015 Feb 1;44:35–62.

89. Wachtel HK. Student evaluation of college teaching effectiveness: A brief review. Assess Eval High Educ. 1998 Jan 1;23(2):191–212.

90. Worthington AC. The Impact of Student Perceptions and Characteristics on Teaching Evaluations: A case study in finance education. Assess Eval High Educ. 2002 Jan 1;27(1):49–64.

91. Blackhart GC, Peruche BM, DeWall CN, Joiner Jr. TE. Factors influencing teaching evaluations in higher education. Teach Psychol. 2006;33(1):37–9.

92. Feistauer D, Richter T. Validity of students' evaluations of teaching: Biasing effects of likability and prior subject interest. Stud Educ Eval. 2018 Dec 1;59:168–78.

93. Erdle S, Murray HG, Rushton JP. Personality, classroom behavior, and student ratings of college teaching effectiveness: A path analysis. J Educ Psychol. 1985;77(4):394–407.

94. Williams WM, Ceci SJ. "How'm I Doing?" Problems with Student Ratings of Instructors and Courses. Change Mag High Learn. 1997 Sep 1;29(5):12–23.

95. Marsh HW, Cooper TL. Prior Subject Interest, Students' Evaluations, And Instructional Effectiveness. Multivar Behav Res. 1981 Jan 1;16(1):83–104.

96. Vasta R, Sarmiento RF. Liberal grading improves evaluations but not performance. J Educ Psychol. 1979;71(2):207–11.

97. Marsh HW, Dunkin MJ. Students evaluations' of university teaching: A multidimensional perspective. In: Smart JC, editor. Higher Education: Handbook of Theory and Research. New York: Agathon Press; p. 143–223.

98. Braskamp LA, Ory JC. Assessing Faculty Work: Enhancing Individual and Institutional Performance. San Francisco: Jossey-Bass Inc; 1994.

99. Chacko TI. Student ratings of instruction: A function of grading standards. Educ Res Q. 1983;8(2):19–25.

100. Nimmer JG, Stone EF. Effects of grading practices and time of rating on student ratings of faculty performance and student learning. Res High Educ. 1991 Apr 1;32(2):195–215.

101. Dornan T, Arno M, Hadfield J, Scherpbier A, Boshuizen H. Student evaluation of the clinical 'curriculum in action.' Med Educ. 2006;40(7):667–74.

102. Anderson KH, Siegfried JJ. Gender Differences in Rating the Teaching of Economics. East Econ J. 1997;23(3):347–57.

103. Watkins D. Student evaluations of university teaching: A cross-cultural perspective. Res High Educ. 1994 Mar 1;35(2):251–66.

104. Centra JA. Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness. The Jossey-Bass Higher and Adult Education Series. Jossey-Bass Inc; 1993.

105. Chae SJ, Kim M, Chang KH, Chung Y-S. Potential bias factors that affect the course evaluation of students in preclinical courses. Korean J Med Educ. 2017 Jun;29(2):73–80.

106. Prave RS, Baril GL. Instructor Ratings: Controlling for Bias from Initial Student Interest. J Educ Bus. 1993 Aug 1;68(6):362–6.

107. Howard GS, Maxwell SE. Correlation between student satisfaction and grades: A case of mistaken causation? J Educ Psychol. 1980;72(6):810–20.

108. Feldman KA. Consistency and Variability among College Students in Rating Their Teachers and Courses: A Review and Analysis. Res High Educ. 1977;6(3):223–74.

109. Uttl B, White CA, Morin A. The Numbers Tell It All: Students Don't Like Numbers! PLOS ONE. 2013 Dec 16;8(12):e83443.

110. Marsh HW, Roche LA. Making Students' Evaluations of Teaching Effectiveness Effective. Am Psychol. 1997;52:1187–97.

111. Lake DA. Student performance and perceptions of a lecture-based course compared with the same course utilizing group discussion. Phys Ther. 2001 Mar;81(3):896–902.

112. Bedard K, Kuhn P. Where class size really matters: Class size and student ratings of instructor effectiveness. Econ Educ Rev. 2008 Jun 1;27(3):253–65.

113. Patrick CL. Student evaluations of teaching: effects of the Big Five personality traits, grades and the validity hypothesis. Assess Eval High Educ. 2011 Mar 1;36(2):239–49.

114. Feldman KA. Course characteristics and college students' ratings of their teachers: What we know and what we don't. Res High Educ. 1978 Sep 1;9(3):199–242.

115. Cashin WE. Student Ratings of Teaching: The Research Revisited. IDEA Paper No. 32. 1995 Sep [cited 2019 Nov 6]; Available from: https://eric.ed.gov/?id=ED402338

116. Koushki PA, Kuhn HAJ. How reliable are student evaluations of teachers? Eng Educ. 1982;72(3):362–7.

117. Marzano MP, Allen R. Online vs. Face-to-Face Course Evaluations: Considerations for Administrators and Faculty. Online J Distance Learn Adm. 2016;19(4).

118. Chingos MM, Griffiths RJ, Mulhern C, Spies RR. Interactive Online Learning on Campus: Comparing Students' Outcomes in Hybrid and Traditional Courses in the University System of Maryland. J High Educ. 2016 Dec 27;88(2):210–33.

119. Hill M, Epps K. The Impact of Physical Classroom Environment on Student Satisfaction and Student Evaluation of Teaching in the University Environment. Acad Educ Leadersh J. 2010 Jan 1;14(4):65–79.

120. O'Dwyer LM, Russell M, Bebell DJ. Identifying Teacher, School and District Characteristics Associated with Elementary Teachers' Use of Technology:A Multilevel Perspective. Educ Policy Anal Arch. 2004 Sep 14;12(0):48.

121. Becker WE, Bosshardt W, Watts M. How Departments of Economics Evaluate Teaching. J Econ Educ. 2012 Jul 1;43(3):325–33.

122. Adams MJD, Umbach PD. Nonresponse and Online Student Evaluations of Teaching: Understanding the Influence of Salience, Fatigue, and Academic Environments. Res High Educ. 2012 Aug 1;53(5):576–91.

123. Peterson DAM, Biederman LA, Andersen D, Ditonto TM, Roe K. Mitigating gender bias in student evaluations of teaching. PLOS ONE. 2019 May 15;14(5):e0216241.

124. Stark P, Freishtat R. An Evaluation of Course Evaluations: "F." Sci Res [Internet]. 2014 Sep 29 [cited 2019 Nov 6]; Available from: https://www.scienceopen.com/document?id=ad8a9ac9-8c60-432a-ba20-4402a2a38df4

125. Lang DL, Walker ER. Innovations in Evaluating and Valuing Public Health Teaching: The Challenge of Course Evaluations. In: Sullivan LM, Galea S, editors. Teaching Public Health. 1 edition. Baltimore: Johns Hopkins University Press; 2019.

126. Beran T, Violato C, Kline D, Frideres J. The Utility of Student Ratings of Instruction for Students, Faculty, and Administrators: A "Consequential Validity" Study. Can J High Educ. 2005 Jun 30;35(2):49–70.

127. American Sociological Association. Reconsidering Student Evaluations of Teaching [Internet]. 2019 [cited 2019 Nov 19]. Available from: https://www.asanet.org/press-center/press-releases/reconsidering-student-evaluations-teaching

128. Boysen GA. Significant interpretation of small mean differences in student evaluations of teaching despite explicit warning to avoid overinterpretation. Scholarsh Teach Learn Psychol. 2015;1(2):150–62.

129. Beran T, Violato C, Kline D, Frideres J. What do students consider useful about student ratings? Assess Eval High Educ. 2009 Oct 1;34(5):519–27.

130. Denson N, Loveday T, Dalton H. Student evaluation of courses: What predicts satisfaction? High Educ Res Dev. 2010 Aug 1;29(4):339–56.

131. Donnon T, Delver H, Beran T. Student and teaching characteristics related to ratings of instruction in medical sciences graduate programs. Med Teach. 2010;32(4):327–32.

132. Mitchell KMW, Martin J. Gender Bias in Student Evaluations. PS Polit Sci Polit. 2018 Jul;51(3):648–52.

133. Sojka J, Gupta AK, Deeter-schmelz DR. Student and Faculty Perceptions of Student Evaluations of Teaching: A Study of Similarities and Differences. Coll Teach. 2002 Apr 1;50(2):44–9.

134. Chen Y, Hoshower LB. Student Evaluation of Teaching Effectiveness: An assessment of student perception and motivation. Assess Eval High Educ. 2003 Jan 1;28(1):71–88.

135. Powell NJ, Rubenstein C, Sawin EM, Annan S. Student evaluations of teaching tools: a qualitative examination of student perceptions. Nurse Educ. 2014 Dec;39(6):274–9.

136. Bandy J. Student Evaluations of Teaching [Internet]. Vanderbilt University. 2010 [cited 2019 Nov 19]. Available from: https://wp0.vanderbilt.edu/cft/guides-sub-pages/student-evaluations/

137. Snooks MK, Neeley SE, Revere L. Midterm Student Feedback: Results of a Pilot Study. J Excell Coll Teach. 2007;18(3):55–73.

138. Lang DL, Walker ER, Steiner RJ, Woodruff RC. Implementation and Mixed-Methods Evaluation of Team-Based Learning in a Graduate Public Health Research Methods Course. Pedagogy Health Promot. 2018 Jun 1;4(2):140–50.

139. Brickman P, Gormally C, Martella AM. Making the Grade: Using Instructional Feedback and Evaluation to Inspire Evidence-Based Teaching. CBE—Life Sci Educ. 2016 Dec 1;15(4):ar75.

140. Vian T, Ashigbie PG. Accelerating Educational Innovation in the MPH Degree Program: What Is the Role of Peer Review of Teaching? J Health Educ Teach. 2015;6(1):43–56.

141. Benton SL, Young S. Best Practices in the Evaluation of Teaching. IDEA Pap. 2018;69:1–18.

142. Seldin P, Miller JE, Seldin CA, McKeachie W. The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions. 4 edition. San Francisco: Jossey-Bass; 2010. 416 p.

143. Newman H, Joyner D. Sentiment Analysis of Student Evaluations of Teaching. In: Penstein Rosé C, Martínez-Maldonado R, Hoppe HU, Luckin R, Mavrikis M, Porayska-

Pomsta K, et al., editors. Artificial Intelligence in Education. Cham: Springer International Publishing; 2018. p. 246–50. (Lecture Notes in Computer Science).

144. Sliusarenko T, Clemmensen L, Ersbøll B. Text mining in students' course evaluations: Relationships between open-ended comments and quantitative scores. In: Proceedings of the 5th International Conference on Computer Supported Education (CSEDU-2013). 2013. p. 564–73.

145. Horigian, Vivana. Population Health Exchange: Teaching Excellence in Public Health [Internet]. Boston University; 2018 [2020 Mar 10]. Available from: https://populationhealthexchange.org/teph-peer-review/

ASPPH Scholarship of Teaching and Learning (SoTL) Task Force Course and Teacher Evaluation Working Group (2019-2020)

Chair:

- Ali A. Weinstein, PhD (George Mason University College of Health and Human Services)

Members:

- Erika (Ela) Austin, PhD, MPH (University of Alabama at Birmingham School of Public Health)
- Donetta Cothran, PhD, MAE (Indiana University School of Public Health - Bloomington)
- Marie Diener-West, PhD (Johns Hopkins University Bloomberg School of Public Health)*
- Nicole Gatto, MPH, PhD (Claremont Graduate University Master of Public Health Program)
- George Hack, PhD, MEd (University of Florida College of Public Health and Health Professions)
- Matt Hayat, PhD (Georgia State University School of Public Health)
- Delia Lang, PhD, MPH (Emory University Rollins School of Public Health)
- Laura Liang, DrPH (Rutgers University School of Public Health)
- Eyal Oren, PhD, MS (San Diego State University School of Public Health)
- Elizabeth Stuart, PhD (Johns Hopkins University Bloomberg School of Public Health)
- Shan Mohammed, MD, MPH, FAAFP (Northeastern University Department of Health Sciences), *ex officio*, chair, ASPPH Education Advisory Committee and chair, ASPPH Scholarship of Teaching and Learning Task Force

ASPPH Staff:

- Elizabeth Weist, MA, MPH, CPH

*indicates an ASPPH teaching award winner