Report and recommendations on student evaluations of teaching

As submitted by the Ad Hoc committee on student evaluations of teaching
Kate Ziemer, Robin Hackett (2021), Barbara Prudhomme White (current), co-chairs

March 23, 2022

Preparation of this Report

Background

This committee was established on 2/22/21 (Senate Motion XXV_02_22_21) and charged with revising teaching evaluations and providing written guidance defining the role of student surveys of teaching and learning in demonstrating teaching effectiveness for the purpose of annual reviews, renewal, promotion, and tenure decisions. We were asked to consider how student surveys of teaching and learning should interact with other methods for evaluating teaching effectiveness, to consider whether guidance should apply across all faculty types, and to be sure these guidelines conformed to current union contract guidelines. Deliverables were to include a new tool to replace the current Student Evaluation of Teaching (SET), and an implementation plan responsive to varying conditions across the university. We were asked to aim for a spring 2022 launch of the new tool.

Process

The committee spent early work sessions completing a comprehensive search into the literature regarding best practices and issues in course **Student Evaluation of Teaching** (SET). The committee identified approximately 20 recent articles that identify and address central issues related to teaching evaluation. The articles note common weaknesses of student evaluations of teaching and contain suggestions for strengthening processes for gathering meaningful, objective feedback from students as well as suggestions for how course evaluations should be considered when used for promotion/hiring/retention decisions. Our committee subsequently branched off into two working groups. The *Tool* group drafted a student survey and explained the logic behind it. The *Use* group described the role of the surveys in promotion, tenure, and retention processes, and recommended methods for increasing response rates. We note that, after much consideration, we came to consensus about how best to focus our labor and accomplish our ends. Rather than try to invent a perfect instrument, we 1) assembled a good-enough instrument based on best practices and UNH needs, and 2) carefully designed specific guidelines and about how the tool can be productively used by all constituents: students, faculty, chairs, administrators.

Research

The committee identified three primary issues relating to student evaluations of teaching: 1) The impact of student bias (race, gender, age, etc.); 2) the use of evaluations in promotion/hiring/retention of faculty decisions; 3) low completion rates, particularly now that they are online.
Specific findings include:

- Studies consistently show bias against women, especially younger women, POC, LGBTQ and disabled populations, and in favor of instructors perceived to be more attractive.

These effects seem to vary by discipline (Lee & Benton, 2017; Kreitzer et al., 2021a, 2021b)
- Students more highly rank instructors of their own gender and race (Kreitzer et al., 2021a, 2021b)
- There is no correlation between instructors ranking and measured student learning (Uttl et al., 2016; Esarey & Valdes 2020; Flaherty 2020; ASA 2019)
- There is positive correlation between students' grades and evaluation scores (Stroebe 2016)
- There is positive correlation between lighter workload and higher scores
- Traditional SET may punish best pedagogical practices. Group or teamwork, peer review, and engaged and active learning receive lower scores.
- Qualitative courses are rated higher than quantitative courses, all else equal (Uttl et al., 2013)
- Students that value SET rate instructors more highly (Spooren & Christiaens, 2017)
- "Evaluations are shaped by discipline, student interest, class level, class difficulty, class meeting time, and other course-specific characteristics, but not generally actual instructor quality" (Kreitzer et al., 2021b) "and they can be influenced by course characteristics like time of day, subject, class size, and whether the course is required, all of which are unrelated to teaching effectiveness" (ASA 2019). Instructor scores can also be improved by bringing cookies to class on evaluation day (Hessler et al., 2018).

Given the literature review, the committee confirmed that these issues are a national phenomenon that many institutions are attempting to address.

Of note, the committee recognizes that there have been several Senate committees over at least the past 8-10 years that have been convened with similar charges. The 2017 senate Ad Hoc committee report (see posted report in OneDrive file) submitted on 5/1/2017 is especially compelling in its thorough research and similar findings. The previous committee's report provides a supportive independent perspective from a different group of faculty that delivered a very similar report as this one, and highlighted nearly identical issues in the literature. Their report resulted in a Senate motion presented below (with friendly amendments included), that passed in favor with 54 votes, 3 votes opposed, and 4 abstentions (Motion XXI M18 date May 8 2017). The 2021-22 Ad Hoc committee extended prior committee work through an updated literature review as well as developing a new instrument for broad application across the institution. The current committee also developed clear recommendations and guidelines for application of the new instrument.

5_1_2017
Motion - That the Faculty Senate support development of a formative teaching assessment process that embodies the goal of continual improvement for all faculty, and which provides earlier and more productive engagement of students in this process within each course. The Senate Agenda Committee will implement assignments to appropriate Senate standing or ad hoc committees. That process should include the recommended actions that follow.

1) That a faculty senate committee, working with the Center for Excellence and Innovation in Teaching and Learning, develop specific recommendations and guidance for instructors regarding early-semester course assessment including professional development on engaging students in that process.
2) That the Center for Excellence and Innovation in Teaching and Learning develop instructional recommendations for students on how to provide constructive professional feedback.
3) That a faculty senate committee, working with the Center for Excellence and Innovation in Teaching and Learning, develop specific recommendations and guidance regarding peer- assessment practices.
4) That a faculty senate committee, working with the Center for Excellence and Innovation in Teaching and Learning and the Provost's office, develop specific recommendations and guidance regarding documentation of teaching activity and effectiveness by portfolio.
5) That a faculty senate committee, working with the Center for Excellence and Innovation in Teaching and Learning and the Provost's office, articulate a set of components of quality teaching, building on the work of the Study Committee 2016-2017.
6) That end-of-course student input be called surveys and not evaluations.
7) That a faculty senate committee take an entirely fresh look at student end-of-course survey prompts to develop a set (numerical and narrative) that provides more insight regarding student perceptions of learning, instructor behaviors, and course learning conditions. In other words, engage students in commenting on what may have affected their learning.
8) That the Center for Excellence and Innovation in Teaching and Learning, working with faculty senate representatives, develop a workshop and set of resources regarding the literature on the validity, bias, and reliability of numerical course surveys, including use of historical UNH data. This workshop may be deployed for Deans, Chairs, and all faculty.

2022 Committee Recommendations

I. Recommended principles

Principle 1: Minimize Bias

To reduce as much bias as possible, several experts suggest that questions should focus on student experiences and opportunities to give feedback rather than questions that infer evaluating teaching effectiveness (e.g., American Sociological Association, 2019; Kreitzer & Sweet-Cushman, 2021). Given that students often do not hold accurate perceptions of their own learning or progress in the classroom – that is, that they often have underdeveloped metacognition (Ehrlinger & Shain, 2014), students may not always be in the best position to give accurate assessments of teaching effectiveness. For example, when students are asked to provide a quantitative rating to a questionnaire item such as "the instructor's teaching methods facilitated my learning," (Bowdoin College Evaluation Questionnaire), how students conceptualize "facilitated my learning" is unclear. One way that students may make this summary judgment is by the grade they receive—higher grades imply greater learning. In addition, some teaching methods that have been shown empirically to facilitate learning, such as taking frequent quizzes (e.g., Roediger, Putnam, & Smith, 2011), are unpopular and may be rated accordingly.

Felkey and Batz-Barbarich (2021) noted that some degree of bias will be inherent in any course evaluation. Acknowledging this bias in promotion decisions is important especially in fields where the faculty member is of a different gender or racial background than the majority of other faculty and students.

Finally, Gravestock, and Gregor-Greenleaf (2008) suggested that students benefit from information on both how to provide useful feedback and how course evaluations are used.

Principle 2: Use student experiences of learning surveys as supplements to other evidence of effective teaching

There is consensus in the literature that course teaching evaluations or student feedback should be only one of many sources of evidence in teaching effectiveness, such as peer review, participation in scholarly activity focused on improvement of teaching, self-reflection, teaching portfolios, etc., (American Sociological Association, 2019; National Research Council, 2003; Kreitzer & Sweet-Cushman, 2021). In addition, evaluators such as chairs, deans, etc. should be informed on the value of SELs in the context of additional sources for evidence of effective teaching (American Sociological Association, 2019; Hanover Research). Adjusting the ways in which SET are used is the primary mechanism for addressing potential bias.

Principle 3: Make efforts to improve student response rates

A key point regarding student response rates and vulnerability to wide variability based on outliers suggests that we need to implement strategies to increase student response rates (Kreitzer & Sweet-Cushman (2021). A threshold response rate of at least two thirds or 66.6% of the course enrollment is recommended (Gravestock & Greg-Greenleaf, 2008). To improve response rates, faculty can provide time during class to complete a brief survey, inform students of the usefulness of their responses for future courses, and give specific examples of how faculty have used feedback in prior semesters (Hoel & Dahl, 2019). We recommend that students be provided with some sort of incentive to fill out surveys. For instance, the date for the release of all grades might be set for January 15; students can obtain their grades released earlier by filling out a survey. Another suggestion is to reward students who fill out surveys with access to aggregate data from those surveys. These data will have value to students for class selection.

II. Recommended structure for survey instrument (tool)

The Committee recommends the following key elements of a new student course feedback survey implemented through the BLUE TEV system.

The proposed instrument asks students to reflect on their learning experiences, rather than rate their professors' behaviors and/or strategies. The 4-part instrument will be named "Student Experiences of Learning" (SEL) and combine qualitative and quantitative questions as well as standardized and customizable segments. The instrument is appropriate for use across all instructional formats, faculty types, and across disciplines, as the survey tools includes broad questions as well as optional sections customizable by instructors, departments, and colleges.

Part 1: Standardized questions

Part one of our proposed instrument is a standardized set of 7 questions administered to students in all courses. It includes five open-ended narrative (qualitative) questions and two universally

applicable Likert Scale and yes/no quantitative questions. These questions were developed from sources of best practice questions that have been tested as statistically valid. The questions ask students about meaningful aspects of their own learning experience rather than asking them to rate/evaluate in ways that requires expertise they do not always have. In addition to improving the value of the response for faculty, there is pedagogical value: student's positive perceptions of their learning and critical reflection on the learning environment are linked to their actual learning and the effectiveness of the learning environment. Questions are designed to be relevant to all course types and disciplines. Furthermore, the instrument does not ask students for an overall quantitative rating. Eliminating the overall quantitative rating (a) reduces the risk that student responses will be used disproportionately in the assessment of teaching for reappointment/promotion decision; and (b) allows faculty to try innovative and/or new pedagogical approaches without concern about impacting the "overall rating" value. The results of this first section will be shared with department heads, administrators, and committees.

We understand the practical desire many have for multiple quantitative questions. While quantitative measures can give quick views of how faculty may be performing, research has identified biases in these questions, as well as the increased chance that participants interpret the question differently than intended (Kreitzer, et al, 2021b). The possibility of bias is not eliminated by using rating scales. However, the instrument is designed to deter using student responses to compare individual faculty to each other or to a department average, and to encourage use in documenting trends for individual instructors over time (ASA, 2019).

Part 2: Optional portions of the instrument

a) Optional instructor-level feedback questions

An optional section that instructors can use to solicit additional feedback on their teaching, and/or students' perceptions of their learning. Faculty who elect to include a part two can select several rating-scale items targeted to their individual professional goals through a database of statistically tested questions, or (if deemed feasible) by adding their own questions. [Note that the ability to add free text questions may not be practical, depending on the software used.]

b) Optional department/college-level questions

An optional section for the department/college to determine up to 5 common questions across a department/college.

Importantly, colleges, programs, and departments can customize questions to the instrument, while instructors can also customize their questions and retain control of data from their own questions. Results of Part 2a customized questions are sent only to the instructor, who can use it both to receive candid feedback on their strategies and approaches and to gather support for their cases for promotion or tenure. The results may be shared with departments for developmental or evaluative purposes at the discretion of the professor.

c) Optional Discovery course questions

An optional section allowing up to 5 Discovery Course questions that will be used to gather university-wide student perception of their learning of general education learning outcomes.

Additionally, additional sections could be used by programs/departments who want to incorporate survey data for program assessments or external accreditation and a separate section for Discovery student learning outcomes to facilitate General Education assessment purposes.

NOTE: For optimal student response, a maximum number of survey questions should be maintained. We recommend no more than 5 questions for each of the optional sections outlined above (Part 2a, b, c above).

III. Recommended guidelines for use of survey instrument (tool)

The Tool Use group reviewed evidence about the objectives of gathering student feedback on courses, best practices for collecting, interpreting, and utilizing such feedback for a variety of purposes, and guidelines for how to prepare students to engage in this process. In doing so, the group identified two fundamental principles guiding the use of student feedback.

First, the purpose of such feedback is to provide formative feedback for the instructor, and second, feedback is valuable as a contributing, not single, piece of information regarding teaching effectiveness. As such, we organized our findings and recommendations into three main areas: guidance for instructors, guidance for administrators and peer review committees (including reappointment, promotion, tenure, post-tenure, analogous reviews for CCLEAR faculty, and reviews for any teaching faculty of record), and guidance for students.

Concerns about these processes that the guidelines are intended to address include the following: student unconscious bias towards instructors of various identities; validity concerns related to low response rates, over-reliance on summary quantitative measures in organizational assessment/promotion processes to the exclusion of other types of data, and the unstated assumption that students can meaningfully evaluate teaching. To address this last point, it is our recommendation that all such data be re-named as Student Perceptions OR Experiences of Learning as they are really data about student perceptions.

GUIDELINES FOR INSTRUCTORS

- Consider implementing a mid-term survey to 1) convey the value of student experience; 2) modify course as appropriate, which shows students that their experience matters and can make the course a better experience for others; 3) create a culture of assessment/responsive feedback, which might incentivize students to complete the end of term surveys. Share with students how and why you made changes (if that was the case)
- Explain to students how understanding their experience in this course is important to you and to them, and how you have used such feedback in the past to make changes to your course/teaching.

- Accompany student course survey information with reflections on what faculty learn from the student ratings/comments and how they have made/will make appropriate modifications to their teaching in response to that feedback; compile this in teaching portfolios to demonstrate the iterative process of improving their teaching as an additional measure of teaching effectiveness.
- Tips for maximizing response rates: Arrange time in class (beginning or middle of your class) for students to complete the survey. Inform students in advance to bring devices to class. Provide time after class before the deadline in case they need to use a computer elsewhere.

GUIDELINES FOR STUDENTS

- Be constructive and specific in your comments
- Comment from your own experience in the course—use 'I' statements to describe your experience
- Think about the course/unit/chapter learning objectives, and comment on *how* the course was taught in ways that helped you to meet those objectives
- Think about and comment on what will help improve the course for future students

GUIDELINES FOR USAGE BY DEANS, P&T COMMITTEES, PROMOTION COMMITTEES, AND AGGREGATE PROGRAM ASSESSMENTS

The committee recommends that each college develop college-wide practices based on student feedback of teaching as a part of several methods for evaluating teaching and teaching development. This recommendation includes encouraging faculty and administrators to embrace a perspective of collecting student feedback on their experiences in courses rather than evaluating faculty performance. This perspective embraces the research that suggests that students are best able to provide personal experiences about what supports their learning thereby help faculty better understand effective teaching strategies, rather than provide reliable evaluative information about individual instructors. In addition to using the proposed instrument, several additional metrics of teaching effectiveness grounded in best-practices are encouraged, including:

a. peer review
b. early-to-mid-term student feedback for in-course adjustments (Templates available through CEITL)
c. self-reflection on teaching philosophy
d. self-study of student performance

- When factored into reviews of teaching, Student Experiences of Learning information should be contextualized as much as possible (e.g., by cross-referencing responses to related questions and/or considering answers to qualitative as well as quantitative questions.) As noted above, the new standardized Student Experiences of Learning instrument will not ask students to provide overall ratings of instructors, the goal being to counteract over-relying on a single question/score in isolation from others.
- Use only as part of larger, broader assessment of teaching process such as peer observation, peer review of syllabi & materials, faculty self-reflection, etc.
- Use only when minimum response rates have been met (literature suggests 65%), and the course enrolls a minimum of 10 students.

- Identify situations (e.g., very large course sections) in which sampling of qualitative comments from large data sets may be acceptable – sample from quartiles based on student's ratings responses. This may minimize sampling error issues
- Provide guidance on how to responsibly interpret data.
- Use aggregate data to demonstrate college and institutional effectiveness of teaching at UNH.

IV. Recommended process for pilot trial

The committee recommends applying the instrument for a 3-year trial during which data will be collected systematically at several levels (e.g., UNH Survey Center, ET&S, AAUP, colleges, departments, faculty, students) to determine the instruments' performance, efficacy, and utility for collecting meaningful information about courses and teaching practices. Pilot steps include:

Step 1. Summer 2022: First roll out to include the common questions section and Discovery optional questions.

Step 2. Fall 2022: End of fall term 2022 will include common questions section and remaining optional questions sections.

Step 3. Spring 2023: (1) continue rollout and (2) engage Survey Center to help understand efficacy of common questions and suggest any needed wording modifications.

Step 4. Summer 2023: Rollout full survey instrument with any modifications to common questions from Survey center.

Step 5. Over 2023-2025 collect information at the end of each semester (fall, J-term, spring, summer) and analyze survey effectiveness.

Step 6. Fall 2024 – Spring 2025: Assess instrument performance and convene an Ad Hoc committee in fall 2024 to assess all data and information about how well the instrument performed in providing information that is useful for students in giving feedback, and in providing meaningful information for P&T and contract reviews and faculty development, with summary report and recommendations due spring 2025.

We recommend that the faculty senate, both student senates (undergraduate and graduate), and administration work together within and across colleges to increase student survey response. An initial first step includes an information and communication campaign, as well as collecting student feedback on the pilot instrument. If unsuccessful in increasing response rates, then other process points (e.g., delaying grade release) should be considered by an ad hoc committee.

Committee Membership

Beemer, Lawrence
Brown, Stephanie
Congalton, Russell
Conrad, Tammy
Collins, Karen

Hackett, Robin (co-chair, 2020-21)
Hall, Caroline
Holtrop, Maurik
Lehmann, Jennifer
Lahiri, Smita
Christine O'Keefe (joined 3/2022)
Overson, Catherine
Parker, Victoria
Petty, Nadine
Pillemer, David
Pimpare, Stephen
Rock, Kylee & Christian Lipovsky
Root, Kelly
Smith, Maddy
Sparrow, Sophie
White, Barbara Prudhomme (co-chair, 1/2022 -)
Ziemer, Kate (co-chair)

Selected References

American Sociological Association (2019). *Statement on Student Evaluations of Teaching* https://www.asanet.org/files/asa-statement-student-evaluations-teaching-sept52019pdf

Ehrlinger, & Shain, (2014) In V.A. Benassi, C.E. Overson, & C.M. Hakala (Eds.). *Applying science of learning in education: Infusing psychological science into the curriculum*. Retrieved from the Society for the Teaching of Psychology web site: http://teachpsych.org/ebooks/asle2014/index.php

Felkey, A. J., & Batz-Barbarich, C. (2021). Can Women Teach Math (and Be Promoted)? A Meta-analysis of Gender Differences across Student Evaluations of Teaching. *AEA Papers & Proceedings, 111*, 184–189.

Gravestock P. & Gregor-Greenleaf, E. (2008). *Student Course Evaluations: Research, Models and Trends*. Toronto: Higher Education Quality Council of Ontario.

Hanover Research: *Best Practices in Reducing Bias in Student Evaluation of Teaching* (date unknown)

Hoel, A., & Dahl, T. I. (2019). Why bother? Student motivation to participate in student evaluations of teaching. *Assessment & Evaluation in Higher Education, 44*(3), 361–378. https://doi-org.unh.idm.oclc.org/10.1080/02602938.2018.1511969

Kreitzer, R. J., & Sweet-Cushman, J. (2021). Evaluating Student Evaluations of Teaching: a Review of Measurement and Equity Bias in SETs and Recommendations for Ethical Reform. *Journal of Academic Ethics*, 1 12.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research, 9*(3), 231–251. https://doi-org.unh.idm.oclc.org/10.1007/s10984-006-9015-7

Li, D., Benton, S. L., & IDEA Center. (2017). The Effects of Instructor Gender and Discipline Group on Student Ratings of Instruction. IDEA Research Report #10. In *IDEA Center, Inc.* IDEA Center, Inc.

Mazurek, R., Arvinen-Barrow, M., Huddleston, W., & Reckelberg, R. (2021). Beyond traditional peer-to-peer teaching evaluation: Using pedagogical theory in conceptualizing a

collaborative teaching development program. *Journal of University Teaching & Learning Practice*, *18*(6), 100–116. https://doi-org.unh.idm.oclc.org/10.53761/1.18.6.08

McCullough, B. D., & Radson, D. (2011). Analysing Student Evaluations of Teaching: Comparing Means and Proportions. *Evaluation & Research in Education*, *24*(3), 183–202.

Murray, D., Boothby, C., Zhao, H., Minik, V., Bérubé, N., Larivière, V., & Sugimoto, C. R. (2020). Exploring the personal and professional factors associated with student evaluations of tenure-track faculty. *PLoS ONE*, *15*(6), 1–21. https://doi-org.unh.idm.oclc.org/10.1371/journal.pone.0233515

National Research Council (2003). *Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics*. Washington, DC: The National Academies Press. https://doi.org/10.17226/10024.

Peterson, D. A. M., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS ONE*, *14*(5), 1–10. https://doi-org.unh.idm.oclc.org/10.1371/journal.pone.0216241

Spooren, P., & Christiaens, W. (2017). I liked your course because I believe in (the power of) student evaluations of teaching (SET). Students' perceptions of a teaching evaluation process and their relationships with SET scores. *Studies in Educational Evaluation*, *54*, 43–49. https://doi-org.unh.idm.oclc.org/10.1016/j.stueduc.2016.12.003

Stroebe, W. (2016). Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations. *Perspectives on Psychological Science*, *11*(6), 800–816. https://doi-org.unh.idm.oclc.org/10.1177/1745691616650284

Uttl B, White CA, Morin A (2013) The Numbers Tell It All: Students Don't Like Numbers! PLoS ONE 8(12): e83443. doi:10.1371/journal.pone. 0083443

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22–42. https://doi-org.unh.idm.oclc.org/10.1016/j.stueduc.2016.08.007

Valencia, E. (2020). Acquiescence, instructor's gender bias and validity of student evaluation of teaching. *Assessment & Evaluation in Higher Education, 45(4), 483–495.* https://doi-org.unh.idm.oclc.org/10.1080/02602938.2019.1666085

Wieman, C. and Gilbert, S. (2017). The Teaching Practices Inventory: A New Tool for Characterizing College and University Teaching in Mathematics and Science**.** *CBE-Life Sciences Education, Vol 13(3), pp. 552–569 (2014). https://doi.org/10.1187/cbe.14-02-0023.*