

# The Validity and Utility of Student Evaluations

**Rex J. Rempel, MSW**

University of St. Thomas and Lake Washington Institute of Technology

*Correspondence concerning this article should be addressed to Rex Rempel, Lake Washington Institute of Technology, 11605 132<sup>nd</sup> Ave NE, Kirkland, WA 98034. (425) 739-8285. Email: ReRex.Rempel@lwtech.edu*

## Abstract

This paper explores the conundrum of student evaluations. At the end of each school term, non-tenured collegiate instructors across disciplines and institutional classifications worry that student evaluations may unfairly derail their careers. Despite the prevalence of published research and opinion pieces, the academy seems far from reaching a consensus on whether or how to use student feedback. This re-examination of claims and the available evidence sets out to ascertain whether student evaluations of teaching provide meaningful information about the quality of teacher performance. Empirical studies reveal problems inherent to professorial evaluation and methodological flaws in the use of these high stakes tools. Nevertheless, the author argues, student evaluations offer useful qualitative and quantitative information about the student experience and the use of such feedback is consistent with social work practice. The author concludes with specific recommendations for the ethical and effective use of student evaluations in higher education.

*Keywords:* student evaluation, teaching, assessment, bias, best practices

---

## Introduction

Institutions of higher learning have used Student Evaluations of Teaching (SET) for nearly a century in hopes of improving the quality of education (Stark & Freishtat, 2014). Colleges and universities of all types survey students about their instructors and courses near the close of school terms and use the results to inform decisions about faculty tenure, promotion, and reward (Johnson, Narayanan, & Sawaya, 2013). Given the place of student reports in such decisions, Nate Kornell is not alone in fearing that “My livelihood depends on what my students say about me in course evaluations” (2013, paragraph 1).

Heated debate about these evaluations continues despite their ubiquity. Some faculty members alter syllabi, assignments, and pedagogical modalities in response to students’ numerical ratings and narrative comments; others refuse to open them. The purpose of this paper

is to present guidance to instructors and institutions alike for the use of student evaluations and to answer basic questions about their accuracy and appropriateness. Do SET reflect the quality of teaching? Does the nature of social work education suggest an appropriate course of action? Can any individual or tool objectively measure faculty performance? Educators argue whether students should be considered customers—whose satisfaction weighs heavily—whether young adults have the information necessary to assess professional performance, and whether the use of SET is just. Many point to the ambiguity in both what student responses measure and in how colleges and universities weigh those results. “The standards for tenure and promotion—teaching, research, and service—are often not clearly delineated, broadly discussed, or systematically evaluated” (Gentry & Stokes, 2015, p. 4). Used in the determination of merit raises, teaching awards, faculty assistance needs, and which faculty to dismiss, “this rating [practice] is perhaps the university’s most important and widely used indicator of a faculty member’s instructional effectiveness” (Thyer, Myers, & Nugent, 2011, p. 275). Given the disagreements about the use of SET, let alone the uncertainty about best practices, stakeholders’ concerns deserve attention. Despite the volume of conversation about SET, the academic community has no consensus about their use. The University of Southern California, for example recently announced that they will continue collecting SET but will no longer use them in tenure and promotion decisions (Flaherty, 2018). Schools need clear guidance whether to abandon evaluations entirely, modify them, or continue using this singular quantitative assessment tool. The following synthesis of quantitative research into correlates of higher and lower student evaluations offers an evidence-based answer. Student evaluations do not accurately measure instructor performance but do convey feedback useful to both programs and instructors. Application of social work’s ethical principles affirms that seeking and using such feedback remains not only reasonable but a responsibility of educators, provided schools address existent systematic flaws.

## Literature Review

Instructors have much to say about student evaluations. Thousands have published discussions on the subject in peer-reviewed academic journals in the past decade. The Chronicle of Higher Education published an average of two online or print articles per week on the issue between 2014 and 2016. Though many of those pieces merely express educators’ feelings, some report new research which can identify the strengths and problems inherent in current practices.

### Strengths of Course Evaluations

Several researchers found correlations between students’ course grades and their instructor ratings. In reviewing 9,240 SET from psychology courses at a large public university, Blackhart, Peruche, Dewart, and Joiner (2006) found significant evidence of higher evaluations for instructors in courses with higher average student grades. In his review of 18,175 students’ grades and evaluations, Zabaleta (2007) found a modest correlation between students’ grades and the evaluations they submitted. In 2010, Carrell and West published unique research: a randomized, controlled study of students and faculty. The United States Air Force Academy randomly assigned 10,534 cadets to instructors for a required course, taught from the same syllabus. Carrell and West established that classes with stronger performances that term—greater success towards course objectives—rated their instructors higher. A few years later Miles and House (2015) ascertained that the correlation with SET extended beyond students’ actual grades to include their *expected* grades. Student evaluations have not demonstrated partiality between

adjunct and full-time faculty, nor between tenured and non-tenured faculty (Thyer et al., 2011; Cheng, 2013). These findings reveal that students recognize and reward classroom success as they perceive it.

Students use these survey tools to express their feelings and reactions to their experiences, particularly about the learning environment. In a laboratory test of reactions to brief videos, students graded a lecturer using eye contact and gestures, speaking fluidly, and moving in space higher than the same lecturer when looking down, reading notes haltingly from behind a podium, even when students learned the material equally well<sup>1</sup> (Kornell, 2013). A seven-semester review of 3,938 engineering courses and 549 instructors at another R1 institution found higher SET for smaller classes and elective courses (Johnson et al., 2013). Miles and House (2015) confirmed that students used SET to express preferences for smaller classes, elective courses, and upper level courses in their review of over 30,000 evaluations. In parallel with informal tools such as direct interaction with instructors, conversation with peers, and web-based forums such as *ratemyprofessor.com*, students continue using SET - the only routine, recognized means at their disposal - to inform their institutions about the type of education they prefer.

Finally, Student Evaluations of Teaching create both a qualitative and quantitative measure of instructor assessment. In many cases, they supply the *only* barometer for evaluation of instructor effectiveness. As Miles and House (2015) put it, “other qualitative metrics such as peer evaluation and educational development activities are included in the [Tenure and Promotion] package, but the [student evaluation] score is the only standard quantitative metric available for measurement” (p. 299), though many question the validity and reliability of student evaluations.

### **Methodological Problems with SET**

Multiple methodological flaws reduce the validity of SET as measures of teacher effectiveness. First, students evaluate partial courses; schools typically solicit student evaluations before completion of a course. Second, sample sizes have shrunk. Schools’ replacement of in-class pencil-and-paper exercises with cheaper outside-class online response systems has been accompanied by a drop of thirty percent or more in response rates (Jan, 2010).

The Director of the University of California Berkeley Center for Teaching and Learning, Richard Freishtat, feared that educators misread SET scores. He and Cal-Berkeley Dean of Statistics Phillip Stark considered the comparison of averages of SET scores foolish.

Personnel reviews routinely compare instructors’ average scores to departmental averages. Such comparisons make no sense, as a matter of Statistics. They presume that the difference between 3 and 4 means the same thing as the difference between 6 and 7. They presume that the difference between 3 and 4 means the same thing to different students. They presume that 5 means the same thing to different students and to students in different courses. They presume that a 3 “balances” a 7 to make two 5s. For teaching evaluations, there’s no reason any of those things should be true. (2014, pp. 5-6)

---

<sup>1</sup> Kornell misrepresents the observable behaviors to which students reacted by characterizing them as *fluent* and *disfluent*, inducing consideration of ethnic bias without language- or ethnicity-based justification from his findings.

Instead, they recommended that users consider the distribution of scores, the number of respondents, and the response rate; the range and scatter of responses tell much more than an average, including the variability of student perceptions and instructors' consistency across time and courses.

Several environmental factors beyond the instructors' control affect SET scores and thus such instructors' professional opportunities. As noted above, students gave higher ratings for smaller classes and elective courses (Johnson et al., 2013; Miles & House, 2015; Voeten & Martin, 2013). In an experimental study of 325 students, female students rated teachers significantly lower on most elements than did their male counterparts (Basow, Codos, & Martin, 2013). Consequently, instructors assigned to larger, female-populated, and required courses fare worse in evaluation, presumably without cause. This may be significant in disciplines such as social work, nursing, and education with majority female student bodies.

Multiple studies demonstrate that SET reflect student grades (Blackhart et al., 2006; Langbein, 2008; Zabaleta, 2007). Instructors who deliver high grades will generally get better evaluations than those who confer a range of grades. When professors and students grade one another and thereby impact each other's career opportunities, instructors benefit from grade inflation rather than honest assessments of student performance (Langbein, 2008). In such conditions, SET can reflect an instructor's willingness to give an A as much as their skill.

Finally, some students make false reports. Up to thirty percent of students included information that they knew to be inaccurate, such as their accounts of the timeliness of grading or the quality of presentations. Students reported intentionally inflating or deflating their evaluations to reward or punish an instructor (Clayson & Haley, 2011). Surveying students does not necessarily mean measuring teacher effectiveness. As Stark and Freishtat (2014) state, however effectively students can assess certain pedagogical practices, SET are only tenuously tied to teaching effectiveness.

### **Bias in Student Evaluations**

Student evaluations seemed to be products of cultural biases as much as objective measures of performance. Weinberg, Fleisher, and Hashimoto (2007) found inconsistent and statistically insignificant evidence of students grading foreign-born instructors lower than those born in the United States. Earlier research asserted but did not prove that teachers' appearances had a marginal impact on their ratings (Ambady & Rosenthal, 1993).

The evidence of sexism and racism is stronger. In reviewing nearly fifty thousand evaluations of a university's economics courses, Weinberg et al. (2007) found that male instructors scored higher than female instructors and saw a "substantial (but statistically insignificant) foreign-domestic gap" with regard to instructors' place of birth (p. 11). Basow et al. (2013) reported surprising results from their experimental study using computer-animated instructors who variously appeared as African-American, White, female, and male.

Contrary to predictions, African American professors were rated higher than White professors on their hypothetical interactions with students. Quiz results, however, supported predictions: higher scores were obtained by students who had a White professor compared to those who had an African American professor, and by students who had a male professor compared to those who had a female professor. These results

may be due to students paying more attention to the more normative professor. Thus, performance measures may be a more sensitive indication of race and gender biases than student ratings. The limited relationship between student ratings and student learning suggests caution in using the former to assess the latter. (p. 352)

Johnson et al.'s (2013) review of engineering courses not only found higher aggregate SET for smaller classes and elective courses but also for male instructors. Studies differed on the size of the effect, but female instructors did tend to receive lower SETs, especially in large classes, thereby introducing a gender penalty (Miles & House, 2015; Voeten & Martin, 2013; Wagner, Rieger, & Voorvelt, 2016). The effect need not be the product of genuine gender-based differences in instructor behavior; student evaluations reward *perceived* maleness. When unseen online instructors variously assumed male and female identities, students rated the male-identified instructors higher regardless of the instructor's actual gender (MacNeill, Driscoll, & Hunt, 2015). Perception mattered, not differences in teaching. Together these findings implicate bias as impacting student behavior and teacher evaluations, revealing as much about the evaluators as they do about instructors.

### **SET within Social Work Education**

Research on students' evaluations of social work instructors raised similar concerns about gender-based differences and the difficulty in making meaningful distinctions between instructors based on student evaluations (Wolfer & Johnson, 2003). Findings point to biases within social work students similar to their peers in other disciplines, even at the graduate level (Perry, Wallace, Moore, & Perry-Burney, 2014; Wolfer & Johnson, 2003). Social work students, however, do consistently rate their instructors near the top of evaluation scales (Steiner, Holley, Gerdes, & Campbell, 2006; Wolfer & Johnson, 2003).

Studies of SET within social work education also legitimized student assessments in ways mirroring other research. Steiner et al. (2006) found that teachers' scores tracked with students' perceptions of their learning. Jirovec, Ramanathan, and Rosegrant-Alvarez (1998) found evidence associating higher SET with perceived fairness, rapport, and perceptions about the organization of the course.

Social work educators may be better equipped than their colleagues in other departments to appreciate and utilize SET. "Social work norms imply the necessity of seeking input from students" (Steiner et al., 2006, p. 355). Social work practice and SET alike proceed from an appreciation for a diversity of voices, evaluation of practice, open communication, professional development, and attention to client or student outcomes (Miller & Wilson, 1977). For these reasons, Miller and Wilson argue that SET can be recognized in particular within social work education as productive and non-threatening. Nevertheless, they note, the complexity of teaching escapes measurement by simple tools; instructor evaluation requires the use of multiple impartial measures.

## **Discussion**

Complaints that student evaluations cannot assess faculty members' true worth are valid. As Boyer (1990) reminded us, the professoriate is a multi-dimensional role. Collegiate instructors are also scholars responsible for the discovery of new knowledge, interdisciplinary

collaboration, and the application of knowledge to significant problems. A fair evaluation requires attention to each of these varied forms of scholarship. Student evaluations cannot assess the quality of an individual's research, contributions to the college community, or engagement with the discipline.

Despite these flaws, student evaluations of courses and instructors are neither meaningless nor inconsequential. The millions completed each year inform innumerable decisions made by fellow students, faculty members, and administrators. They impact enrollment, reputations, syllabi, work opportunities, pedagogy, and professional advancement, and not without reason.

These end-of-term surveys give voice to student views. Instructors and institutions vary significantly in their use of this data—such as the significance of student evaluations for tenure and promotion decisions (Voeten & Martin, 2013)—but the information appears to reflect what postsecondary students want from their education as well as their reactions to faculty behavior. This communication tool provides students opportunities to explicate their experience of higher education, an endeavor expensive in time, money, and energy, one that will significantly impact their futures. Publicly accessible evaluations facilitate informed student choice of instructors. Students also benefit in so far as evaluations lead to practice changes by instructors, such as increased efforts at student engagement or the use of active pedagogical methods. Finally, the promotion and retention of faculty members with higher evaluation scores increases the portion of instructors on faculty who deliver closer contact, dynamic presentation, focused coursework, choices within coursework, and who communicate fluently.

Teachers benefit, as well, whenever institutions attempt to make merit-based decisions if sound methodologies are used. Student evaluations can identify strengths in instructors' performance and areas for further attention. The repeated use of the same tool creates a measure of change over time, a means of gauging improvement through specific practice changes.

These qualitative and quantitative feedback mechanisms supply programs and schools with economical, accessible, easily digested data about the relationships between those seeking higher education and those furnishing it. In particular SET can benefit institutions through better understandings of the strengths or weaknesses of their curriculum, as reflected in student scores. These tools might more appropriately be named Student Evaluations of an Educational Experience, due to the significance of matters outside instructor control such as course size and degree requirements. Consistent feedback from students of multiple instructors and courses can highlight whether curricular changes are warranted. Finally, schools win whenever the use of student evaluations encourages instructors to engage students as partners in education. Satisfied students may be more likely to remain enrolled, graduate, and encourage others' support of the school (Pascarella, Salisbury, & Blaich, 2011).

Benefits extend beyond campuses. This process of evaluation girds the unwritten compact between educators and the public. The academic profession's social contract allows for (and requires) academic freedom in correspondence with peer review, shared governance, and a focus on public service (Hamilton & Gaff, 2009). Professional rights sit upon responsibilities such as review and evaluation. The last half-century's views of postsecondary students as vested stakeholders, partners in education, or customers made those students a new type of peer and thus reviewers.

Social work's values and client-centered traditions imply that its educators should be especially open to student feedback. Social workers are trained to listen carefully to marginalized voices, to focus explicitly on clients' goals, to regard clients as experts on their own lives. Social workers are well prepared to openly discuss bias, differences in power, and structural barriers. Seeking and receiving student feedback should be second nature to experienced social workers.

Student evaluations deliver a basic indication of achievement of course objectives. They recognize student engagement (or disengagement) in response to instructors' behavior. Student evaluations highlight the types of courses and communication which energize students. Comments and scores raise awareness of the impacts—positive and negative—of course schedules, modalities, *and* faculty behaviors.

It is fair and appropriate for educators to question the validity and reliability of student responses; after all, students lack key information and are hardly impartial respondents. Nevertheless, higher education cannot dismiss student evaluation altogether by claiming that immature young adults lack the necessary skills. Liberal education is explicitly designed to develop critical thinking skills; students' inability to assess methods or outcomes is itself an indictment of their education. Instructors have responsibilities to help their students situate and contextualize the content learned, including its significance and its relationships to broader themes.

Social work, of all academic disciplines, has reason to validate student input. National and international social work education standards call for both self-evaluation and external evaluation of programs (CSWE, 2015; Sewpaul & Jones, 2007). The Council on Social Work Education also demands that schools of social work develop students' ability to evaluate and assess practice, outcomes, and effectiveness. Social work educators and their colleagues alike are expected to produce critical thinkers and facilitate the exercise of assessment skills.

It is similarly incongruous to dismiss student evaluations due to bias. No individual or tool can objectively measure faculty performance. Evaluation is inherently flawed and prejudiced, whether it be peer review or student evaluation (Greenwald & Krieger, 2006). Faculty are no better than students in this regard. Many researchers have concluded that faculty biases related to race, gender, and socioeconomic status influence the grades they award students (Guskin, Peng, & Simon, 1992; King, 1998; Page & Rosenthal, 1990; Tourmaki, 2003). Teachers continue to grade student work, nevertheless.

## **Recommendations**

Student evaluations can help improve the quality of education if schools clarify what is being measured and address factors known to skew assessments. SET are not evaluations of teaching per se. Student evaluations are likely less valid measures of teaching skill than genuine peer review but they do offer valuable feedback and assessment of students' learning and educational experience, something which faculty and schools should consider carefully. Student Evaluations of an Educational Experience may be as or more useful as institutional quality assurance tools than SET are as measures of individual faculty teaching.

### **Recommendations for Collegiate Faculty**

All faculty, regardless of discipline, rank, or educational system, can benefit from the use of student evaluations. Kornell (2013) and other college instructors have expressed fear that student reactions will cause them irreparable harm. College professors are not passive victims of student evaluations, however. They can and do influence the outcomes, as has been demonstrated by empirical results. It may be advisable for faculty members to invite students early on to share their goals for a course, to articulate their measures of success. Doing so can inform the instructor of their students' expectations, allowing for clarification of what is and is not feasible. It also offers opportunities to shape the direction of the course in satisfying ways. Instructors can encourage student use of evidence in assessment by discussing the group's progress towards students' measures of success before their use of school-mandated evaluation surveys. Students might be more inclined to make reasonable assessments when they are encouraged to present a basis for their assessment, particularly one that they identified. Beyond preparing for a final, review sessions that reference the group's goals can help students appreciate their progress.

Instructors can also help students understand the nature and use of the tool. They can discuss the application, methodology, benefits, and weaknesses of student evaluations. College students can understand the limits on instructors in larger classes and survey courses, for example, but may need assistance in seeing how such factors influence their evaluations. Higher education is predicated on the belief that adult learners can be lead to growth in their knowledge, skills, and values. Faculty from any discipline can teach evidence-based evaluation skills to students, as part of their education in critical thinking.

Those committed to the quality of their craft would be well served by considering their student evaluations as one of many data sources. Synthesis of self-evaluation, peer observations, student evaluation, and other sources of information should produce a rounded understanding; trends across informants deserve particular attention. Instructors and students likely benefit when instructors seek student feedback, even informal feedback, during the term while changes remain possible. As with grading student papers, timely and specific feedback leads to more significant outcomes. Faculty who set goals to improve their performance in particular domains can use SET as a tool to note change over time, and thus the efficacy of their improvement efforts.

Finally, instructors and their unions may want to advocate for appropriate changes in how their institutions use student evaluations, to ensure reasonable evaluation of the quality of their work.

### **Recommendations for Administrators in Higher Education**

SET's methodological problems can be addressed. Colleges should allow students to complete a course, see its outcomes, and move beyond the pressure of final exams—and an instructor's influence—before evaluating that course. Ask students about specific faculty behaviors. Eliminate global assessments of teaching skills and other queries for which students lack adequate information. Institutions should avoid the use of averages to compare instructors. Consider the response rate and the distribution of responses, instead, using only those scores generated by a sample with enough participants to be meaningful.

Administrators may wish to adjust scores for known biases, including gender and class type, adjusting scores slightly upward for larger, mandated, female-populated, or female-led



classes.<sup>2</sup> Such adjustments may prove important in assessing the faculty of introductory survey courses, for example. Any modifications, however, should be made with attention to school norms, given the real possibility of setting- or population-specific differences. Whether such corrections would be acceptable to faculty senates, unions, or the legal system is essential to implementation but beyond the scope of this paper.

All involved may need to minimize the use of SET for intra-faculty comparisons. Deans can consider whether an instructor awards a range of grades. Meanwhile, educators must use other appropriate sources of information, including self-assessment and peer observation, to determine for themselves whether an individual's work represents quality scholarship (Miller & Wilson, 1977; Wolfer & Johnson, 2003). As the Russian proverb says “Доверяй, но проверяй;” trust, but verify.

Finally, those responsible for tenure and promotion decisions must evaluate candidates justly. The quality of a candidate's research, publications, program administration, contributions to the community, and performance as a member of the college community must be assessed alongside their teaching, through alternate means. Reserve SET for consideration of teaching efficacy and adjustments to the implicit or explicit curriculum.

## References

- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*(3), 431-441. doi: 10.1037/0022-3514.64.3.431
- Basow, S. A., Codos, S., & Martin, J. L. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal*, *47*(2), 352-363.
- Blackhart, G. C., Peruche, B. M., Dewall, C. N., & Joiner Jr, T. E. (2006). Factors influencing teaching evaluations in higher education. *Teaching of Psychology*, *33*(1), 37-39.
- Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching.
- Carrell, S. E. & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, *118*(3), 409-432. doi: 10.1086/653808
- Cheng, D. A. (2013). Effects of professorial tenure on undergraduate ratings of teaching performance. *Education Economics*, *23*(3), 1-20. doi: 10.1080/09645292.2013.826632
- Clayson, D. E., & Haley, D. A. (2011). Are students telling us the truth? A critical look at the student evaluation of teaching. *Marketing Education Review*, *21*(2), 101-112. doi: 10.2753/mer1052-8008210201

---

<sup>2</sup> Female or female-perceived instructors received SET from 1.2% to 7% lower, depending on the study, with an average effect of 4.5%. The two studies noted in this article which addressed class size saw an average difference of 1.6% per ten students. Two two studies noted here which measured SET for elective and required courses showed an average of 2.7% higher scores for elective courses. Any one course may have a combination of positive and negative bias effects.

- Council on Social Work Education. (2015). Educational policy and accreditation standards for baccalaureate and master's social work programs. Retrieved from <http://www.cswe.org/File.aspx?id=81660>
- Flaherty, C. (2018, May 22). Teaching eval shake-up. *Inside Higher Ed*. Retrieved from <https://www.insidehighered.com>
- Gentry, R., & Stokes, D. (2015). Strategies for professors who service the university to earn tenure and promotion. *Research in Higher Education Journal*, 29, 1-15.
- Greenwald, A., & Krieger, L. (2006). Implicit bias: Scientific foundations. *California Law Review*, 94(4), 945-967. doi:10.2307/20439056
- Guskin, S. L., Peng, C. J., & Simon, M. M. (1992). Do educators react to "multiple intelligences"? Effects of educators' stereotypes on judgments and expectancies for students with diverse patterns of giftedness/talent. *Gifted Child Quarterly*, 36(1), 32-37. doi: 10.1177/001698629203600108
- Hamilton, N. W., & Gaff, J. G. (2009). *The future of the professoriate: Academic freedom, peer review, and shared governance*. Washington, D.C: Association of American Colleges and Universities.
- Jan, T. (2010). Colleges see decline in evaluations after going online. *The Boston Globe*. Retrieved from [http://archive.boston.com/news/local/massachusetts/articles/2010/04/07/colleges\\_see\\_decline\\_in\\_evaluations\\_after\\_going\\_online/](http://archive.boston.com/news/local/massachusetts/articles/2010/04/07/colleges_see_decline_in_evaluations_after_going_online/)
- Jirovec, R. L., Ramanathan, C. S., & Rosegrant-Alvarez, A. (1998). Course evaluations: What are social work students telling us about teaching effectiveness? *Journal of Social Work Education*, 34(2), 229-236. doi: 10.1080/10437797.1998.10778919
- Johnson, M. D., Narayanan, A., & Sawaya, W. J. (2013). Effects of course and instructor characteristics on student evaluation of teaching across a college of engineering. *Journal of Engineering Education*, 102(2), 289-318. doi: 10.1002/jee.20013
- King, J. L. (1998). The effects of gender bias and errors in essay grading. *Educational Research Quarterly*, 22(1), 13-25.
- Kornell, N. (2013). Do the best professors get the worst ratings? *Psychology Today*. Retrieved from <https://www.psychologytoday.com/blog/everybody-is-stupid-except-you/201305/do-the-best-professors-get-the-worst-ratings>
- Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review* 27, 417-428. doi: 10.1016/j.econedurev.2006.12.003
- MacNell, L., Driscoll, A., & Hunt, A. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291-303. doi: 10.1007/s10755-014-9313-4
- Miles, P., & House, D. (2015). The tail wagging the dog: An overdue examination of student teaching evaluations. *International Journal of Higher Education*, 4(2), 116-126. doi: 10.5430/ijhe.v4n2p116
- Miller, D. A., & Wilson, C. P. (1977). Student evaluation of faculty: The other side--corruption of faculty and social work responsibility. *Journal of Social Welfare*, 4(2 & 3), 99-107.
- Page, S., & Rosenthal, R. (1990). Sex and expectations of educators and sex and race of students as determinants of teaching behavior and student performance. *Journal of School Psychology*, 28(23), 119-131. doi: 10.1016/0022-4405(90)90003-p
- Pascarella, E. T., Salisbury, M. H., & Blaich, C. (2011). Exposure to effective instruction and college student persistence: A multi-institutional replication and extension. *Journal of College Student Development*, 52(1), 4-19. doi: 10.1353/csd.2011.0005
- Perry, A. R., Wallace, S. L., Moore, S. E., & Perry-Burney, G. D. (2014). Understanding student evaluations: A black faculty perspective. *Reflections: Narratives of Professional Helping*, 20(1), 29-35.
- Sewpaul, V., & Jones, D. (2007). Global standards for social work education and training. *Social Work Education*, 23(5), 493-513. doi: 10.1080/0261547042000252244
- Stark, P. B. & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen*. Retrieved from <http://www.stat.berkeley.edu/~stark/Preprints/evaluations14.pdf>. doi: 10.14293/S2199-1006.1.sor-edu.aofrq.v1

- Steiner, S., Holley, L. C., Gerdes, K., & Campbell, H. E. (2006). Evaluating teaching: Listening to students while acknowledging bias. *Journal of Social Work Education, 42*(2), 355–376. doi: 10.5175/jswe.2006.200404113
- Thyer, B. A., Myers, L. L., & Nugent, W. R. (2011). Do regular social work faculty earn better student course evaluations than do adjunct faculty or doctoral students? *Journal of Teaching in Social Work, 31*(4), 365–377. doi: 10.1080/08841233.2011.597674
- Tourmaki, N. (2003). Effect of student characteristics on educators' predictions of student success. *Journal of Educational Research, 96*(5), 310–320. doi: 10.1080/00220670309597643
- Voeten, E., & Martin, L. (2013). Student evaluations of teaching are probably biased. Does it matter? *The Washington Post*. Retrieved from <https://www.washingtonpost.com/news/monkey-cage/wp/2013/10/02/student-evaluations-of-teaching-are-probably-biased-does-it-matter/>
- Wagner, N., Rieger, M., & Voorvelt, K. (2016). Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams. *Economics of Education Review, 54*, 79–94. doi: 10.1016/j.econedurev.2016.06.004
- Weinberg, B.A., Fleisher, B.M., & Hashimoto, M. (2007). Evaluating methods for evaluating instruction: The case of higher education (NBER Working Paper No. 12844). Retrieved from <http://www.nber.org/papers/w12844>. doi: 10.3386/w12844
- Wolfer, T. A., & Johnson, M. M. (2003). Re-evaluating student evaluation of teaching. *Journal of Social Work Education, 39*(1), 111–121. doi: 10.1080/10437797.2003.10779122
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education, 12*(1), 55–76. doi: 10.1080/13562510601102131

**Rex Rempel, LICSW**, is an Associate Professor and Department Director for Behavioral Health and Social Services at Lake Washington Institute of Technology, a community college near Seattle. He previously served as a field instructor for the University of Washington School of Social Work and an adjunct instructor of Social and Behavioral Science at Northwest University. Professor Rempel earned a baccalaureate degree in philosophy degree from the University of Notre Dame and a Masters of Social Work from the University of Washington. Mr. Rempel has worked as a shelter counselor, mental health case manager, gerontological social worker, mental health program supervisor, and crisis diversion facility program manager. He is completing a Doctorate of Social Work at the University of St. Thomas. His dissertation on social work education at community colleges in the United States will be published in the coming year.