

# Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related



Bob Uttl<sup>\*</sup>, Carmela A. White<sup>1</sup>, Daniela Wong Gonzalez<sup>2</sup>

Department of Psychology, Mount Royal University, Canada

## ARTICLE INFO

### Article history:

Received 23 February 2016

Received in revised form 10 August 2016

Accepted 17 August 2016

Available online 19 September 2016

### Keywords:

Meta-analysis of student evaluation of teaching

Multisection studies

Validity

Teaching effectiveness

Evaluation of faculty

SET and learning correlations

## ABSTRACT

Student evaluation of teaching (SET) ratings are used to evaluate faculty's teaching effectiveness based on a widespread belief that students learn more from highly rated professors. The key evidence cited in support of this belief are meta-analyses of multisection studies showing small-to-moderate correlations between SET ratings and student achievement (e.g., Cohen, 1980, 1981; Feldman, 1989). We re-analyzed previously published meta-analyses of the multisection studies and found that their findings were an artifact of small sample sized studies and publication bias. Whereas the small sample sized studies showed large and moderate correlation, the large sample sized studies showed no or only minimal correlation between SET ratings and learning. Our up-to-date meta-analysis of all multisection studies revealed no significant correlations between the SET ratings and learning. These findings suggest that institutions focused on student learning and career success may want to abandon SET ratings as a measure of faculty's teaching effectiveness.

© 2016 Elsevier Ltd. All rights reserved.

*"For every complex problem there is an answer that is clear, simple, and wrong."* H. L. Mencken

Student Evaluation of Teaching (SET) ratings are used to evaluate faculty's teaching effectiveness based on an assumption that students learn more from highly rated professors. Although SET were used as early as 1920's, their use expanded across the USA in the late 1960's and early 1970's (Murray, 2005; Wachtel, 1998). Today, nearly all colleges and universities in North America use SET to evaluate their faculty's teaching effectiveness (Murray, 2005; Wachtel, 1998). Typically, SET are conducted within the last few weeks of courses, before the final grades are assigned. Students are presented with rating forms that ask them to rate their perceptions of instructors and courses, often on a 5-point Likert scale ranging from Strongly Disagree to Strongly Agree. The rating forms may ask students to provide overall ratings of instructor and/or course and they may also ask students to rate numerous specific characteristics of teachers (e.g., knowledge, clarity of explanation, organization, enthusiasm, friendliness, fairness, availability, approachability, use of humor, contribution to students' learning)

and courses (e.g., organization, difficulty) (Feldman, 1989; Spooen, Brockx, & Mortelmans, 2013). The ratings for each course/class are summarized, typically by calculating mean ratings across all responding students for each rated item and across all rated items, and these mean class SET ratings are then used to evaluate professors' teaching effectiveness by comparing them, for example, to department or university average ratings. Although use of SET as a feedback for professors' own use is not controversial, the use of SET as a measure of professors' teaching effectiveness for making high stakes administrative decisions about instructors' hiring, firing, merit pay, and promotions is highly controversial (e.g., Emery, Kramer, & Tian, 2003; Spooen, Brockx, & Mortelmans, 2013; Stark & Freishtat, 2014; Wachtel, 1998).

Proponents of SET as a measure of instructor teaching effectiveness have put forward a number of reasons for their use: (1) SET are cheap and convenient means to evaluate faculty's teaching, (2) SET are very useful to demonstrate administrators' concerns with public accountability and public relations, (3) SET allow students to have say in evaluation of faculty's teaching, and (4) students are uniquely positioned to evaluate their experiences and perceptions of instructors as they are teaching classes (Murray, 2005; Wachtel, 1998). The last reason on this list is the SET proponents' main rationale for why SET ought to measure instructor's teaching effectiveness. The SET proponents assume that students observe instructors' behavior, assess how much they learned from the instructor, rate the instructor according to how

<sup>\*</sup> Corresponding author at: Department of Psychology, Mount Royal University, 4825 Mount Royal University Gate, Calgary, AB, T3E 6K6, Canada.

E-mail address: [buttl@mtroyal.ca](mailto:buttl@mtroyal.ca) (B. Uttl).

<sup>1</sup> Department of Psychology, University of British Columbia, Okanagan, Canada.

<sup>2</sup> Department of Psychology, University of Windsor, Canada.

much the instructor's contributed to their learning, and thus, high correlation between SET and measures of learning should follow. In contrast, the opponents of SET as measure of teaching effectiveness argue that SET are primarily measures of student satisfaction, that is, "a happy or pleased feeling because of something that you did or something that happened to you" ([www.merriam-webster.com](http://www.merriam-webster.com)). Clearly, whether a student is overwhelmed by "happy or pleased feeling" at the end of the course is likely to depend on many factors that have nothing to do with instructor's teaching effectiveness, for example, whether or not a student was getting grades that he or she thought she deserved to be getting throughout a course, whether or not a course was forced on a student by being required, whether or not a student was reported by an instructor for cheating or plagiarism, whether or not a student found instructor's accent or looks pleasant, etc. The opponents of SET as measures of teaching effectiveness argue that SET have no or only limited validity as a measure of instructor teaching effectiveness because both SET and measures of learning are influenced by teaching effectiveness irrelevant factors (TEIFs) such as academic discipline/field of study, student interest, student motivation, instructor sex, instructor accent, class level, class size, class meeting time, etc. (Franklin & Theall, 1995; Hoyt & Lee, 2003; Spooren, Brockx, & Mortelmans, 2013; Uttl et al., 2012; Wachtel, 1998). Although thousands of studies have examined validity of SET, including influence of various TEIFs on SETs, the gulf between the proponents and opponents of SET is as wide as ever.

However, the well established findings in cognitive psychology and intelligence literature suggest that any substantive correlations between SET and learning are likely to be a fluke or an artifact rather than due to students' ability to accurately assess instructor's teaching effectiveness. First, how well students do on measures of learning is dependent to large degree on students' intelligence or ability to learn, prior relevant knowledge, and motivation to learn. Second, students' ability to judge how much they learned is also dependent on their intelligence or ability. One of the well-established findings in cognitive psychology is so called Dunning-Kruger effect (Kruger & Dunning, 1999) showing that unskilled persons assess their ability to be much higher than it really is and that highly skilled persons underestimate their ability and assume that tasks they found easy were also easy for others. In one set of studies, Dunning and Kruger examined Cornell University students' self-assessment of logical reasoning skills, grammatical skills, and humor. When the students were showed their scores and asked to estimate their own rank in the class, the competent students estimated their rank accurately whereas the incompetent students overestimated their rank to such a degree that they

believed their work deserved B or better grade even though their work received Ds or Fs.

The key evidence cited in support of the belief that SET measure instructor's teaching effectiveness are multisection studies showing correlations between SETs and student achievements, the correlations that have been acknowledged and accepted as true by both proponents and opponents of SETs. Fig. 1 describes the logic of multisection studies. An ideal multisection study design includes the following features: a course has many equivalent sections following the same outline and having the same assessments, students are randomly assigned to sections, each section is taught by a different instructor, all instructors are evaluated using SETs at the same time and before a final exam, and student learning is assessed using the same final exam. If students learn more from more highly rated professors, sections' average SET ratings and sections' average final exam scores should be positively correlated. However, random assignment of students to sections is rarely possible. Accordingly, some multisection studies control for prior learning/ability by, for example, regressing individual students' achievement scores on measures of students' prior learning/ability and using residual gains in achievement, averaged across all students within sections, as measures of achievement/learning. In general, researchers have agreed that multisection study designs are the best for determining the relationship between SET ratings and student learning facilitated by professors. For example, Abrami, d'Appollonia, and Cohen (1990) summarize this view as follows: "The multisection validation design is the strongest design for assessing the degree to which student ratings predict teacher-produced student learning" (p. 230).

More than three decades ago, Cohen (1981) conducted the first meta-analysis of multisection studies and reported that SET ratings correlate with student learning with  $r = .43$ , a small-to-moderate correlation. Cohen wrote: "The results of the meta-analysis provide strong support for the validity of student ratings as a measure of teaching effectiveness" (p. 281) and "... we can safely say that student ratings of instruction are a valid index of instructional effectiveness. Students do a pretty good job of distinguishing among teachers on the basis of how much they have learned" (p. 305). Since that time, Cohen's meta-analysis has been frequently cited hundreds of times in support of using SETs to evaluate faculty's teaching effectiveness (see Table 1) and the view that multisection studies have demonstrated validity of SET ratings as a measure of teaching effectiveness – that students learn more from highly rated professors – has been accepted as the established fact in various research summaries and widely disseminated to faculty members, administrators, and general public. Even self-help books designed to improve teaching of beginning faculty members inform them that the research has established that SET ratings measure teaching effectiveness. A few quotes from these reviews and materials will suffice:

"We (d'Apollonia & Abrami, 1977) reviewed the research evidence from multisection validity studies and found that ratings explain instructor impacts on student learning to a moderate extent (corrected  $r = .47$ )" (Abrami & d'Apollonia, 1999, p. 519)

"Overall, multisection validity studies have shown substantial correlation with student achievement as measured by examination performance. (Abrami, d'Apollonia, and Cohen, 1990; Abrami and d'Apollonia, 1997)" (Ory & Ryan, 2001, p. 43)

"... meta-analyses of multisection validity research have supported the validity of SETs by demonstrating that the sections that evaluate the teaching as most effective are also the sections that perform best on standardized final examinations (Cohen, 1981, 1987; Feldman, 1989). This research demonstrates that teachers who receive better SETs are also the

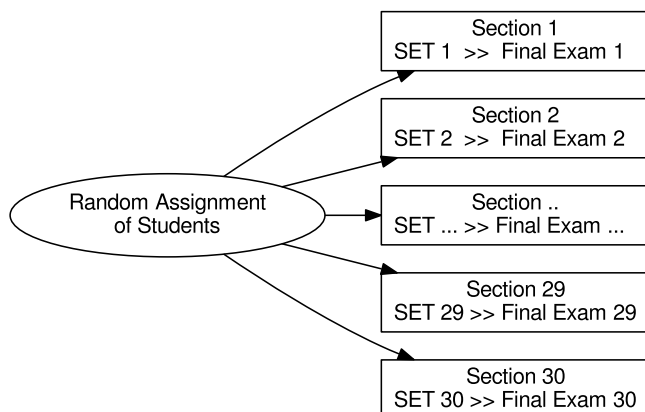


Fig. 1. Multisection SET validity study design.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات