

Student Ratings of Teaching Effectiveness: Student Engagement and Course Characteristics

Tanya Beran
Claudio Violato
University of Calgary

ABSTRACT

Characteristics of university courses and student engagement were examined in relation to student ratings of instruction. The Universal Student Ratings of Instruction instrument was administered to students at the end of every course at a major Canadian university over a three-year period. Using a two-step analytic procedure, a latent variable path model was created. The model showed a moderate fit to the data (Comparative Fit Index = .88), converged in 10 iterations, with a standardized residual mean error of .03, $\chi^2(149) = 1988.59$, $p < .05$. The model indicated that course characteristics such as status and description are not directly related to student ratings. Rather, they are mediated by student engagement, which is measured by student attendance and expected grade. It was concluded that, although the model is statistically adequate, many other factors determine how students rate their instructors.

RÉSUMÉ

Cette recherche évalue les caractéristiques des cours universitaires ainsi que la participation des étudiants en rapport avec l'évaluation estudiantine de l'enseignement. Pendant trois ans, dans une grande université du Canada, on a administré un outil d'évaluation appelé

« Universal Student Ratings of Instruction » [évaluation estudiantine globale de l'enseignement] auprès des étudiants, à la fin de chaque cours. On créa un modèle d'analyse à variables latentes en utilisant une méthode d'analyse à deux étapes. Le modèle s'accordait passablement aux données (Indice d'ajustement comparatif = .88), a convergé dans dix itérations avec une déviation standard à résidu normalisé de .03, $\chi^2(149) = 1988.59$, $p < .05$. Le modèle démontre que les caractéristiques des cours tels que le statut et la description ne sont pas directement liés aux évaluations des professeurs. Elles sont plutôt influencées par la participation des étudiants, et mesurées par l'assiduité des étudiants ainsi que par les notes anticipées. Bien que le modèle soit adéquat statistiquement, les résultats suggèrent qu'il existe plusieurs autres éléments influençant l'évaluation des professeurs par les étudiants.

INTRODUCTION

Since student evaluations of teaching effectiveness can have a significant impact on instructors' careers (Sprinkle, 2008), considerable research has been conducted to ensure that these ratings are valid. Greenwald (1997) concluded that these studies "give a clear impression that major questions of the 1970s about ratings validity were effectively answered and largely put to rest by subsequent research" (p. 1184). Nevertheless, researchers suggest that various characteristics about the courses and the students themselves may, in part, contribute to these ratings. Course and student characteristics have typically been studied for their impact on student ratings of instruction, while studies examining the relative importance of several of these characteristics simultaneously are rare. Accordingly, in the present study, we set out to determine the extent to which characteristics about the course and student engagement in the course influence how students rate their instructors.

Validity

Validity refers to the extent to which a measure accurately quantifies the construct being measured (Messick, 1989). In the context of teaching evaluation, when student ratings reflect the process of instruction (i.e., what teachers do when they teach) and the impact of teachers on the desired products of instruction (i.e., student learning), the ratings are said to be valid (Abrami, d'Apollonia, & Cohen, 1990). Over the past several decades, a considerable body of research, commentary, and criticism has been focused on issues related to the validity of student ratings (Abrami, 2001; Ory & Ryan, 2001). Having explored historical trends on the validity of student ratings, Greenwald (1997) reported that over a 25-year period beginning in the 1970s, more publications favoured evidence for, rather than against, validity. These conclusions about the adequate validity of student ratings have been supported by other reviewers (Cohen, 1981; McKeachie, 1979; Murray, 1984).

Although teaching effectiveness is one factor that determines student ratings, additional factors outside of teaching, such as the characteristics of the courses being evaluated and the engagement of students conducting the evaluations, may influence how students rate their instructors.

Course Characteristics

University and college courses vary according to several characteristics. Courses can be described by their type and length. Types of courses typically provided in higher education include lectures, labs, practica, tutorials, and distance learning. Little research has examined differences in student ratings according to these types of courses, but one study found that labs received higher ratings than lectures or tutorials (Beran & Violato, 2005). It is possible that hands-on application of theory and research leads to greater satisfaction in learning, and, hence, higher student ratings for instruction. In terms of course length, little research has examined whether longer courses receive higher ratings than shorter courses. In a preliminary study by Beran and Violato (2005), course duration was related to student ratings, but the relationship was small (effect size—Cohen's d —of .15). Given the greater opportunity for contact and learning from an instructor in a full-year rather than a half-year course, students may gain more knowledge and skills, which, in turn, may yield higher student ratings.

Courses also vary according to their status, that is, whether they are required, are within the student's program, and have a heavy workload. Arreola (1995) and McKeachie (1979) each examined the role of course requirement and found that student ratings were lower for required courses, compared to elective courses. Similarly, students may give higher ratings to courses outside their area of study than to courses within their program if those courses are taken for interest. Theall and Franklin (2001) also suggested that courses within the student's program area tend to be given low ratings. Perhaps the opportunity to select elective courses and courses outside their program provides students with options rather than requirements. Moreover, those courses offered outside of student programs and selected as electives may have a lower workload requirement.

Empirical evidence of this relationship between workload and student ratings is mixed, however. Some research suggests a positive and direct relationship between workload and student ratings (Marsh & Roche, 2000), perhaps because it presents a greater challenge and a sense of value in learning for the investment of time and effort. Students in these courses may also feel that they have learned more and, thus, give high ratings. Other studies, however, report that higher workload is related to lower student ratings (Greenwald, 2002; Kulik, 2001); if the course workload seems excessive and causes students a high degree of stress, it may result in low ratings.

Student Engagement

Student engagement is a multi-dimensional construct that includes active learning and collaboration (National Survey of Student Engagement, 2005). When students participate in class discussions and presentations, as well as converse with their instructors, they can develop and enrich their knowledge and skills (Brint, Cantwell, & Hanneman, 2008). These activities occur when students attend class regularly and are likely to create high expectations for good marks. These student-engagement behaviours can also lead to enhanced learning, which, in turn, may result in positive student ratings for the course. Indeed, Greenwald (2002) concluded that students who receive high marks are likely to give high ratings. Student attendance as another student-engagement characteristic that may influence student ratings has not been studied extensively, however. Students who attend classes frequently may be more motivated and interested in the course and may rate the teaching effectiveness as high, compared to students who attend only sporadically. Beran and Violato (2005), for example, found this to be the case.

Various course characteristics and student engagement may relate to student ratings of instructors. Course descriptions such as type and length may influence ratings; for instance, applied courses such as labs and long courses may receive high ratings. Course status may also affect ratings. That is, elective courses, courses outside the student's program of study, and courses requiring a moderate workload are likely to receive high student ratings. Given the small and sometimes mixed results of previous research on the relationship between course characteristics and student ratings, student engagement was introduced into the present study to determine its relative importance to student ratings. When students are actively involved in their learning, as demonstrated by frequent class attendance and high expectations for marks, they should give their instructors high ratings. The main purpose of the present study, therefore, was to test a latent variable path analysis (LVPA) model that integrates several course characteristics, student engagement, and student ratings of instruction. An advantage of applying LVPA to this problem is that it allowed us to identify and measure several latent variables and determine their interrelationships in a full model of student ratings of instruction.

METHODOLOGY

Sample and Procedure

A sample of 371,131 student ratings across all faculties at a major Canadian university over a three-year period (from the 1999 to the 2002 winter term) was obtained. Ratings were conducted at the end of every course, session, and term, with the majority of responses obtained during the fall term (i.e., December), followed by the winter term (i.e., April), which is consistent with student enrolment. The average response rate was 61%.

Instrument

Student ratings were derived with the Universal Student Ratings of Instruction (USRI) instrument, which consists of 12 items that were constructed based on other published student-rating measures used in research (e.g., Marsh, 1991). The responses provided in Table 1 were coded on a 7-point response scale, ranging from strongly disagree to strongly agree, with a higher score indicating a more positive rating. The internal consistency alpha reliability coefficient of the 12 items was .92, indicating that the scale is internally consistent. The structure of the USRI has been examined in previous research and is considered to be a unidimensional measure of teaching (Beran & Violato, 2005) that is consistent with other measures of teaching effectiveness (e.g., Greenwald & Gillmore, 1997). To ensure anonymity, student identification numbers were not recorded; therefore, the number of individual students who completed the ratings is unknown.

Table 1
Mean and Standard Deviation of the USRI items (n = 371,131)

	M	SD
The overall quality of instruction	5.65	1.24
Student questions and comments were responded to appropriately	5.94	1.16
The course content was communicated with enthusiasm	6.00	1.07
Students were treated respectfully	5.83	1.32
Opportunities for course assistance were available	6.04	1.22
The course outline or other descriptive information provided enough detail about the course	6.07	1.24
The course as delivered followed the outline and other course descriptive information	5.96	1.15
The course material was presented in a well-organized manner	6.24	1.08
The evaluation methods used for determining the course grade were fair	5.72	1.38
Students' work was graded in a reasonable amount of time	6.01	1.19
I learned a lot in this course	5.76	1.37
The support materials used in this course helped me to learn	5.61	1.39

In addition to rating their courses, students were asked about the characteristics of each course, specifically, its type (1 = lecture, 2 = lab, 3 = practicum, 4 = tutorial, 5 = distance), its duration (0 = half year, 1 = full year), and its status, that is, whether the course was in their program of study (1 = in department, 2 = not in department, 3 = unknown), whether it was required (1 = required course, 2 = required choice, 3 = elective), and how its workload compared to similar courses (1 = much lower, 2 = lower, 3 = same, 4 = higher, 5 = much higher).

In terms of their engagement with the course, students were asked to indicate their attendance rate for each course they were rating (1 = 0–20%; 2 = 21–40%; 3 = 41–60%; 4 = 61–80%; 5 = 81–100%) and the grade they expected to receive at the end of the course (11 = A, 10 = A-, 9 = B+, 8 = B, 7 = B-, 6 = C+, 5 = C, 4 = C-, 3 = D+, 2 = D, 1 = F). These characteristics are shown in Table 2.

RESULTS

When analyses of variance were conducted to examine student ratings against the variables listed in Table 2, the resulting effect sizes (η^2) were all low (.00 to .07). Thus, according to these characteristics, the ratings did not differ to a large extent. To further explore the relationships among course characteristics, student engagement, and student ratings, a two-step analytic procedure using two separate random subsamples of 2,000 was conducted. In the first step, principal component extraction with varimax rotation was used. Three factors emerged after five iterations, explaining 55.9% of the variance. The first factor, course status, which accounted for 23% of this variance, was measured by workload (factor loading = .54), course within department (factor loading = .78), and course required (factor loading = .82). The second factor accounted for 16% of the variance and reflected course description, with factor loadings of .67 for course type and .74 for duration. The third factor, student engagement, accounted for 17% of the variance and was measured by attendance (.75) and expected grade (.75).

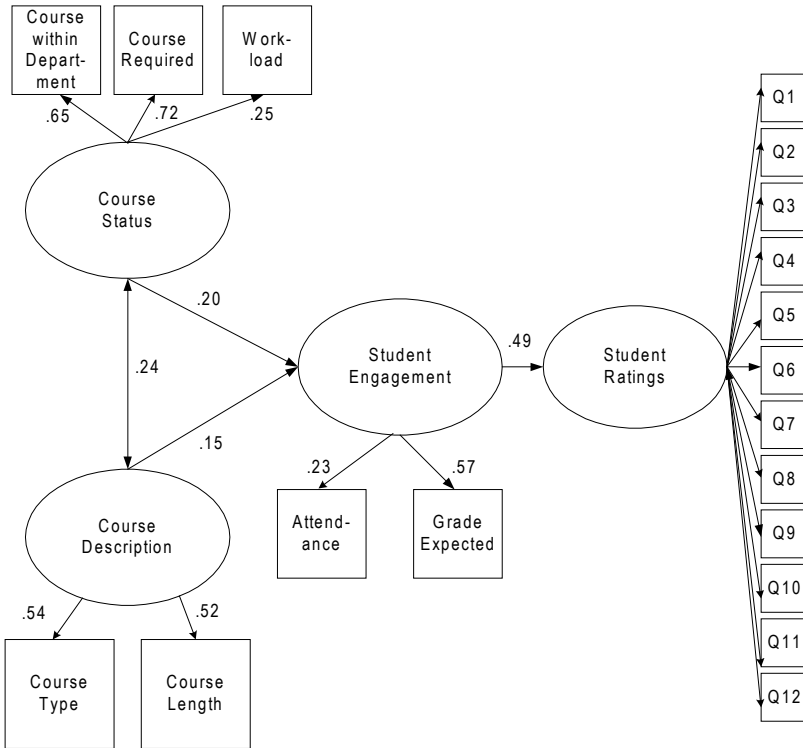
These three factors were then used to derive latent variables in the latent variable path analysis, using a second randomly drawn sample of 2,000 (see Figure 1). The model was re-specified until the best statistical fit was obtained, following the principle of parsimony. The model converged in nine iterations and provided a moderate fit for the data, $\chi^2(149) = 2304.40$, $p < .05$; standardized root mean square residual (SRMR) = .03; and comparative fit index (CFI) = .86. Thus, the model accounted for 86% of the variance-covariance in the data. There was a range of residual coefficients: .57 to .84 for the 12 USRI items; .70 to .97 for course status; .84 to .85 for course description; and .82 to .97 for student engagement.

As shown in Figure 1, student ratings differed according to student engagement and course characteristics. Student ratings are depicted on the right side of the model, as measured by all 12 USRI items. The latent variable “student engagement” was measured by the grade students expected to receive in the course and their frequency of attending the course. This variable was directly related to student ratings—students expecting high grades and attending class on a regular basis provided high ratings for the course. Two types of course characteristics are shown on the left side of the model. Course description was measured by length and type. This variable was related to course status, measured by course requirement, department status, and workload. Although all of these course variables influenced student engagement unidirectionally, they were not directly related to student ratings. Thus, student engagement mediated

Table 2
Course Characteristics and Student Engagement (n = 371,131)

Characteristics	n*	Percentage
Student		
Grade	54,148	14.6
A	75,875	20.4
A-	78,251	21.1
B	70,619	19.0
B-	33,309	9.0
C+	16,930	4.6
C	13,939	3.8
C-	5,446	1.5
D+	1,174	0.3
D	1,040	0.3
F	361	0.1
Missing	20,039	5.3
Attendance		
0-20	769	0.2
21-40	1,163	0.3
41-60	4,816	1.3
61-80	27,904	7.5
81-100	332,847	89.7
Missing	3,632	1.0
Course		
Length		
Half	352,209	94.9
Full	18,922	5.1
Type		
Lecture	330,927	89.2
Lab	7,510	2.0
Practicum	1,097	0.3
Tutorial	5,232	1.4
Distance	2,839	0.8
Other	3,717	1.0
Missing	19,809	5.3
Workload		
Much lower	6,457	1.7
Lower	37,997	10.2
Same	209,947	56.6
Higher	87,164	23.5
Much higher	27,057	7.3
Missing	2,509	0.7
Program		
In department	218,121	58.8
Not in department	126,417	34.1
Department unknown	23,116	6.2
Missing	3,477	0.9
Required		
Required course	190,102	51.2
Required choice	80,054	21.6
Elective	98,613	26.6
Missing	2,362	0.6

*Note.** number of responses



Note. Coefficient loadings for the 12 USRI items on Student Ratings range from .57 to .84.

Figure 1. Latent variable path model of student ratings employing Maximum Likelihood estimation ($n = 2,000$).

the effect of course description and course status on student ratings. However, the residual variances for student engagement and student ratings were high (.96 and .87, respectively). In other words, course description, course status, and student engagement were significant in the model, but there was considerable residual variance in student ratings. The same is true for the residual variance in student engagement. This result is consistent with the low effect sizes from the ANOVA analyses, which showed that student ratings did not differ greatly according to course characteristics and student engagement.

DISCUSSION

The main finding in the present latent variable path model is that course characteristics (course description and course status) are indirectly related to student ratings since they are mediated through student engagement, which directly influences student ratings. However, the large residual variances and modest model fit suggest that these characteristics explain student ratings to a limited extent only.

Course Characteristics

Course characteristics were not directly related to student ratings in the model. Thus, ratings did not vary as a direct function of the length and type of course; instead, the influence of the latter was mediated through student engagement. Although shorter courses provide fewer opportunities for learning than longer courses, learning may be condensed and students may be motivated to work harder over a shorter duration, compared to longer courses, which have more opportunities to procrastinate (Schraw, Wadkins, & Olafson, 2007). Also, compared to lecture-type courses, applied courses (e.g., labs, tutorial) did not receive higher ratings in the present study. Some research has shown that the invitation for student response within a lecture-based class increases learning (Blood & Neel, 2008). Thus, rather than measure differences across courses according to their type, future research should assess instructors' teaching styles within courses in relation to student ratings and student engagement. Our study may not have shown differences in student ratings across courses because these influences are felt only when mediated through student engagement.

The present study also showed that course-status characteristics are not directly related to student ratings but are mediated through student engagement. Courses outside the student's program of study and those taken as electives do not necessarily result in higher student ratings. Presumably, courses taken within a program area are meaningful because students selected their program of study. Thus, students may have similar interests in program courses, compared to out-of-program courses. Also, within programs there is often flexibility in choosing courses or course sections.

In terms of course workload, some previous research has shown that students give high ratings for high-load courses (Marsh & Roche, 2000); however, other research has shown that they give low ratings for high-load courses (Greenwald, 2002; Kulik, 2001). This was not the case in the present study (we found a small effect size). The discrepancies in the results from previous research and the small effect size in our study suggest that workload may be minimally related to student ratings and that other factors may mediate this relationship.

Student Engagement

Student engagement, as measured by attendance and expected grade, was related to student ratings in the model. Accordingly, students with frequent attendance and high grade expectations give course instructors high ratings. When students are motivated and interested in the course, they are likely to participate in classes and miss few of them. By asking and answering questions, discussing concepts, sharing examples, evaluating ideas, and applying knowledge, students are likely to develop mastery in their learning. Such accomplishment will likely create expectations for high marks.

Course status and description also predicted student engagement. Their engagement was higher in longer and more applied courses, as well as in those

courses that were outside their department, were electives, and had a higher workload. Perhaps greater interest in these courses increased the students' engagement in their learning. Indeed, courses that students selected and that were applied, longer, and required more work were likely to involve students more in learning about the subject.

Although the path coefficients suggest that student engagement mediates the relationship between course characteristics and student ratings, the residual variances of student engagement and student ratings were high. This suggests that ratings are determined to only a limited extent by student engagement. Similarly, student engagement is not largely determined by course characteristics. Thus, students who are not actively engaged in their learning may not necessarily give lower ratings to their instructors. Also, they may be engaged in all types of courses regardless of their characteristics. Perhaps students who find the class entertaining and informative without having to put effort into learning give positive ratings. Additionally, students who are interested in a particular topic may be actively engaged regardless of how the course is offered.

Other Factors

The present model suggests that factors other than course characteristics and student engagement determine how students rate their instructors. It is likely that characteristics about the instructor have more to do with effective teaching. Although a specific definition of effective teaching has not been identified (Goodwin & Stevens, 1993; Johnson & Ryan, 2000; Kulik, 2001), many descriptions focus almost exclusively on the instructional process (e.g., preparation of material, content knowledge). Arreola (1984), for example, regarded teaching as encompassing three broad dimensions: content expertise, instructional delivery skills, and instructional design skills. These dimensions relate to the instructional process as they reflect the skills and characteristics that promote or facilitate student learning. Lowman (1984) specified effective teaching according to intellectual excitement, which encompasses clarity and presentation of current materials, and to interpersonal rapport, which includes showing interest in students as individuals, encouraging creative and independent thought, and being warm, open, predictable, and student-oriented. Although not measured in the present study, these factors are likely stronger determinants of student ratings than student engagement and course characteristics.

Implications for Future Research

The present study has implications for future research. The other factors discussed above need to be examined in relation to student ratings, as well as in comparison to student and course characteristics, to determine their relative importance. Also, alternative measures of student ratings should be employed. The USRI instrument is unidimensional, and it is possible that multi-dimensional ratings will yield different results. Moreover, the residual variances of the latent

variables in our model indicate that these course and status descriptions are not good representations of the various types of courses taken by students.

Summary

Most researchers agree that teaching effectiveness may be defined as the degree to which an instructor facilitates student achievement (McKeachie, 1979). Despite instructors' anecdotal concerns about the potential influence of various factors outside their control (i.e., course characteristics and student engagement) on the amount that students learn, our study suggests that these play a small role in the ratings students give their instructors. Future research may well employ teacher characteristics and incorporate these into our proposed LVPA model, which includes student engagement as a latent variable. ♣

REFERENCES

- Abrami, P. C. (2001). Improving judgments about teaching effectiveness using teacher rating forms. In M. Theall, P. C. Abrami, & L. A. Mets (Eds.), *New directions for institutional research* (No. 109, pp. 59–87). San Francisco: Jossey-Bass.
- Abrami, P. C., d'Apollonia, S., & Cohen, P. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82*(2), 219–231.
- Arreola, R. A. (1984). Evaluation of faculty performance. In P. Seldin, *Changing practices in faculty evaluation: A critical assessment and recommendations for improvement* (pp. 79–85). San Francisco: Jossey-Bass.
- Arreola, R. A. (1995). *Developing a comprehensive faculty evaluation system*. Bolton, MA: Anker Publishing.
- Beran, T., & Violato, C. (2005). Ratings of teacher instruction: How much do student and course characteristics really matter? *Assessment and Evaluation in Higher Education, 30*(6), 593–601.
- Blood, E., & Neel, R. (2008). Using student response systems in lecture-based instruction: Does it change student engagement and learning? *Journal of Technology and Teacher Education, 16*(3), 375–383.
- Brint, S., Cantwell, A. M., & Hanneman, R. A. (2008). The two cultures of undergraduate academic engagement. *Research in Higher Education, 49*, 383–402.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51*, 281–309.
- Goodwin, L. D., & Stevens, E. A. (1993). The influence of gender on university faculty members' perceptions of "good" teaching. *Journal of Higher Education, 64*(2), 167–182.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11), 1182–1186.

Greenwald, A. (2002). Constructs in student ratings of instructors. In H. I. Braun & D. N. Jackson (Eds.), *The role of constructs in psychological and educational measurement* (pp. 277–297). New York: Erlbaum.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209–1217.

Johnson, T. D., & Ryan, K. E. (2000). A comprehensive approach to the evaluation of college teaching. *New Directions for Teaching and Learning*, 83, 109–123.

Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. In M. Theall, P. C. Abrame, & L. A. Mets (Eds.), *New Directions for Institutional Research* (No. 109, pp. 9–25). San Francisco: Jossey-Bass.

Lowman, J. (1984). *Mastering the techniques of teaching*. San Francisco: Jossey-Bass.

McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, 62, 384–397.

Marsh, H. W. (1991). A multidimensional perspective on students' evaluations of teaching effectiveness: A reply to Abrami and d'Apollonia (1991). *Journal of Educational Psychology*, 83, 416–421.

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92, 202–228.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (4th ed., pp. 13–103). New York: Macmillan Publishing.

Murray, H. G. (1984). The impact of formative and summative evaluation of teaching in North American universities. *Assessment and Evaluation in Higher Education*, 9(2), 117–132.

National Survey of Student Engagement (NSSE). (2005). *NSSE 2005 Annual Report: Exploring different dimensions of student engagement*. Bloomington, IN: NSSE.

Ory, J. C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? In M. Theall, P. C. Abrami, & L. A. Mets (Eds.), *New directions for institutional research* (No. 109, pp. 27–44). San Francisco: Jossey-Bass.

Schraw, G., Wadkins, T., & Olafson, L. (2007). Doing the things we do: A grounded theory of academic procrastination. *Journal of Educational Psychology*, 99(1), 12–25.

Sprinkle, J. E. (2008). Student perceptions of effectiveness: An examination of the influence of student biases. *College Student Journal*, 42(2), June, 276-293.

Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction. In M. Theall, P. C. Abrame, & L. A. Mets (Eds.), *New directions for institutional research* (No. 109, pp. 45-56). San Francisco: Jossey-Bass.

CONTACT INFORMATION

Tanya Beran
c/o Department of Community Health Sciences
Faculty of Medicine
University of Calgary, AB
T2N 4N1
tnaberan@ucalgary.ca

Tanya Beran is an associate professor in the Medical Education and Research Unit of the Faculty of Medicine at the University of Calgary. She teaches courses on research, assessment, and measurement. She has won awards for her research and has published several studies on teaching evaluation.

Claudio Violato is a professor in and the director of the Medical Education and Research Unit of the Faculty of Medicine at the University of Calgary. He specializes in medical education and educational psychology. In addition to 10 books, Dr. Violato has published more than 200 scientific and technical articles and reports in journals such as the *Canadian Journal of Surgery*, *Teaching and Learning in Medicine*, *Academic Medicine*, *Medical Education*, *Canadian Journal of Psychiatry*, *British Medical Journal*, and *Pediatrics*.