

A Review of Student Evaluation of Teaching: Applications to Veterinary Medical Education

Tanya N. Beran ■ Tyrone Donnon ■ Kent Hecker

ABSTRACT

Student evaluation of teaching is ubiquitous to teaching in colleges and universities around the world. Since the implementation of student evaluations in the 1970s in the US, considerable research has been devoted to their appropriate use as a means of judging the effectiveness of teaching. The present article aims to (1) examine the evidence for the reliability, validity, and utility of student ratings; (2) provide seven guidelines for ways to identify effective instruction, given that the purpose of student evaluation is to assess effective teaching; and (3) conclude with recommendations for the integration of student ratings into the continuous evaluation of veterinary medical education.

Key words: student ratings of instruction, student evaluation, teaching evaluation, veterinary medical education, higher education

Veterinary medical education (VME) is generally concerned with teaching the knowledge as well as professional and clinical skills necessary to provide care for animals, eradicate animal disease, and examine links among human, animal, and environmental health.¹ The effectiveness of teaching is typically evaluated through the use of student ratings. Given their widespread use in colleges and universities around the world, researchers since the 1970s have asked whether students' ratings of their courses and instructors are valid, reliable, and, indeed, useful.² The present article reviews the results of these early studies to determine whether there is sufficient evidence for the validity, reliability, and utility of student ratings. Seven guidelines for ways to define effective instruction are then presented, followed by specific recommendations for the integration of student ratings into the continuous evaluation of VME.

Accountability in education has become paramount.³ More than ever before, teachers must provide evidence of effective instruction. Indeed, most universities and colleges have introduced standard evaluation methods. In Canada and the United States ratings evaluate courses and their instructors, whereas in Europe and Australia ratings generally assess programs in their entirety. According to Marsh, student rating forms are arguably "the most thoroughly studied of all forms of personnel evaluation, and one of the best in terms of being supported by empirical research."⁴(p.369) Educational psychologists such as Herbert Marsh, Peter A. Cohen, Kenneth A. Feldman, and Philip Abrami have made significant contributions to the field of teaching evaluation. Evaluation methods typically include a questionnaire that students complete about the characteristics of the instructor and the course that they find most conducive to learning. These measures tend to assess classroom-based teaching through questions about lectures, problem-based learning, and seminars. Teaching evaluation methods have not been studied in the context of VME; rather, to date, students'

ratings have merely been reported.⁵ This article provides a summary of the research conducted on student evaluation and presents implications that are relevant for VME.

VALIDITY

Validity is a psychometric property of measures that is used to judge their adequacy.⁶ According to the seminal work of Cronbach,⁷ there are several forms of validity that provide conceptual and statistical evidence that student ratings accurately assess teaching. First, we review evidence of content validity, which refers to expert and/or stakeholder agreement on whether the questions accurately measure teaching.

Content Validity

Many student rating questionnaires, some of which have been researched, have been developed at individual institutions. In his review of the student ratings literature published from 1971 to 1995, which is when the majority of content validity studies were conducted, Greenwald concluded that most of the questionnaires demonstrate evidence of content validity.⁸

In recognition of teaching as a complex, multi-dimensional task,⁹⁻¹¹ many questionnaires include single items or groups of items that are meant to measure distinct dimensions. The most thoroughly studied student rating questionnaire was developed by Marsh.¹² Based on 30 published studies and the use of factor analyses, 35 items related to teaching can be divided into nine separate factors that appear to measure different dimensions of teaching. These include (1) learning, (2) instructor enthusiasm, (3) organization, (4) group interaction, (5) rapport with students, (6) breadth of coverage, (7) examinations/grading, (8) assignments/readings, and (9) workload/difficulty. Typically, however, rating questionnaires tend to be short and, thus, may not be capable of statistically identifying the multiple dimensions of teaching.¹³ Or, they may

include only a few dimensions such as the process (i.e., instructors' behaviors) and outcome of teaching (i.e., student learning).⁹ Others may include a list of teaching behaviors and characteristics of teaching based on instructors' and students' perceptions.¹⁴ Still others measure characteristics related to instructors (e.g., "Student questions and comments were responded to appropriately") separately from characteristics related to course (e.g., "The support materials help me to learn").¹⁵ For the majority of these rating scales, researchers have shown that ratings are unidimensional and that a halo effect occurs whereby students who are generally satisfied with the course give satisfactory ratings to all dimensions.^{15,16} We can conclude that, although teaching may very well be a multi-dimensional task, students' evaluations of it, particularly evaluations using short rating scales, are often global. Moreover, these findings suggest that items tend to focus more on the process of teaching rather than any theories of learning.¹⁷ In lieu of mapping specific teaching strategies to learning strategies, student rating scales tend to list various "ingredients" or traits of teaching, particularly those that are easily identifiable, such as instructor enthusiasm.¹⁸

Concurrent Validity

Concurrent validity examines the degree to which two different forms of student evaluation yield similar results. In a typical study of concurrent validity, researchers use two separate methods (e.g., student ratings and peer ratings) to measure teaching. If the results from the two methods are similar, the methods are determined to have adequate concurrent validity. Indeed, Kulik's review of several studies that used correlation analyses demonstrated evidence of concurrent validity for student ratings.¹⁹ For example, student ratings are similar to ratings given by alumni, colleagues, administrators, and neutral observers of teaching.^{14,20–22} In the context of VME, once effective teaching is specified according to expected learning outcomes for veterinary students and student rating forms are developed, these student ratings can be compared to another criterion, or indicator, of teaching effectiveness—such as teaching awards, anecdotal feedback, peer feedback, and observer ratings of teaching (e.g., teaching consultants)—to ensure that they demonstrate concurrent validity.

Predictive Validity

Predictive validity is similar to concurrent validity in that a measure of student ratings is compared to another measure. Predictive validity, in contrast, refers to a comparison with an expected outcome, such as student learning. Presumably, students who achieve the learning objectives are those whose teachers provide effective instruction.⁹ Indeed, Cohen's review of 41 such studies indicates that student achievement is related to student ratings.²³ That students learn more in courses that provide more effective instruction serves as the strongest evidence for the validity of student ratings.^{21,24} The implication for VME is that student ratings can be compared with measures of student outcome success, such as scores on licensing exams and clinical practice. In other words, those students who are receiving effective instruction would, presumably, be performing well on licensing exams and in clinical encounters.

Each content, concurrent, and predictive validity study that shows that student ratings are a valid indicator of teaching quality results in converging evidence for construct validity. That is, the construct (e.g., teaching effectiveness) can then be said to consist of the various behaviors (e.g., showing enthusiasm, organizing information, modeling clinical skill) that are assessed in the student rating measure used in the studies. The extent to which measures are able to accurately quantify these constructs defines the term construct validity.

Although numerous studies have presented evidence of several types of validity for student ratings, beliefs that they are biased persist.^{25,26} A common misunderstanding is that students misjudge teaching according to various factors. These include course characteristics, such as its status (required/elective), type (science/humanities), duration (half year/full year), time of year (fall/winter/spring), year of program (undergraduate/graduate), or mode of presentation (lab/lecture).^{4,27} Another factor is related to student characteristics, such as the grade they expect to achieve, their perceived workload, attendance, and lack of knowledge about teaching.¹⁶ Characteristics of the evaluation procedures, such as whether they are paper-based or electronic or whether students are required to include their student numbers on the evaluations, have also been suggested as biasing factors.²⁸ Moreover, instructor characteristics, such as rank and personality style, may also influence ratings.²⁹ Empirical results do not support these views.^{30–32} For example, Beran and Violato¹⁵ found significant differences in ratings for many of these characteristics; however, these differences affect student ratings only to a minimal extent. In other words, although significant, these characteristics have a small impact on student ratings. Beran and Violato conclude that students appear to base their ratings primarily on the quality of teaching instruction. Further analysis of teaching instruction indicates that organized delivery of the course, rather than any biasing characteristics, results in the highest ratings.¹³

RELIABILITY

In addition to the confirmation of validity, research has shown teaching evaluation measures to also be reliable indicators of effective instruction.^{2,5,9,15,16,18,23,26,27,33} Reliability refers to the consistency of scores in quantifying the construct being measured.⁶ Studies on consistency, stability, and generalizability have examined reliability across questions, students, raters, time, and contexts.³⁴ When these scores are similar, student ratings are judged to be reliable. Indeed, studies show that these forms of reliability for student evaluation are high.^{11,34} There is demonstrated stability, for example, between student ratings at the end of the course and ratings completed several years later.²² Moreover, students and graduates have been found to provide similar ratings of teaching, and instructors teaching different courses are found to receive consistent ratings.⁴ Thus, ratings across contexts are also shown to be reliable. In sum, there is considerable evidence for the various forms of reliability for student ratings.

Generalizability studies have shown that student ratings are more similar for an instructor teaching different

classes than for different instructors teaching the same class.^{4,11} A recent study indicates that ratings are required from at least five students to obtain a reliable estimate of instruction within a clinical encounter.³⁵ A different study suggests that 7 to 10 ratings are required.³⁶ It is recommended, therefore, that student evaluations in VME aim to obtain a minimum of five to seven ratings for each instructor.

UTILITY

Despite evidence of their validity and reliability, the utility—or usefulness—of student ratings is less accepted.³⁷ The utility of student evaluations has been examined for students, administrators, and instructors.^{38–41} Utility refers to the use or application of a particular measure.⁴² For a measure to have a high degree of utility, it must provide the type of information required to be used for its intended purpose, which varies across user groups. Students presumably use ratings when selecting courses and instructors; administrators use ratings as a summary measure of teaching quality which is used in making certain decisions, such as decisions about promotion and tenure; and instructors use the ratings to improve their teaching as well as the content and structure of their courses.¹⁴ In other words, the utility of student ratings refers to the extent to which they benefit their users. Although originally implemented with the intention of providing feedback to instructors to improve their teaching, instructors are the least likely of the three groups to use this feedback.¹⁵ Students, in addition, are generally not aware that they can access the feedback and are not likely to use the information for course selection.^{41,43} Administrators, on the other hand, are the most likely group to use the ratings and consider them helpful when making decisions about promotion, tenure, salary increments, and teaching awards.^{19,39} Instructors' general rejection of student ratings might arise from concerns about students' intentions, students' ability to judge effective teaching, and the lack of specific information that is derived from these questionnaires.^{44–46} Although alternative methods of evaluation, such as peer review and observations, do exist, they are generally more time consuming.

RECOMMENDATIONS FOR STUDENT RATINGS RESEARCH FOR VME

The primary consideration in evaluating instruction is defining its characteristics. Certainly, well-designed rating forms will contain questions modeled upon specific approaches to teaching, which are based on theories of learning.⁴⁷ If, for example, an instructor adopts a behaviorist approach, teaching methods will likely include a focus on clarifying course objectives and matching these objectives to course content.¹⁸ Questions asking whether instruction met course objectives would be included in student ratings. Instructors who espouse cognitive approaches will likely focus on classifying and ordering content; thus, student ratings would assess whether course material was well organized.¹⁸ An exhaustive review of learning theories is beyond the scope of this article (readers should refer to works by Bower and Hilgard⁴⁸ as well as Klein⁴⁹). There is an extensive number of items that can be included on student evaluation questionnaires. To guide the selection and development of such items,

we recommend using the seven principles for effective instruction that Kaufman and Mann⁵⁰ provided in a recent review of several widely adopted learning theories:

1. Learners are motivated to actively engage in their learning experiences and must be considered responsible for maximizing the benefits derived from the education process. This principle suggests that student evaluation should include questions about how students felt inspired, encouraged, and stimulated to learn as a result of the course.⁵¹
2. Learning is impacted by many characteristics within the learning environment, such as the availability of resources, organization of course content, and interactions with the instructor. Student evaluation should include, therefore, ratings of these types of factors that influence learning within the classroom.
3. Teaching must be relevant to the learner's needs. The knowledge, skills, and qualities of an effective veterinarian need to be identified and addressed in each course. The degree to which instructors teach and demonstrate these characteristics can be evaluated through student ratings.
4. Teaching must build on knowledge and experience so that existing understanding can be extended and re-organized to adapt to new information. This process serves to reduce redundancy and maximize learning within the finite duration of a course, and it can be evaluated by asking students to report the degree to which courses are well organized and incorporate concepts from previous courses.
5. Learning is more than the acquisition of knowledge and skills. It involves the learners' values, attitudes, and beliefs, which "are central to learners' willingness to attempt new actions."^{50(p.33)} Teaching activities must include reflective responses and interactions between students and teachers. Extended to student evaluation, this principle suggests that students evaluate their experiences in presenting their ideas and perceptions as well as the quality of the interactions they have had with faculty and other students.
6. Another theoretical approach to learning is the student's self-monitoring or self-regulation. These processes include determining learning needs, developing professional and learning objectives, and judging personal success. They can be evaluated by asking students to rate the degree to which they are self-motivated to learn and attain educational and professional goals.
7. Lifelong learning is the goal for most, if not all, theories of learning. Learning, thus, takes place in any environment and is driven by the learner. At the core of this principle is personal reflection. It involves self-awareness and a critical analysis of one's limitations. This self-evaluation then motivates plans for learning and progress. Applied to student ratings, it suggests that students should be asked to judge the extent to which they are aware of their progress in learning while attending the course.

Table 1: Examples of recommendations in practice

Steps	Examples
Planning phase	
1. Identify program goals	Program goals may be to provide a positive learning environment that fosters collegiality among students, faculty, and veterinarians, while promoting self-esteem, confidence, and personal and professional ethics.
2. Specify course objectives	The primary objective of a course may be for students to demonstrate the ability to understand when referral is required or additional professional input is needed to serve the needs of animals, clients, and society.
3. Identify effective instruction	Effective instruction occurs when the communication between instructors and students is positive and constructive (this step addresses Kaufman and Mann's principles 1, 3, and 5 ⁵⁰).
4. Identify user needs	Instructors may want to know which methods of teaching in a course were most helpful for student learning. Administrators may want to know whether students would select a course instructor for a teaching award. Students may want to know how an instructor was ranked in comparison to others who taught the same course.
5. Student input	Students can be asked "what type of feedback would you like instructors to receive?"
6. Consideration of value of posting	Faculty may want to discuss the benefit of posting ratings to students.
Implementation phase	
7. Other measures of student ratings	A well-researched instrument that would be useful for comparison is the Student Evaluation of Educational Quality. ⁶⁸
8. Consultant/facilitator support	University faculty development programs may provide this form of support upon request.
9. Student consultation	Students can be asked "have you seen a benefit from the evaluation methods you have participated in?"
10. Improve response rate	The administration of evaluations on paper toward the end of the course during class time may result in a higher response rate than electronic administration.
11. Midterm evaluation	Students can be asked to provide an evaluation half way through the course.

Overall, these attributes of learning focus on the role of the learner, the instructor, and the interaction, in addition to the specific education environment. These principles are also relevant to the learning experiences that students encounter outside of courses and while participating in any activity within a VME program. Student ratings, accordingly, must address the above seven principles of learning within specific courses and, more broadly, within the program.

Considering that there is extensive controversy about student ratings and yet tremendous opportunity for instructors to gain valuable feedback from students about teaching, it is important that evaluation procedures be carefully and deliberately established. Here are some additional strategies, recommended in student evaluation research, for effective use of student evaluations in VME:

Planning

1. Before implementing student evaluations, the program's goals must be identified.⁵²
2. Course objectives must also be clearly stated.^{53,54}
3. Faculty must be aware of what constitutes effective instruction.⁵⁴

4. The key finding from research on the utility of student ratings is that user groups have different needs. Instructors require feedback in the form of a formative assessment, which gives them information about areas in which to improve. Administrators, however, require a summative evaluation to know the outcome of teaching. Accordingly, one questionnaire that asks about general teaching experiences may be administered to all instructors for accountability purposes (these can be readily found in research on student ratings), and individualized measures can be developed based on instructors' specific goals for teaching. Students, on the other hand, may want to know how the instructor compares to other instructors.
5. Before implementing an evaluation system, instructors can be asked about the type of information they would like to receive from students about their teaching. If instructors perceive information to be valuable, they are likely to accept and consider it. In addition, students' perceptions of effective teaching can be surveyed to include them in rating scales. For example, students consider qualities such as organization, clarity, fairness, and respect toward students as the most important indicators of effective teaching.⁴¹

- At some universities, students are able to view instructors' ratings. Considering that many students may not be aware of these results⁴¹ and that instructors consider this publication to be a breach of confidentiality,⁵⁵ the value of posting student ratings should be scrutinized. Moreover, if students have little or no choice in their course selection, knowing individual instructors' ratings may be of little benefit to them.

Implementation

- Student evaluation questionnaires need to be pilot-tested and administered alongside other indicators of teaching effectiveness to determine the reliability and validity of their scores.⁵⁶
- Resources such as a consultant/facilitator to help instructors identify specific problem areas and improve teaching have been shown to significantly increase the usefulness of student feedback.⁵⁷
- Students can be consulted periodically during student evaluation planning because, as Marlin suggests, evaluation results are meaningless if students do not take the process seriously.⁵⁸
- A minimum of five to seven ratings from each course is needed to obtain a reliable estimate. In addition, responses must represent the views of all students in the class. Although no specific minimum response rate has been established in the research for the results to be considered reliable, some institutions set minimum rates according to class size.⁵⁹
- Studies suggest that to increase student satisfaction with the process, student rating questionnaires should be administered at midterm to allow instructors the opportunity to implement changes in their teaching.^{60,61}

The above list is not exhaustive; rather, it outlines key findings from the student evaluations research. Examples of how each recommendation can be implemented are presented in Table 1. These were obtained from discussions about student evaluations with VME faculty at the authors' institution. They may not be applicable to every program, but they do exemplify the integration of strategies into VME.

Student ratings of instruction have become integral to accountability in higher education,^{44,62} and they are not likely to be discontinued. Thus, university education programs must develop systems that optimize the usefulness of this exercise. The reliance on student ratings data in making personnel decisions such as promotion and tenure, however, concerns instructors who are not convinced of the reliability and validity of the ratings. Based on empirical research, most factors that are believed to influence student ratings actually have relatively little or no influence.^{21,32,60} Thus, it is important for instructors to be informed through teaching resources, such as workshops, about the research on student ratings so that their use does not result in reduced faculty morale or grading standards.^{15,25,63-66} Student ratings are intended to be used as one component of a multi-dimensional evaluation system

that may also include peer, self, and administrative evaluations.⁶⁷ Considering that veterinary medicine is critical to animal and human health, it is essential that instructors deliver effective instruction within a well-organized program, so that student learning can be maximized.

REFERENCES

- North American Veterinary Medical Education Consortium (NAVMEC). Roadmap for veterinary medical education in the 21st century: responsive, collaborative, flexible. Washington, DC: Association of American Veterinary Medical Colleges; 2010.
- Bembentuy H. The last word: an interview with Herbert W. Marsh; a leading voice on self-concept, teaching effectiveness, and a force in quantitative analysis (Part II). *J Adv Acade*. 2009;20(3):536-44.
- Altbach PG, Gumport PJ, Berdahl RO, editors. American higher education in the twenty-first century: social, political, and economic challenges. 3rd ed. Baltimore: Johns Hopkins University Press; 2011.
- Marsh HW. Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. *Int J Educ Res*. 1987;11(3):253-388. [http://dx.doi.org/10.1016/0883-0355\(87\)90001-2](http://dx.doi.org/10.1016/0883-0355(87)90001-2).
- Powers DE. Student perceptions of the first year of veterinary medical school. *J Vet Med Educ*. 2002;29(4):227-30. <http://dx.doi.org/10.3138/jvme.29.4.227>. Medline:12717641
- Hecker K, Violato C. Validity, reliability, and defensibility of assessments in veterinary education. *J Vet Med Educ*. 2009;36(3):271-5. <http://dx.doi.org/10.3138/jvme.36.3.271>. Medline:19861713
- Cronbach L. Essentials of psychological testing. 1st ed. New York: Harper; 1949.
- Greenwald AG. Constructs in student ratings of instructors. In: Braun HI, Jackson DN, Wiley DE, editors. The role of constructs in psychological and educational measurement. New York: Lawrence Erlbaum; 2002. p. 277-98.
- Abrami PC, d'Apollonia S, Cohen P. Validity of student ratings of instruction: what we know and what we do not. *J Educ Psychol*. 1990;82(2):219-31. <http://dx.doi.org/10.1037/0022-0663.82.2.219>.
- Feldman KA. Identifying exemplary teachers and teaching: evidence from student ratings. In: Perry RP, Smart JC, editors. Effective teaching in higher education: research and practice. New York: Agathon Press; 1997. p. 368-95.
- Marsh HW, Roche LA. Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. *Am Psychol*. 1997;52(11):1187-97. <http://dx.doi.org/10.1037/0003-066X.52.11.1187>.
- Marsh HW, Roche LA. The use of student evaluations and an individually structured intervention to enhance university teaching effectiveness. *Am Educ Res J*. 1993;30(1):217-51.

- 13 Donnon T, Delver H, Beran TN. Student and teaching characteristics related to ratings of instruction in medical sciences graduate programs. *Med Teach*. 2010;32(4):327-32. <http://dx.doi.org/10.3109/01421590903480097>. Medline:20353330
- 14 Feldman KA. Effective college teaching from the students' and faculty's view: matched or mismatched priorities? *Res Higher Educ*. 1988;28(4):291-329. <http://dx.doi.org/10.1007/BF01006402>.
- 15 Beran TN, Violato C. Ratings of university teacher instruction: how much do student and course characteristics really matter? *Assess Eval High Educ*. 2005;30(6):593-601. <http://dx.doi.org/10.1080/02602930500260688>.
- 16 Greenwald AG, Gillmore GM. Grading leniency is a removable contaminant of student ratings. *Am Psychol*. 1997;52(11):1209-17. <http://dx.doi.org/10.1037/0003-066X.52.11.1209>. Medline:9357333
- 17 Beran TN, Violato C. Student engagement and course characteristics. *Can J High Educ*. 2009;39(1):1-13.
- 18 Centra JA. Reflective faculty evaluation: enhancing teaching and determining faculty effectiveness. San Francisco: Jossey-Bass; 1993.
- 19 Kulik JA. Student ratings: validity, utility, and controversy. *New Dir Inst Res*. 2001;109:9-25. <http://dx.doi.org/10.1002/ir.1>.
- 20 Marsh HW. Student evaluation of university teaching: dimensionality, reliability, validity, potential biases, and utility. *J Educ Psychol*. 1984;76(5):707-54. <http://dx.doi.org/10.1037/0022-0663.76.5.707>.
- 21 Murray HG. Low inference classroom teaching behaviours and student ratings of college teaching effectiveness. *J Educ Psychol*. 1983;71:856-65.
- 22 Overall JU, Marsh HW. Students' evaluations of instruction: a longitudinal study of their stability. *J Educ Psychol*. 1980;72(3):321-5. <http://dx.doi.org/10.1037/0022-0663.72.3.321>.
- 23 Cohen PA. Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies. *Res Higher Educ*. 1981;51(3):281-309.
- 24 Patrick CL. Student evaluations of teaching: effects of the big five personality traits, grades, and the validity hypothesis. *Assess Eval High Educ*. 2011;36(2):239-49. <http://dx.doi.org/10.1080/02602930903308258>.
- 25 Gray M, Bergmann BR. Student teaching evaluations. *Academe*. 2003;89(5):44-6. <http://dx.doi.org/10.2307/40253388>.
- 26 Theall M, Franklin J. Looking for bias in all the wrong places: a search for truth or a witch hunt in student ratings of instruction. *New Dir Inst Res*. 2001;109:45-56. doi:10.1002/ir.3
- 27 Scherr FC, Scherr SS. Bias in student evaluations of teacher effectiveness. *J Educ Bus*. 1990;65:356-8.
- 28 Blunt A. The effects of anonymity and manipulated grades on student ratings of instructors. *Comm Coll Review*. 1991;18(4):48-54. <http://dx.doi.org/10.1177/009155219101800409>.
- 29 Feldman KA. The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: a review and synthesis. *Res Higher Educ*. 1986;24(2):139-213. <http://dx.doi.org/10.1007/BF00991885>.
- 30 Chen GH, Watkins D. Stability and correlates of student evaluations of teaching at a Chinese university. *Assess Eval High Educ*. 2010;35(6):675-85. <http://dx.doi.org/10.1080/02602930902977715>.
- 31 Marsh HW. Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *J Educ Psychol*. 1983;75(1):150-66. <http://dx.doi.org/10.1037/0022-0663.75.1.150>.
- 32 Seldin P. Changing practices in faculty evaluation: a critical assessment and recommendations for improvement. San Francisco: Jossey-Bass; 1984.
- 33 Lowman J. Mastering the techniques of teaching. San Francisco: Jossey-Bass; 1984.
- 34 Hobson SM, Talbot DM. Understanding student evaluations: what all faculty should know. *Coll Teach*. 2001;49(1):26-31. <http://dx.doi.org/10.1080/87567550109595842>.
- 35 Cottrell S, Smith T, Ferrari N, et al. A generalizability study of student ratings in logbooks to assess the educational process of clinical learning. *Acad Med*. 2010;85(7):1237-41. <http://dx.doi.org/10.1097/ACM.0b013e3181e00712>. Medline:20592522
- 36 Stalmeijer RE, Dolmans DHJM, Wolfhagen IHAP, et al. The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med*. 2010;85(11):1732-8. <http://dx.doi.org/10.1097/ACM.0b013e3181f554d6>. Medline:20881822
- 37 Ory JC, Ryan K. How do student ratings measure up to a new validity framework? *New Dir Inst Res*. 2001;109:27-44. <http://dx.doi.org/10.1002/ir.2>
- 38 Beran T, Rokosh J. Instructors' perspectives on the utility of student ratings of instruction. *Instr Sci*. 2009;37(2):171-84. <http://dx.doi.org/10.1007/s11251-007-9045-2>.
- 39 Beran T, Violato C, Kline D. What's the "use" of student ratings of instruction for administrators? One university's experience. *Can J High Educ*. 2007;37(1):27-43.
- 40 Beran T, Violato C, Kline D, et al. The utility of student ratings of instruction for students, faculty, and administrators: a "consequential validity" study. *Can J High Educ*. 2005;35(2):49-70.
- 41 Spencer KJ, Schmelkin LP. Student perspectives on teaching evaluation. *Assess Eval High Educ*. 2002;27(5):397-409. <http://dx.doi.org/10.1080/0260293022000009285>.
- 42 Hellman CM. Faculty evaluation by students: a comparison between full-time and adjunct faculty. *J Appl Res Community Coll*. 1998;6(1):45-50.

- 43 Beran T, Violato C, Kline D, et al. What do students consider useful about student ratings? *Assess Eval High Educ.* 2009;34(5):519–27. <http://dx.doi.org/10.1080/02602930802082228>.
- 44 Nasser F, Fresko B. Faculty views of student evaluation of college teaching. *Assess Eval High Educ.* 2002;27(2):187–98. <http://dx.doi.org/10.1080/02602930220128751>.
- 45 Schmelkin LP, Spencer KJ, Gellman ES. Faculty perspectives on course and teacher evaluations. *Res Higher Educ.* 1997;38(5):575–92. <http://dx.doi.org/10.1023/A:1024996413417>.
- 46 Wachtel HK. Student evaluation of college teaching effectiveness: a brief review. *Assess Eval High Educ.* 1998;23(2):191–212. <http://dx.doi.org/10.1080/0260293980230207>.
- 47 Marsh HW, Hocevar D. Students' evaluations of teaching effectiveness: the stability of mean ratings of the same teachers over a 13-year period. *Teach Teach Educ.* 1991;7(4):303–14. [http://dx.doi.org/10.1016/0742-051X\(91\)90001-6](http://dx.doi.org/10.1016/0742-051X(91)90001-6).
- 48 Bower GH, Hilgard ER. *Theories of learning.* Englewood Cliffs, NY: Prentice-Hall; 1981.
- 49 Klein B. *Contemporary learning theories.* Hillsdale, NJ: Lawrence Erlbaum; 1989.
- 50 Kaufman DM, Mann KV. Teaching and learning in medical education: how theory can inform practice. In: Swanick T, editor. *Understanding medical education.* New Jersey: Wiley-Blackwell; 2010. p. 16–36.
- 51 Hatzipanagos S, Rochon R. Introduction to the special issue: approaches to assessment that enhance learning. *Assess Eval High Educ.* 2010;35(5):491–2. <http://dx.doi.org/10.1080/02602938.2010.493700>.
- 52 Crosson AC, Boston M, Levison A, et al. Beyond summative evaluation: the instructional quality assessment as a professional development tool. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing; 2006. CSE Technical Report No.: 691.
- 53 Lewis KG. Using midsemester student feedback and responding to it. *New Dir Teach Learn.* 2001;87:33–44. <http://dx.doi.org/10.1002/tl.26>.
- 54 Ory JC. Faculty thoughts and concerns about student ratings. *New Dir Teach Learn.* 2001;87:3–15. <http://dx.doi.org/10.1002/tl.23>.
- 55 Howell AJ, Symbaluk DG. Published student ratings of instruction: revealing and reconciling the views of students and faculty. *J Educ Psychol.* 2001;93(4):790–6. <http://dx.doi.org/10.1037/0022-0663.93.4.790>.
- 56 Anderson HM, Cain J, Bird E. Online student course evaluations: review of literature and a pilot study. *Am J Pharm Educ.* 2005;69(1):34–43. <http://dx.doi.org/10.5688/aj690105>.
- 57 Brinko KT. The practice of giving feedback to improve teaching: what is effective? *J Higher Educ.* 1993;64(5):574–89. <http://dx.doi.org/10.2307/2959994>.
- 58 Marlin JW. Student perceptions of end of course evaluations. *J Higher Educ.* 1987;58(6):704–16. <http://dx.doi.org/10.2307/1981105>.
- 59 McGill University. Course evaluation policy and historical background [Internet]. Montreal: McGill University; 2011 [cited 2011 Mar 23]. Available from: <http://www.mcgill.ca/tls/courseevaluations/policy/>.
- 60 Abbott RD, Wulff DH, Nyquist JD, et al. Satisfaction with processes of collecting student opinions about instruction: the student perspective. *J Educ Psychol.* 1990;82(2):201–6. <http://dx.doi.org/10.1037/0022-0663.82.2.201>.
- 61 Dwinell PA, Higbee JL. Students' perceptional of the value of teaching evaluations. *Percept Mot Skills.* 1993;76:995–1000.
- 62 D'Apollonia S, Abrami PC. Navigating student ratings of instruction. *Am Psychol.* 1997;52(11):1198–208. <http://dx.doi.org/10.1037/0003-066X.52.11.1198>.
- 63 Armstrong JS. Are student ratings of instruction useful? *Am Psychol.* 1998;53(11):1223–4. <http://dx.doi.org/10.1037/0003-066X.53.11.1223>.
- 64 Eiszler CF. College students' evaluations of teaching and grade inflation. *Res Higher Educ.* 2002;43(4):483–501. <http://dx.doi.org/10.1023/A:1015579817194>.
- 65 Sproule R. Student evaluation of teaching: a methodological critique of conventional practices. *Educ Policy Anal Arch* [Internet]. 2000 Nov 2 [cited 2008 Jul 5];8(50):[about 23 p.]. Available from: <http://epaa.asu.edu/ojs/article/view/441>.
- 66 Trout P. Flunking the test: the dismal record of student evaluations. *Academe.* 2000;86(4):58–61. <http://dx.doi.org/10.2307/40251901>.
- 67 McGee R. Faculty evaluation procedures in 11 western community colleges. *Community Coll J Res Pract.* 1995;19(4):341–8. <http://dx.doi.org/10.1080/1066892950190406>.
- 68 Marsh HW. Student evaluation of educational quality (SEEQ): short version [Internet]. [cited 2011 Mar 12]. Available from: http://www.mta.ca/pctc/TONI_SEEQ/Student%20Evaluation%20of%20Educational%20Quality%20SHORT.pdf.

AUTHOR INFORMATION

Tanya N. Beran, BA, MSc, PhD, is Associate Professor, Medical Education Research Unit and Department of Community Health Sciences, University of Calgary, 3330 Hospital Dr. NW, Calgary, AB T2N 4N1 Canada. E-mail: tnaberan@ucalgary.ca. Her areas of research interests include evaluation and measurement as applied to teaching evaluation as well as psychological factors that influence student learning and career decision making.

Tyrone Donnon, BSc, BEd, MEd, PhD, is Associate Professor, Medical Education Research Unit and Department of Community Health Sciences, University of Calgary 3330 Hospital Dr. NW, Calgary, AB T2N 4N1 Canada. E-mail: tldonnon@ucalgary.ca. His research interests focus on teaching and assessment in medical education, learning theory and quantitative research methods as they pertain to measurement and evaluation in post-secondary education.

Kent Hecker, BSc, MSc, PhD, is Assistant Professor, Department of Veterinary Clinical and Diagnostic Sciences, Faculty of Veterinary Medicine, and Department of Community Health Sciences and Medical Education Research Unit, University of Calgary 3330 Hospital Dr. NW, Calgary, AB T2N 4N1 Canada. E-mail: kghecker@ucalgary.ca. Dr. Hecker holds a PhD in Medical Education Research with a focus on measurement and assessment.