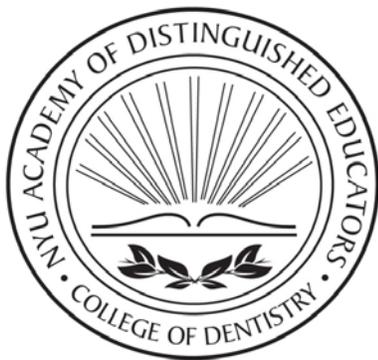


Considerations in evaluating teaching effectiveness in higher education

A White Paper developed by

New York University College of Dentistry
Academy of Distinguished Educators
Think Tank Committee on Student Evaluations

January 2015



Considerations in evaluating teaching effectiveness in higher education: A White Paper developed by New York University College of Dentistry Academy of Distinguished Educators Think Tank Committee on Student Evaluations. New York: 2015.

Copyright ©2015 New York University College of Dentistry. All Rights Reserved.

Table of Contents

Preamble	4
Introduction	4
Evidence from the Literature	6
Discussion	6
Suggestions	7
Environment: Control and Safety	7
Process and Tools	7
Testing and Research Opportunities	8
Conclusions	9
References	10
Acknowledgments	12

Preamble

The NYU College of Dentistry (NYU Dental) Academy of Distinguished Educators is a group of faculty members committed to elevating the quality of teaching and learning at the College. The Academy promotes innovation, experimentation, and evaluation of methods used for instruction and assessment and encourages approaches that are reasonable and evidence-based.

Past efforts of the Academy have focused on engaging and enabling faculty members to translate their teaching experiences to scholarship. This paper is part of the Academy's "think-tank" function in which conditions, policies, and practices at the College are examined and contextualized with suggestions to guide policy makers.

The process of topic selection comes from within the Academy or at the request of the leadership of the College. Academy members elect to participate on a topic specific working committee. Committee members reference the literature and offer suggestions that are reasonable and evidence-based, in a working draft. After the draft paper is developed, it is shared with communities of interest in an iterative process, where feedback and comments are received and incorporated. This paper was distributed to the general membership of the Academy, members of faculty council, and of the Executive Management Council---a group of deans and other leader-administrators from the Colleges of Dentistry and Nursing. Editorial decisions were made by consent of the committee.

Introduction

What is evaluation? The dictionary definition is judgment of the value or condition of (someone or something) in a careful and thoughtful way. So why do we need evaluations in education? Perhaps the best answer lies in the idea that we constantly strive to be better and need guidance (from feedback) to achieve our goals.

It is important for organizations to have regular "performance evaluations." This entails looking back and appraising positive and negative events to enhance the function of the organization. Evaluation of the educational curriculum is important and necessary at all levels of academia. The monitoring of programs and faculty is imperative to assure the adequacy of teaching and learning environments for students.

The most frequent sources of data to document faculty teaching effectiveness include peer review, self-evaluation, administrative evaluation, and student evaluation. However, how do we know the most effective way to evaluate? It is common practice to ask students to participate in course evaluations. Yet there is often confusion among students as to what a "course evaluation" really entails. Then there is the question of the accuracy of results. How can we assure participation that yields meaningful outcomes?

This paper summarizes a review of the evidence concerning evaluation of teaching effectiveness in higher education and offers suggestions to guide practice.

Evidence from the Literature

Annual evaluations to assess the function of the organization (including curriculum, programs, and faculty) are common in academic institutions.¹ Monitoring teaching effectiveness reflects the institution's commitment to continual improvement.²

Despite thousands of publications over the past 90 years on the assessment of teaching effectiveness, confusion remains.³ Unfortunately, many of the instruments used in faculty evaluation systems are “homemade,” and, therefore, suspect in their validity and reliability.⁴ Faculty tend to agree that the student evaluations weigh heavily in administrative decision making.⁵ However, reliance on student ratings alone to measure teaching effectiveness is not adequate.³ Valid inferences (i.e., judgments or interpretations from assessment) may rely on convergence from several different approaches, including student ratings, peer observation, and self-assessment. Without additional evidence, student ratings alone can lead to incorrect inferences and unfair decisions.⁶

There are advantages and disadvantages in using surveys completed by students to evaluate faculty teaching effectiveness. From an institutional perspective, the datasets can be used in the process of promotion and tenure as well as to guide faculty members toward improved pedagogical performance.⁷ Rater surveys provide direct feedback to the faculty and minimize bias compared to one-time outside reviews. Surveys are anonymous---a clear benefit considering Family Educational Rights and Privacy Act (FERPA) concerns and the potential for bias. Administratively, these tools are efficient and easy to implement. Operational advantages include low cost, rapid turnaround, and ease of scale revision. These benefits are magnified when surveys are delivered online.⁸ There are, however, concerns involving reliability and validity of these tools.

A number of confounders have been identified that distort evaluation results. Course evaluations are affected by the expected grade (students rated instructors better before they knew their grades). “Easy A” graders are essentially assured they will receive good evaluations. Instructor's gender, class size, and the physical environment (including ambience of the classroom)^{7,9} also distort evaluation results. Mandatory courses had lower student ratings than courses where there was a choice involved (e.g., electives). Larger classes tended to have lower student evaluation scores.¹⁰ Courses with higher workloads also had lower student rating scores.⁴ Moreover, students who attended lectures in person rated lecturers significantly higher than those who viewed podcasts of them. This phenomenon occurred regardless of students' year in [medical] school, the rated faculty member's degree, or the student's class grade or rank.¹⁰

Student rater evaluations reflect attitudinal perspectives and are not a performance measurement system for teaching or learning outcomes. There is a low to moderate correlation between student ratings of instruction and student achievement. The relationship between evaluations and teaching effectiveness is not clear, since grades are a confounder. There are, however, reports that support the use of student rater evaluation. One suggests that teachers whose students do well in achievement measures receive higher instructional ratings than teachers whose students do poorly, and there is a strong tendency for students to rate most highly those teachers from whom they learned the most.¹¹ Others after many years of research concluded that, when appropriately measured, student rater evaluations are reliable and stable and could be relatively unaffected by a variety of potential biases such as grading leniency, class size, workload, and prior subject interest.⁹ Several reports, including the Advisory Committee on Educational

Outcome Assessment, have proposed that assessor training may be the “missing link” in improving assessment quality.¹²

Discussion

Surveys are limited in providing feedback toward self-improvement in teaching skills and performance as they tend to focus on shortcomings rather than solutions. Summative evaluations are most commonly administered at the end of the semester or degree program and thus could be argued to be more suited to meeting the needs of the organization over those of the faculty member.

Effective evaluation entails the use of scientifically sound and practically feasible measurement instruments and processes. The conditions of administration should be controlled to minimize bias. Students are often not qualified to rate a professor’s knowledge and content expertise, teaching methods, use of technology, course materials, assessment instruments, and grading practices.¹³ These limitations may be addressed through evaluations by peers. Regardless, each evaluation tool should be tested for reliability and validated. Evaluation also involves faculty’s reflection on the evaluation results (preferably with colleagues) followed by tailor-made individual enhancement trajectories. In questionnaires completed by faculty in response to student rater evaluations, it appears that these evaluations can have a negative effect on morale and job satisfaction, and overall had a counterproductive [redundant] effect.⁵ When this paper was reviewed by communities of interest at NYU Dental, concerns were expressed by several stakeholders regarding the use of student rater evaluations for promotion and tenure decisions.

Acceptance of feedback is influenced by a recipient’s belief that the feedback is (1) credible, (2) accurate, (3) offered in a nonthreatening manner, and (4) from someone whom the recipient trusts. Although there is little empirical evidence to support the proposition that reflective practice improves the quality of teaching, faculty members who demonstrate higher levels of reflection had, on average, higher increases in student evaluations of teaching year to year.¹³ This finding supports the concept that reflective practice increases teaching quality. A small-to-medium size effect between improvements in teaching (based on consecutive student rater evaluations) and consultative feedback interventions was reported. Importantly, teaching performance is most positively affected by the learning climate of the educational program.¹⁴

Suggestions

- **Environment: Control and Safety**

Even before the process begins, faculty members should be included. What are their questions and concerns? Once criteria, methods, and evaluation processes are established, faculty members should be informed *before being evaluated*. This information should be available and accessible to all faculty members. Faculty members should feel safe knowing that the evaluation process and the datasets are temporal and instructive, not fixed or punitive. The emphasis should be on building constructive alliances for giving and receiving feedback, rather than assessing job performance.

- **Process and Tools**

Evaluation process should be recast in terms of an improvement process: student rater evaluations contextualized in a holistic process that pools data from multiple sources to enhance validity, including peer evaluation, self-evaluation, and student evaluation of teaching. The role of faculty members has shifted from that of an instructor to a facilitator, mentor, or coach. Some faculty members function as curators of learning environments. Others are involved in the development of learning environments. These learning environments often include active and interactive experiences that may be virtual and asynchronous. As education has shifted focus from teaching to learning, it makes sense to link both course and faculty evaluations to other assessments demonstrating student outcomes and progression toward competence.

Results of evaluation delivered in a timely fashion at multiple consecutive points give the object of evaluation opportunities to receive and reflect on feedback and potentially adjust/modify practices toward improvement. These results may be presented (in conjunction with other sources of feedback) by a supportive peer or mentor, guiding reflection and assisting/directing faculty members to develop a personalized improvement plan.

Instruments need to be validated and updated for their continuous use in various contexts as well as for specific groups, embedded in a system of feedback, support, and learning.

The forms or tools used should be clear in their purpose and intention. The test construction process should include considerations for reliability, internal consistency, and validity. The conditions for evaluation must identify and minimize confounders and sources of bias. Areas evaluated by students should be limited to those where a student is qualified to judge.

If institutions do not have the resources/expertise to develop reliable psychometric measurement tools, available validated resources can be used. For example, the System for Evaluation of Teaching Qualities (SETQ) was developed to evaluate feedback and enhance teaching performance of individual faculty. The SETQ instruments were initially modeled on the Stanford Faculty Development Program (SFDP26)¹⁵ instruments developed in the United States. Additionally, there are validated tools/questionnaires to evaluate clinical teachers.^{16, 17}

However, a systematic review concluded that the available instruments should be used carefully, as some still lacked consistency and reliability.¹⁸

- **Testing and Research Opportunities**

There are problems with current methods for evaluation and there will likely be problems with new methods employed. Therefore, in conjunction with evaluations, there should be ongoing efforts to test the test. Considering our large group sizes and the enormity of the task of evaluating so many objects (e.g., courses, instructors), this could be an opportunity for research to inform and guide best practices. Such research could include experiments in survey design and alternate modalities (e.g. focus groups, expert panels, essays, etc.) using qualitative methods.

Conclusions

There are no perfect tools for course or faculty evaluations. Even best practice models acknowledge difficulties in implementation, administration, and precision as well as doubt in interpreting results. However, being more mindful of elements of construction, design, and conditions for implementation with regard to reliability, accuracy, and bias while continually testing the process and tools should, with time, result in meaningful inferences. We hope that we can start the process afresh with clear intentions and a shared understanding of what is being evaluated.

As higher education has moved to focus on learning outcomes, perhaps course and faculty evaluations should be connected with assessment of learning. And, if so, our radar should now be directed at these assessments: Are they meaningful, reliable, valid measures of student's knowledge, skills, and achievement? Without processes to evaluate the tests, it is impossible to know what the totality of the educational program (including assessments, grades, evaluations, among others) really means in terms of students' progression toward competence and the effectiveness of the learning environment.

Finally, a cautionary note about student evaluations: An improperly designed process and/or tools may result in a dumbing down of standards as teachers chase favorable student ratings, resulting in the "easy A." True learning takes hard work, which may not always be the most enjoyable experience for students. Raising the bar can be unpopular and difficult for a faculty member if the culture places too much value on students' impressions.

References

1. Chambers DW, Boyarsky H, Peltier B, Fendler F. Development of a mission-focused faculty evaluation system. *J Dent Educ.* 2003 Jan;67(1):10-22.
2. Springfiled E, Gwozdek AE, Peet M, Kerschbaum WE. Using multiple methods to assess learning and outcomes in an online degree-granting dental hygiene program. *J Dent Educ* 2012 Apr;76(4):414-26.
3. Berk RA. Top five flashpoints in the assessment of teaching effectiveness. *Med Teach* 2013;35(1):15-26. doi: 10.3109/0142159X.2012.732247.
4. Annan SL, Tratnack S, Rubenstein C, Metzler-Sawin E, Hulton L. An integrative review of student evaluations of teaching: implications for evaluation of nursing faculty. *J Prof Nurs* 2013 Sep-Oct;29(5):e10-24. doi: 10.1016/j.profnurs.2013.06.004.
5. Ryan JJ, Anderson JA, Birchler AB. Student evaluations: the faculty responds. *Research in Higher Education* 1980;12(4): 317-333.
6. Jahangiri L, Mucciolo TW, Choi M, Spielman AI. Assessment of teaching effectiveness in U.S. Dental schools and the value of triangulation. *J Dent Educ.* 2008 Jun;72(6):707-18.
7. Sarkis J, Inshik S. Course evaluation validation using data envelopment analysis. *The Accounting Educators' Journal.* 2010;20:21-32.
8. Turhan K, Yaris F, Nural E. Does instructor evaluation by students using a WEB-based questionnaire impact instructor performance? *Adv Health Sci Educ Theory Pract.* 2005;10(1):5-13.
9. Marsh, HW, Roche LA. Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *Am Psychol.* 1997 Nov;52(11):1187-1197. doi: 10.1037/0003-066X.52.11.1187.
10. Martin SI, Way DP, Verbeck N, Nagel R, Davis JA, Vandre DD The impact of lecture attendance and other variables on how medical students evaluate faculty in a preclinical program. *Acad Med.* 2013 Jul;88(7):972-7. doi: 10.1097/ACM.0b013e318294e99a.
11. Chen Y, Hosshower B. Student evaluation of teaching effectiveness: an assessment of student perception and motivation. *Assessment & Evaluation in Higher Education.* 2003;28(1):71-88.
12. Dudek NL, Marks MB, Bandiera G, White J, Wood, TJ. Quality in-training evaluation reports--does feedback drive faculty performance? *Acad Med.* 2013 Aug;88(8):1129-34. doi: 10.1097/ACM.0b013e318299394c.
13. Winchester TM, Winchester MK. A longitudinal investigation of the impact of faculty reflective practices on students' evaluations of teaching. *British Journal of Educational*

- Technology. 2014 Jan;45(1):112-124. doi: 10.1111/bjet.12019.
14. Lombarts KM, Heineman MJ, Scherpbier AJ, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. *PLoS One*. 2014 Jan 28;9(1):e86512. doi: 10.1371/journal.pone.0086512. eCollection 2014.
 15. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med*. 1998 Jun;73(6):688-95.
 16. Fluit C, Bolhuis S, Grol R, Ham M, Feskens R, Laan R, Wensing M. Evaluation and feedback for effective clinical teaching in postgraduate medical education: validation of an assessment instrument incorporating the CanMEDS roles. *Med Teach*. 2012;34(11):893-901. doi: 10.3109/0142159X.2012.699114. Epub 2012 Jul 20.
 17. McGrath C, Wai Kit Yeung R, Comfort MB, McMillan AS. Development and evaluation of a questionnaire to evaluate clinical dental teachers (ECDT). *Br Dent J*. 2005 Jan 8;198(1):45-8.
 18. Fluit CR, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers. *J Gen Intern Med*. 2010 Dec;25(12):1337-45. doi: 10.1007/s11606-010-1458-y. Epub 2010 Aug 12.

Acknowledgements

This White Paper was developed by the NYU Dental ADE Think Tank Committee on Student Evaluations.

The paper was a collaborative endeavor. Committee members spent countless hours in research, manuscript preparation, and working with large and diverse groups of colleagues. This is the first think-tank paper developed and as such was a special challenge---especially in pioneering methods to facilitate dialogue with the broader community.

Committee members (in no particular order):

ANALIA VEITZ-KEENAN, DDS
Clinical Associate Professor
Oral and Maxillofacial Pathology, Radiology and Medicine
E-mail: av244@nyu.edu

SILVIA E. SPIVAKOVSKY, DDS
Clinical Associate Professor
Oral and Maxillofacial Pathology, Radiology and Medicine
E-mail: ss3617@nyu.edu

MITCHELL J. LIPP, DDS
Clinical Associate Professor
Orthodontics
E-mail: mitchell.lipp@nyu.edu

Although the paper was reviewed by many, special acknowledgments to the following individuals for their detailed analysis and thoughtful suggestions: Dean Charles Bertolami, Dr. Ivy Peltz, Dr. Leila Jahangiri, Dr. Mark Wolff, and Dr. William Eidtson.

