Assessment of Teaching Effectiveness in U.S. Dental Schools and the Value of Triangulation


Abstract: The routine evaluation of teaching effectiveness is important in improving faculty, departmental, and institutional efforts. There are three main categories of assessments: those performed by students, peers, and self. Although each category is independently valid, a collection of data from all three categories leads to a more comprehensive outcome and a creation of a triangulation model. The purpose of this study was to identify commonly used methods of assessing teaching effectiveness and to suggest the use of a triangulation model, which has been advocated in the literature on performance assessment as an optimal approach for evaluating teaching effectiveness. A twelve-question survey was sent to all U.S. dental schools to identify evaluation methods as well as to find evidence of triangulation. Thirty-nine out of fifty-seven schools responded. The majority of the schools used student evaluations (81 percent) and peer reviews (78 percent). A minority of schools reported using self-evaluations (31 percent). Less than one in five dental schools reported using all three strategies to achieve triangulation (19 percent). The three most commonly used evaluation methods (“performed routinely”) were all in the student evaluation category. Less than half of the schools routinely evaluated clinical teaching effectiveness by any means (42 percent). In conclusion, dental schools should implement a triangulation process, in which evaluation data are obtained from students, peers, and self to provide a comprehensive and composite assessment of teaching effectiveness.

Teaching effectiveness can be defined as the extent to which the teaching activity fulfills its intended purpose, function, and goal. Although goals can vary across institutions and may be specific or global, evaluating a teacher’s effectiveness is proof of the academic institution’s commitment to continual improvement. This study reviews survey findings to determine the most commonly used methods for assessing teaching effectiveness in U.S. dental schools. An evaluation process is a review of faculty performance; when appropriately implemented, the process results in objective decisions regarding promotions and recommendations. The outcomes from assessments of teaching effectiveness can be used for formative and summative purposes. Formative assessments are used to develop or improve persons or programs. Summative assessments are used to determine “final” achievement and to provide accountability, as the results are more directly related to the efficacy of teaching. Summative evaluations focus on the outcome of teaching rather than the process. A comprehensive method of assessing teaching effectiveness should include various types of evaluation strategies, specifically those of the formative nature.

A wide-ranging collection of data or evidence from multiple and diverse sources is described in the education literature as triangulation. This composite assessment methodology is thought to be more valuable because it compensates for the inadequacies that may be related to any one assessment method. According to Berk, using a triangulation methodology provides “a more accurate, reliable, and comprehensive method of teaching effectiveness than just one source.” Many other experts on performance assessment in the academic...
arena recognize that the measurement of teaching effectiveness is a complex process, and they strongly endorse using multiple sources of evidence. In an attempt to balance out some of the shortcomings of each corner of the “triangle.” The three most frequently used data sources that comprise the triangle are student ratings, peer observations, and self-assessments. In summary, teaching effectiveness can be assessed using diverse evaluation strategies, the combination of which is referred to as triangulation. For a triangulation method to be most effective, data from different sources should ideally overlap to some extent. For example, similar questions could be asked of students, peers, and self. The extent to which triangulation models are used to evaluate teaching effectiveness in dental education has not been determined. The purpose of this study was to identify commonly used methods of assessing teaching effectiveness in U.S. dental schools and to propose a model, based on the concept of triangulation, which includes evaluations from students, peers, and self.

Methods of Measuring Teaching Effectiveness

Student Assessment

Teaching invariably involves interaction with students; therefore, a huge emphasis has been given to student evaluations, ratings, feedback, and performance. The validity of student evaluations has been a subject of much debate (Table 1). A meta-analysis of forty-one research studies indicated that student evaluations/ratings should be an essential component of evaluating teaching effectiveness. A large number of students can respond to a survey on teaching effectiveness, making the results reliable. Moreover, the respondents are those who have observed the teachers on a number of occasions and have been personally affected by the teaching, giving this type of evaluation a high face validity.

Studies also support the use of retrospective “student” evaluations, such as exit surveys and alumni ratings, as additional methods of assessing past teaching effectiveness. Presumably, there is less of a correlation with exams or assignments and more of a relationship to the value of the education, the individual’s future goals, and the real world. However, these ratings are not necessarily any more reliable than current student evaluations.

Some have questioned whether students have the wisdom and the experience to make qualified observations and evaluations about teaching quality. In addition, it has been suggested that students’ evaluations may be overly influenced by their perceptions of faculty members’ leniency or severity in assigning grades (e.g., lenient, “easy” graders receive higher ratings than strict graders), which may undermine the value of students’ course and teacher evaluations and induce skepticism among the faculty in the value of these ratings. In fact, there is conflicting evidence and opinion in the faculty assessment literature concerning the effect that students’ expectations for high grades have on their evaluation of teachers. Some studies totally support the leniency hypothesis, while others found no evidence that leniency affects ratings. There appears to be little or no support in the literature that evaluations are affected by course content-related elements such as the course workload or difficulty. or even the student’s grade point average. Yet, there is some evidence that students’ assessment of teaching and teachers may be somewhat influenced by students’ prior interest in the content or subject.

A number of studies find little or no evidence that gender differences (instructor or student) play a role in determining evaluation outcomes. Some studies note that the characteristics of the instructor, such as personality and popularity, may have an effect on the evaluations, although the “popularity” myth has been debunked, for the most part. There are even reports that ratings are influenced by the presence of the instructor in the room while evaluations are being completed. Some studies recommend that a third party administer and collect course/teacher evaluation forms to protect student anonymity and reduce the “psychological presence” of the instructor upon students’ decision making about ratings.

Evidence suggests that past or present students should not be asked to comment on the adequacy or timeliness of a course or curriculum. Nor should they be required to judge the appropriateness of related assignments or exams. Faculty (peers) are considered more suitable for such evaluations. While evaluations are completed more accurately when students are given adequate time and instruction, the content
Whether the extent of learning, as indicated by test performance, is directly related to ratings and whether student feedback actually improves teaching are both debated in the literature. In spite of the controversy, the current belief is that student ratings and evaluations are highly valuable as determinants of teaching effectiveness and that the ratings provide a feedback mechanism for continuous improvement. It appears that the majority of researchers and experts in academic performance assessment believe in the reliability and validity of student evaluations as a means of evaluating teaching.
Peer Assessment

Peer review by faculty colleagues is considered another valuable assessment method for measuring teaching effectiveness. Peers can be either formative or summative. Peer review is used to enhance good teaching practices, develop faculty, and share resources. Peer evaluation has a high validity for judging course goals and objectives, course content, and appropriateness of tests or assignments, something that students are not able to effectively assess. Formative peer reviews are viewed more favorably than summative reviews by both faculty who participate in colleague assessment programs and by experts in performance assessment. Consequently, most assessment authorities recommend that peer evaluation should be formative—namely, a mentoring process that

<table>
<thead>
<tr>
<th>Peer-Related Issue</th>
<th>Agree</th>
<th>Disagree</th>
<th>Open to Debate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friendship bias</td>
<td>Berk et al.</td>
<td>Cederblom and Lounsbury</td>
<td>Love, Cederblom and Lounsbury</td>
</tr>
<tr>
<td>Ratings are very generous, skewed to the upper scale</td>
<td>Berk et al., Root</td>
<td>Braskamp and Ory</td>
<td></td>
</tr>
<tr>
<td>Rating scales do not measure effective teaching characteristics</td>
<td>Berk et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal collegial relationship</td>
<td>Shortland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value of the feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promotes self-knowledge and personal development through observation of others</td>
<td>McKeachie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations of others (reflective approach)</td>
<td>Shortland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualifications or competence of observers</td>
<td>Berk et al., Root</td>
<td>Martin and Double</td>
<td></td>
</tr>
<tr>
<td>No evidence of instructor self-improvement</td>
<td>Cosh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad feelings, defensiveness, interpersonal conflict</td>
<td>Cosh, Bridges</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of confidence in the outcomes or reliability of results</td>
<td>Braskamp et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear and resistance</td>
<td>Jarzabkowski and Bone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal subjectivity</td>
<td>Berk et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited number of observers</td>
<td>Berk et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation is time-consuming</td>
<td>DeZure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation of a single class is not representative of an entire course</td>
<td>Braskamp et al., Centra</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reputation, tenure, or faculty status influences observer bias (&quot;halo&quot; effect)</td>
<td>Berk et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender, race, sexual orientation bias</td>
<td>Berk et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observers exchange positive ratings for one another (back scratching)</td>
<td>Martin and Double</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function or use of peer input or feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous peer ratings</td>
<td>H. Hoyt and Pallet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevancy of teaching portfolio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings related to learning or achievement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Literature findings of factors influencing peer evaluations
focuses on providing feedback to a colleague for purposes of enhancing teaching, but the content of the feedback should not be used as a data source for formal performance evaluations or promotion and tenure decisions.

It has been suggested that friendship bias, reputation, tenure or faculty status, informal collegial relationship, and personal subjectivity can result in the skewing of faculty peer evaluations.\textsuperscript{9,47,48,60,98,101-106} Therefore, faculty tend to be more accepting of peer evaluations when the outcomes are used for feedback purposes and collaborative reflection rather than as a basis for formal performance assessment, which may have wage implications for promotion.\textsuperscript{34,98,103,107,108}

Studies have suggested that peer evaluation can create bad feelings, defensiveness, and interpersonal conflict among faculty.\textsuperscript{5,103,104,109} This can also result in fear and resistance from those being evaluated.\textsuperscript{8,9,98,104,107,110} The limited number of observers who are involved in the faculty peer review process has been described as a disadvantage,\textsuperscript{9} although recent use of videos of teaching in action allows for more observers with a greater scheduling flexibility for review.\textsuperscript{111,112}

Others question the qualifications or competence of peers to function as reviewers of teaching effectiveness.\textsuperscript{5,9,101,103,106,107} There is also debate as to whether the rating scales actually measure effective teaching characteristics\textsuperscript{9,16} or result in instructor improvement.\textsuperscript{9,102-104,107,110,113}

Although there are reports that peer evaluations are biased, unreliable, unfair, and limited in scope and may lead to overly positive reviews if used for formal (summative) performance evaluations,\textsuperscript{60,103,104,107,114,115} peer reviews are necessary to augment other assessment methods, such as student ratings or self-evaluations.\textsuperscript{1,5,47,98,104,107} Where summative peer evaluations are performed, they should be done with the same rigor as a peer review process for assessment of research and scholarly activity, and they may result in teaching awards.\textsuperscript{3} For this purpose, many academic institutions encourage the use of teaching portfolios to formalize the process.\textsuperscript{8,9,111,116-119} In summary, peer evaluations, especially formative ones, are valuable when conducted in conjunction with other types of evaluations to counter the potential bias of the peer evaluator.

**Self-Assessment**

Self-evaluation or self-assessment is another method used in evaluating teaching effectiveness (Table 3).\textsuperscript{1,56,112,116,117,120} Self-assessment involves a faculty self-report that describes teaching, scholarship, service, and practice. Although this self-report is not a true assessment of teaching, it has merits mainly by enforcing a reflective approach for self-development.\textsuperscript{15,34,103,104,107,108,116,121} One process used for self-evaluation, specifically for formative purposes, is the review of videos of teaching in action.\textsuperscript{1,112} Lee and Wu\textsuperscript{112} have suggested five potential benefits of the use of videos: 1) easy access to video content, 2) comprehensive view of teaching styles, 3) shared expertise, 4) venue for constructive feedback, and 5) greater utilization of peer-to-peer learning.

However, some investigators contend that the reliability of self-assessment is low because individuals have been shown to provide overly positive self-appraisals in comparison to other data sources\textsuperscript{47,56,122,123} and because assessments may be inaccurate, overstated, or self-serving, especially when it comes to summative decisions.\textsuperscript{14,16,56,116,120,124}

<table>
<thead>
<tr>
<th>Table 3. Literature findings of factors influencing self-evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-Related Issue</strong></td>
</tr>
<tr>
<td>Gain awareness of different teaching styles through observation (reflective approach) for self-development</td>
</tr>
<tr>
<td>Subjective view of ability (inflated view of skill) or leniency effect</td>
</tr>
<tr>
<td>Validity and fairness of using own portfolios or self-reports</td>
</tr>
<tr>
<td>Efficacy—reliance on own skill</td>
</tr>
<tr>
<td>Influence from peers (peer input and feedback)</td>
</tr>
<tr>
<td>Ratings related to learning or achievement</td>
</tr>
</tbody>
</table>
The real value of self-reflection is in the opportunity to demonstrate one’s own perception of teaching effectiveness.\textsuperscript{15} An underexplored question in the literature on assessment of teaching effectiveness is: does teacher self-assessment support or contradict student and peer evaluations? Self-evaluation is a common form of assessment among universities and higher education, and has a greater validity when used in conjunction with other teaching effectiveness strategies, such as peer and student evaluations.\textsuperscript{1}

**Triangulation Model**

Amidst all of the debate regarding methods of assessing teaching effectiveness (Tables 1, 2, and 3), there is recognition of independent value in using student, peer, and self-evaluations as methods of assessing teaching effectiveness. However, when these methods are combined, a triangulation model is achieved, resulting in a more comprehensive assessment outcome.\textsuperscript{9,14-16}

Berk has identified twelve different strategies to measure teaching effectiveness.\textsuperscript{1} These include elements from students, peers, and self. Based on a modification of these strategies, we developed a survey for applications specific to dental education in order to identify the incidence of commonly used methods of assessing teaching effectiveness in U.S. dental schools and to find evidence of use of triangulation.

**Methods**

A twelve-question multiple-choice survey was created with four options per question (yes; if yes, performed routinely; no; not sure), plus a comments section (Table 4). The survey was sent electronically via the American Dental Education Association.
(ADEA) Academic Affairs listserv to academic deans of all U.S. dental schools in July and again in August of 2007. Each school was allowed only one submission. Out of fifty-seven U.S. dental schools, thirty-nine submitted survey responses. Three submissions were eliminated because, in each case, the schools had answered only one of the twelve questions. Results of the remaining thirty-six surveys were tabulated and analyzed to determine commonly used methods of teaching effectiveness for an actual response rate of 63 percent.

Additionally, the following process was used to determine the percentage of schools that employ a triangulation model (Table 5). Where a school had a response of “yes, performed routinely” in each category of student (question 1, 2, or 8), peer (question 3, 4, 5, 9, 10, or 11), and self (question 6 or 7) review, the school was identified as having achieved triangulation (Table 6). Question 12, pertaining to a formal mechanism of assessment of clinical teaching effectiveness, was excluded in the definition of triangulation because it was not specific to any single defined category (student, peer, or self).

### Results

The survey results are shown in Table 4, including the written comments that were provided by seven of the thirty-six schools. For all questions, there was a higher response to an evaluation method simply being used, in contrast to its being “performed routinely.” Of the eleven strategies (questions 1-11), the most commonly used methods “performed routinely” were student written evaluations or ratings (81 percent); exit survey of students (75 percent); and learning outcome measures, exams, and national boards (72 percent). The three least routinely used assessment methods were self-evaluation by reviewing videos of teaching in action (0 percent); peer ratings (6 percent); and peer observation of lectures, seminars, and/or videos (17 percent). Less than half (42 percent) of the schools routinely performed clinical teaching effectiveness evaluations.

To determine the extent of triangulation, the survey questions were assigned to one of the three evaluation categories (Table 5). Review of the survey results (Table 6) illustrates that, of the three evaluation categories (student, peer, and self), the most commonly used method (“performed routinely”) was student evaluation. Nearly 20 percent of the responding schools reported using all three categories of student, peer, and self, and thus potentially had implemented triangulation in the performance assessment process.

### Discussion

Routine performance of evaluations is important for the development of the faculty, the department, and the institution. To optimize the value of the assessment outcomes, such evaluations should include elements from student, peer, and self-evalu-

---

**Table 5. Questions assigned to categories for determining triangulation**

<table>
<thead>
<tr>
<th>Evaluation Category</th>
<th>Related Question</th>
</tr>
</thead>
</table>
| Student             | • Student written evaluations or ratings  
|                     | • Student interviews, verbal evaluations  
|                     | • Exit survey of students  
|                     | • Learning outcome measures: exams, national boards, etc.  
| Peer                | • Peer review of written documents such as course syllabi, course goals and objectives, lectures, outlines  
|                     | • Peer observation of lectures, seminars, and/or videos  
|                     | • Peer ratings  
|                     | • Teaching awards by institution  
|                     | • Teaching portfolios  
| Self                | • Faculty self-evaluation in a written form  
|                     | • Self-evaluation by reviewing videos of teaching in action  

**Table 6. Number of schools and related percentages for those responding “Yes, performed routinely” to each evaluation category: student, peer, self**

<table>
<thead>
<tr>
<th>Evaluation Category</th>
<th>Responded “Yes, performed routinely” Number of Schools (n=36)</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>29</td>
<td>80.56%</td>
</tr>
<tr>
<td>Peer</td>
<td>28</td>
<td>77.78%</td>
</tr>
<tr>
<td>Self</td>
<td>11</td>
<td>30.56%</td>
</tr>
<tr>
<td>Triangulation (Student and Peer and Self)</td>
<td>7</td>
<td>19.44%</td>
</tr>
</tbody>
</table>
Student evaluations are valuable when related to course organization and instructor delivery, but not when students are asked to evaluate course content, relevancy, or appropriateness of goals and objectives. Many dental schools, however, value student views regarding curriculum. The ADEA Senior Survey is primarily devoted to asking students their opinions about curricular content, and this evaluation tool is widely used in dental schools to justify curriculum decisions. Although the ADEA survey is valuable in providing a scope of different curricula nationally, its interpretation for all curriculum changes should be restricted because of student limitations in judging the appropriateness of content while they are still in school and prior to experiencing the realities of actual dental practice in the community.

Peer reviews are more appropriate than student assessment for content evaluation. However, sustaining a viable peer review system may be time-consuming and resource-intensive. Many faculty members may be reluctant to embrace another evaluation process, considering that they are overwhelmed with existing job responsibilities and tenure requirements.

Another challenge to a peer review process is related to the current culture and environment in dental schools. Faculty are likely to hesitate in critiquing one another for fear of creating bad feelings that can damage future working relationships. In reality, however, the promotion and tenure system at most major universities now expect submission of peer appraisals of various aspects of academic performance—more commonly on the research side, but increasingly for teaching. Dental schools will need to determine a way to successfully implement peer reviews of teaching for summative purposes. For example, when selecting individuals to perform summative peer reviews, if faculty are allowed to identify their own reviewers, then “buddy bias” is a significant issue. If department chairs designate peer reviewers, there is less potential for friendship bias, but faculty may fear that chairs are out to “get them” by purposely choosing reviewers who have interpersonal conflicts, biases, or other issues with the reviewed faculty member. These realities need to be considered when using peer evaluations for summative purposes.

Peer assessments, together with self-evaluations, have the greatest value in focusing faculty on a process of routine self-reflection and determining or redirecting faculty efforts. One method of evaluation—review of video of teaching in action—warrants a discussion. The review of videos is a well-described mechanism for self- or peer evaluation. Based on the findings from this survey, it does not appear that any U.S. dental schools utilize “videos of teaching in action” as an evaluation mechanism. Video review may be more timely now than ever before. Students today are not limited to note-taking or traditional transcription. There is an increasing use of technology for recording voice and video of lectures and other teaching activities. These digital recordings can be stored,
reviewed, shared, analyzed, and used as tools for training and development. The fact that students are using this technology for study or review purposes makes it even more critical that the faculty evaluate these videos for teaching effectiveness, prior to or in conjunction with the student review.

While not directly related to the main issue in this study, the fact that only 42 percent of the responding schools reported that they routinely evaluated clinical teaching effectiveness is disconcerting. This is a low percentage considering that more than half of the students’ total time in dental school is in the clinic/patient care environment and this phase of the curriculum certainly is the students’ most powerful learning experience. It is in the best interest of dental schools to focus on evaluating clinical teaching effectiveness with the same level of commitment that is given to evaluating classroom teaching effectiveness.  

The process of evaluating teaching effectiveness can only have value if there is a forum for formalized feedback of these outcomes to create action plans for continuous improvement. In this manner, the students, the faculty, and the institution will greatly benefit from these evaluations.

Future studies are indicated to determine an optimal combination of evaluation methods for each category of student, peer, and self in achieving triangulation.

Limitations of the Study

The survey sent to the U.S. dental schools was used to identify the incidence of commonly used methods of assessing teaching effectiveness and to find evidence of triangulation. The survey was intended to collect data on methods used, from which a triangulation model could be proposed. However, none of the questions in the survey directly or indirectly asked respondents whether the school used a triangulation method. At the time of survey construction, it was perceived that the term “triangulation” was not commonly used or understood among dental faculty. Although triangulation is well defined in the education literature, its use in the dental literature is uncommon. For those schools that showed potential evidence of triangulation, there is no way to determine whether their collection of data from multiple sources was consciously done to implement a triangulation strategy.

Conclusion

The results of this study are as follows:

1. It was determined that a triangulation method for assessing teaching effectiveness was potentially performed by only 19 percent of the responding schools, although there was no evidence derived from this survey to determine if the one in five schools that reported multiple data sources have consciously determined to implement a triangulation process.

2. The three most commonly used evaluation methods, identified as “performed routinely,” were all from the student evaluation category (written evaluations, exit surveys, and exam results).

3. The self-evaluation category was found to be the least utilized area for assessment (31 percent) as compared to the peer (78 percent) and student (81 percent) categories.

4. Fewer than half of the responding schools routinely evaluated clinical teaching effectiveness (42 percent).

5. To improve the assessment of teaching effectiveness, the use of a triangulation model is suggested, whereby evaluation methods from all three categories of students, peers, and self are performed routinely.

REFERENCES


84. Svinicki MD. Encouraging your students to give feedback. New Dir Teach Learn 2001;87(3):17–24.