Title
Lecture Evaluations by Medical Students: Concepts That Correlate With Scores

Authors
Jen, A
Webb, EM
Ahearn, B
et al.

Publication Date
2015

DOI
10.1016/j.jacr.2015.06.025

Peer reviewed

Lecture Evaluations by Medical Students: Concepts That Correlate With Scores

Aaron Jen, BA, Emily M. Webb, MD, Bren Ahearn, MA, David M. Naeger, MD

Abstract

Purpose: The didactic lecture remains one of the most popular teaching formats in medical education; yet, factors that most influence lecturing success in radiology education are unknown. The purpose of this study is to identify patterns of narrative student feedback that are associated with relatively higher and lower evaluation scores.

Methods: All student evaluations from our core radiology elective during 1 year were compiled. All evaluation comments were tagged, to identify discrete descriptive concepts. Correlation coefficients were calculated, for each tag with mean evaluation scores. Tags that were the most strongly associated with the highest- versus lowest-rated (> or < 1 SD) lectures were identified.

Results: A total of 3,262 comments, on 273 lectures, rated by 77 senior medical students, were analyzed. The mean lecture score was 8.96 ± 0.62. Three tags were significantly positively correlated with lecture score: "interactive"; "fun/engaging"; and "practical/important content" ($r = 0.39$, $r = 0.34$, and $r = 0.32$, respectively; all $P < .001$). More tags (n = 12) were significantly negatively correlated with score; the three tags with the strongest such correlation were: "not interactive"; "poorly structured or unevenly paced"; and "content too detailed or abundant" ($r = -0.44$, $r = -0.39$, and $r = -0.36$, respectively; all $P < .001$). Analysis of only the highest- and lowest-rated lectures yielded similar results.

Conclusions: Several factors were identified that were strongly associated with lecture score. Among the actionable characteristics, interactive lectures with appropriately targeted content (ie, practical/useful) were the most highly rated.

Key Words: Medical students, medical education, lectures, teaching, feedback, evaluations

J Am Coll Radiol 2016;13:72-76. Copyright © 2016 American College of Radiology

INTRODUCTION

The didactic lecture is one of the most time-tested and popular teaching methods used today. As a simple method of transferring knowledge from instructor to student, its benefits are many. The format is scalable, allowing for instruction of tens of students up to thousands, particularly when information is disseminated via the Internet [1]. In addition, the format is very familiar, and can be time efficient for preparing and delivering content.

The traditional lecture format, however, is prone to several pitfalls [2]. Students' attention spans for passive learning are extremely limited, often to just 20 minutes [3]. Lecturer styles of presentation are unique, and sometimes ineffective, requiring learners to adapt to each presenter. Additionally, lecturers choose the level of information and the pace of presentation, which may not be ideally suited to all learners. Finally, lectures are limited in what they can teach; they often focus on facts and concepts, rather than on skills or creativity [4].

Due to these challenges, some educators have argued that alternative teaching formats should have a greater role in education. Problem-based and team-based learning are implemented widely in current medical school curricula [5-7]. Blended and "flipped" learning models are popular [8-10]. Particularly now, in the digital age, online interactive modules increasingly are being promoted.

Each method offers its own unique advantages and disadvantages, yet none has completely replaced traditional lecturing. For that reason, educators should continue to improve the effectiveness of lecture delivery. Efforts to improve lecturing have come far in recent years. Several authors have recently shared insightful techniques to promote active learning, including incomplete outlines,

Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, California.

Corresponding author and reprints: David M. Naeger, MD, Department of Radiology and Biomedical Imaging, 505 Parnassus Ave, M-391, Box 0628, San Francisco, CA 94143-0628; e-mail: david.naeger@ucsf.edu.

The authors have no conflicts of interest related to the material discussed in this article.

1546-1440/15/$36.00 ■ http://dx.doi.org/10.1016/j.jacr.2015.06.025

break activities, relatable examples, and teachable skills [4,11-14].

Kessler et al [15] compiled presentation techniques from highly rated faculty at the American College of Emergency Physicians, focusing on key objectives, increased audience participation, and manageable slide content. Multimedia design principles have been applied to lecture creation. Issa et al [16], for example, found that audiences preferred a lecture design focused on visual representation, and without bullet points. To our knowledge, very little of the radiology literature has examined what characteristics are most desired by learners, particularly medical students.

In this study, we analyzed medical student feedback on lectures in our core senior radiology elective. Specifically, we analyzed all narrative comments collected during 1 year, and determined the comment phrasing that was most associated with various average numeric evaluation scores. Our hypothesis was that high- versus low-scoring lectures would be associated with a definitive set of unique adjectives.

METHODS

This study is exempt from institutional review board requirements. Only anonymized lecture comments and scores, extracted from an existing course-evaluation database, were reviewed.

Lecture Evaluations: Scores and Comments

Feedback is routinely obtained for all lectures in the core senior student radiology elective at our institution. Anonymized scores and comments from all iterations of the course taught in 2014 were included in this retrospective analysis. Each lecture's numeric score (range: 1 [worst] to 10 [best]) was an average rating from all students in attendance; the free-text comments consisted of all written comments. Providing both a numeric score and a written comment (even if just one word) was required. The actual content of the evaluations, which were originally collected for the purpose of course improvement, was anonymous, a fact known to the students providing the evaluations and comments.

Data Extraction

Deidentified free-text comments for each lecturer were reviewed for specific adjectives or singular concepts, and semantically tagged by one investigator. A list of tags was generated based on the content encountered. If a sentence contained an adjective or concept that was the same as or synonymous with one tagged in a previously reviewed comment, it was tagged with the same label. If a concept was entirely new, it was added to the list. The final list of discrete adjective families contained 42 tags. In cases in which comments were difficult to tag, all the investigators reviewed the comments and decided on an appropriate categorization, by consensus. This process was needed for approximately 5% of comments.

Each anonymized student comment (a sentence, sentence fragment, or rarely, multiple sentences) for each anonymized lecture was given a "yes or no" value for each of the 42 tags. Most comments were only one sentence, so most tags were not represented by any given comment. Each lecture was assigned a percentage frequency for each of the 42 tags. For example, if a lecture was evaluated by 12 students, and 4 stated that it was "interactive," and 3 stated that it was "too long," then 33% of the comments were positive for the tag "interactive," and 25% were positive for the tag "too long."

Statistical Analysis

The correlation between lecture score (range: 5.7-10.0) and tag percentage (range: 0%-80%) was analyzed using the nonparametric Spearman rank-order correlation coefficient. To determine a statistical significance cutoff, a Bonferroni correction was applied to account for the many analyses. A total of 42 correlation analyses were assessed, so an overall significance level of $P < .05$ was selected and divided by 42; individual analyses with $P < .0012$ were considered statistically significant.

We sought to confirm the correlation analysis by analyzing only the most-extreme groups of lectures, specifically, the highest- and lowest-scoring lectures (as defined by a score >1 SD away from the mean). The frequencies of each tag for lectures in the highest- versus lowest-scoring group were compared using the Mann-Whitney U test. A Bonferroni correction was applied as well, so individual analyses with $P < .0012$ were considered statistically significant.

RESULTS

The feedback from 317 lectures given by 54 different lecturers was available in the course-evaluation database. Forty-four sessions were excluded from analysis, owing to nontraditional lecture formats (eg, orientations, tours, and hands-on modules). These formats differed fundamentally from the traditional lecture model, and could, theoretically, have been evaluated differently by the students. The final sample included 273 lectures given by 48 lecturers. These lectures were evaluated by a total of 77

Table 1. Complete list of tags in order of correlation with lecture scores

Comment Tags	Spearman Correlation Coefficient	*P* value*
Interactive†	0.3937	**<.0001**
Fun or engaging†	0.3410	**<.0001**
Practical or important content†	0.3180	**<.0001**
Well-organized, systematic, or straightforward	0.1901	.0016
Instructor is clear	0.1488	.0139
Instructor is concise	0.1231	.0421
Focused topic	0.1030	.0895
Utilizing repetition via review or quiz	0.0839	.1670
Detailed, informative, or thorough	0.0617	.3100
Taught to appropriate level	0.0329	.5885
Relaxed or casual	0.0294	.6287
Good reference materials	0.0230	.7055
Funny	0.0211	.7285
Good images	0.0202	.7397
Including key points or objectives	0.0181	.7659
Good slides	0.0175	.7738
Clinical or case-based content	−0.0012	.9838
Challenging	−0.0089	.8836
Receptive to questions	−0.0156	.7978
Well-paced	−0.0201	.7414
Enthusiastic	−0.0301	.6203
Friendly	−0.0302	.6189
Biased or unprofessional	−0.0723	.2341
Instructor lacks expertise in topic	−0.0944	.1197
Broad in scope (introductory, review of topic)	−0.0948	.1182
Images are lacking or unclear	−0.1180	.0515
Lecturer is unfriendly	−0.1548	.0105
Timing issues (including starting late or running long)	−0.1615	.0075
Slides require improvement	−0.1692	.0051
Pace is too fast	−0.1795	.0029
Instructor is unclear	−0.1991	**.0009**
Repeated content from other lectures	−0.2036	**.0007**
Interesting or unfamiliar content	−0.2255	**.0002**
Difficulty too basic	−0.2477	**<.0001**
Lacking key points or objectives	−0.2498	**<.0001**
Lacking clinical correlation or examples	−0.2753	**<.0001**
Too advanced or requiring introductory material	−0.2880	**<.0001**
Content is not relevant	−0.3050	**<.0001**
Boring or slow	−0.3102	**<.0001**
Content is too detailed or abundant	−0.3577	**<.0001**
Poorly structured or unevenly paced†	−0.3896	**<.0001**
Not interactive†	−0.4367	**<.0001**

*Statistically significant correlations with lecture score, as defined by a Bonferroni-corrected *P* value < .0012, are indicated in boldface type.
†Comments that had statistically significant associations (*P* < .0012) with either the highest- or lowest-rated lecture group compared with the other.

senior medical students (each belonging to one of six cohorts that enrolled in the course). A total of 3,262 comments were analyzed, each representing a unique combination of lecture topic, lecturer, and student. The mean lecture score was 8.96 (SD ± 0.62), with a mean number of comments per lecture of 11.95 (range: 4-15 [SD ± 2.83]).

Forty-two tags were generated based on the content of feedback (a complete list is presented in Table 1). Several tags are opposites, such as "organized" and "disorganized." Although two concepts may be opposites, students did not always employ only one or the other (ie, they were not mutually exclusive), so both the positive and negative versions of such tags were included separately. For some tags, no distinct opposite appeared in the comments. For example, students mentioned that some lectures were "funny" (one of the 42 tags), but none was described as "humorless" (which therefore is not on the list).

Association Between Lecture Scores and Specific Medical Student Comments

Three tags were positively associated with the numeric scores (ie, they were more frequently encountered in lectures with high scores). These were: "interactive" ($r = 0.39$); "fun/engaging" ($r = 0.34$); and "practical/important content" ($r = 0.32$), respectively; all $P < .001$.

A longer list of 12 tags was significantly negatively associated with lecture score (ie, they were more frequently encountered in lectures with *lower* scores). The three most extreme examples were: "not interactive" ($r = -0.44$); "poorly structured or unevenly paced" ($r = -0.39$); and "content is too detailed or abundant" ($r = -0.36$), respectively; all $P < .001$. The full list of tags is presented in Table 1, which indicates those that are statistically associated and those that are nonassociated.

Comments With Greatest Discrepancy Between High- and Low-Scoring Groups

Evaluating only the subset of lectures with scores >1 SD away from the mean (39 with high ratings; 40 with low ratings), associations between specific tags and high- or low-rated lectures identified an overlapping, but much smaller, set of tags. Three positive tags and two negative tags, respectively, were in the high- or low-scoring lectures significantly more often (Table 1).

DISCUSSION

Via a systematic review of 3,262 medical student comments, we examined the association between specific phrasing in evaluation comments and the overall numeric rating for a lecture. We found that students described high-scoring lectures as interactive, engaging, and practical. We found that a longer list of descriptors was associated with low-scoring lectures, including lack of interactivity, poor structure, and an overabundance of content.

The descriptors most associated with positively rated lectures shared two themes: attention-sustaining characteristics and practicality. The "attention sustainment" component comprised, first, the lecture being "interactive." Interactivity can take on many forms, but nearly all are skills that can be taught to educators and honed through time. The second "attention sustainment" component was the degree to which a lecture was considered "fun or engaging." This concept is more difficult to define and teach to educators, and it may in part reflect individual teaching style and personality.

The second positive theme, "practicality," should not be overlooked. Many students gave positive feedback when lecturers covered useful and practical topics, such as commonly seen conditions. Learners are more motivated to absorb information when the content seems immediately useful and pertinent. Medical students, in particular, are soon to face residency training that requires an enormous skillset and dauntingly high stakes. Practical information is highly valued. Honing a lecture to deliver practical information is an achievable skill, much like incorporating interactivity. Surveying and conversing with students about what topics are most helpful can provide invaluable guidance, as can reviewing such items as the content of licensing examinations.

Twelve comment tags were associated with poorly received lectures. The most strongly negative characteristic was "not Interactive." Students seemed to expect some degree of interactivity, and they were willing to complain when it was absent as much as they were willing to praise its use. Additional negatively associated tags that were the opposite of a positive tag included "boring or slow" (opposite of "fun/engaging"), and a slew of tags that were the opposite of "practical or important content," including content described as too detailed, not relevant, too advanced, too basic, lacking clinical importance, and repeating previously covered content.

The value of organization and clarity in lectures was revealed most clearly through students' negative comments (the positive version of the tag was just below the threshold for statistical significance, as we defined it). Giving a well-organized and clear lecture does not make content more useful, or a presentation more engaging; its presence may not make a lecture, but its absence can doom one.

Several of our tags were descriptors that were not statistically significantly associated with lecture score. Examples include enthusiasm, humor, and image quality. Although these may be positive qualities, they were less associated with the overall score, compared with the concepts described earlier.

Our results can guide educators who wish to improve existing presentations or create new ones. First, lecturers should seek out any and all opportunities to be interactive. Holding a back-and-forth conversation, and posing questions to the audience, are tactics possible with smaller groups. For larger groups, audience response systems may be considered [17]. Second, speakers should aim to present useful and practical content. Understanding audience knowledge level, and targeting learning objectives appropriately, can guide educators. Finally, lectures should be clear and organized. Centering a talk on a few main points, and providing a simple and sensible outline, can help.

We found the majority of the statistically significant positive and negative tags to be reasonable, although we were surprised that "interesting or unfamiliar content" was considered negative. We suspect that students used the word "interesting" as a euphemism for "not useful," and considered unfamiliar content to have little practical value. In addition, we were surprised that "funny" as a standalone adjective did not have a significantly positive association. Students preferred "engaging" lecturers, which may in part reflect humor when used to maximize learning, but humor on its own did not seem to be valued as highly.

Audience feedback, if properly applied, is a valuable resource for improving medical education [18,19]. In addition, it has played an important role in the advancement of lecture theory [20-24]. For example, Sherbino and Bandiera [20] analyzed feedback from emergency medicine

residents and faculty and found that faculty members focused their comments on increased clarity, engagement, and enthusiasm; whereas residents prioritized content and difficulty. Cao et al [23] used feedback to find that residents preferred lectures with higher image quality, but fewer images per case. Similarly, recent work by Larocque et al [24] compared feedback with presentation details, finding that students preferred larger font sizes and a higher proportion of text-only slides. In general, acquiring and using feedback from all educational activities allows for an ongoing feedback cycle and continuous improvement.

Our study has several limitations. First, we have implied here that a positive lecture score indicates a better educational outcome. Although the fact that a lecture is well received may reveal some degree of teaching success, the student perspective alone does not tell the whole story, or reflect any quantitative assessment of learning. Feedback systems ideally should prioritize the educational value of an activity, although sometimes, feedback is undoubtedly influenced by instructor popularity. Additionally, students may not possess the experience to judge fully what content is truly practical. Finally, our results may pertain to only radiology lectures, given the emphasis on solely imaging rather than lectures in other medical specialties.

In summary, our study identifies positive and negative phrases, as used in medical student evaluations of radiology lectures, that were strongly associated with lecture scores. "Interactive," "engaging," "useful," and "organized" lectures tend to receive the highest scores. Educators can be mentored to create lectures with these characteristics in an effort to improve their repertoire of teaching materials.

TAKE-HOME POINTS

- Teaching evaluations often contain free-text narrative comments, which can offer a wealth of information to the educator.
- By analyzing more than 3,000 narrative comments regarding lectures in our main radiology course, we identified certain phrases that are most associated with well received and poorly received lectures.
- "Interactive," "engaging," "useful," and "organized" lectures tend to receive the highest scores.
- Educators can be mentored to create lectures with these characteristics in mind; in particular, methods of adding interactivity and selecting level-appropriate material can be taught to educators as a way to improve their teaching effectiveness.

REFERENCES

1. Martin FG. Will massive open online courses change how we teach? Commun ACM 2012;55:26-8.
2. Flexner A. Medical education in the United States and Canada: a report to the Carnegie Foundation for the Advancement of Teaching. Bulletin No. 4., 1910. Available at: http://archive.carnegiefoundation.org/publications/medical-education-united-states-and-canada-bulletin-number-four-flexner-report-0. Accessed May 11, 2015.
3. Mustafa T, Farooq Z, Asad Z, et al. Lectures in medical educaton: what students think? J Ayub Med Coll Abbottabad 2014;26:21-5.
4. Webb EM, Naeger DM, Fulton TB, Straus CM. Learning objectives in radiology education: why you need them and how to write them. Acad Radiol 2013;20:358-63.
5. Koh GCH, Khoo HE, Wong ML, Koh D. The effects of problem-based learning during medical school on physician competency: a systematic review. CMAJ 2008;178:34-41.
6. Neville AJ. Problem-based learning and medical education forty years on. A review of its effects on knowledge and clinical performance. Med Princ Pract 2009;18:1-9.
7. Hunt DP, Haidet P, Coverdale JH, Richards B. The effect of using team learning in an evidence-based medicine course for medical students. Teach Learn Med 2003;15:131-9.
8. Belfi LM, Bartolotta RJ, Giambrone AE, Davi C, Min RJ. "Flipping" the introductory clerkship in radiology: impact on medical student performance and perceptions. Acad Radiol 2015;22:794-801.
9. Sadeghi R, Sedaghat MM, Sha Ahmadi F. Comparison of the effect of lecture and blended teaching methods on students' learning and satisfaction. J Adv Med Educ Prof 2014;2:146-50.
10. Mehta NB, Hull AL, Young JB, Stoller JK. Just imagine: new paradigms for medical education. Acad Med 2013;88:1418-23.
11. Jakee K. Overhauling technical handouts for active student participation: a model for improving lecture efficiency and increasing attendance. Int J Teach Learn High Educ 2011;23:98-108.
12. Graffam B. Active learning in medical education: strategies for beginning implementation. Med Teach 2007;29:38-42.
13. Richardson D. Don't dump the didactic lecture; fix it. Adv Physiol Educ 2008;32:23-4.
14. Sandhu S, Afifi TO, Amara FM. Theories and practical steps for delivering effective lectures. J Community Med Health Educ 2012;2:158.
15. Kessler CS, Dharmapuri S, Marcolini EG. Qualitative analysis of effective lecture strategies in emergency medicine. Ann Emerg Med 2011;58:482-489.e7.
16. Issa N, Mayer RE, Schuller M, Wang E, Shapiro MB, DaRosa DA. Teaching for understanding in medical classrooms using multimedia design principles. Med Educ 2013;47:388-96.
17. Moss K, Crowley M. Effective learning in science: the use of personal response systems with a wide range of audiences. Comput Educ 2011;56:36-43.
18. Richardson JTE. Instruments for obtaining student feedback: a review of the literature. Assess Eval High Edu 2005;30:387-415.
19. Kember D, Leung DYP, Kwan KP. Does the use of student feedback questionnaires improve the overall quality of teaching? Assess Eval High Educ 2002;27:411-25.
20. Sherbino J, Bandiera G. Improving communication skills: feedback from faculty and residents. Acad Emerg Med 2006;13:467-70.
21. Burgess A, Wright C, Qasabian R, O'Mara D, Mellis C. Surgical teaching program for our senior medical students: room for improvement. Adv Med Educ Pract 2014;5:369-75.
22. Chapman T, Chew FS. Introductory lecture series for first-year radiology residents: implementation, investment and assessment. Acad Radiol 2013;20:332-7.
23. Cao L, McInnes MDF, Ryan JO. What makes a great radiology review course lecture: the Ottawa radiology resident review course experience. BMC Med Educ 2014;14:22.
24. Larocque N, Kenny S, McInnes MDF. Medical school radiology lectures: What are determinants of lecture satisfaction? AJR Am J Roentgenol 2015;204:913-8.