

A curious case of the phantom professor: mindless teaching evaluations by medical students

Sebastian Uijtdehaage & Christopher O'Neal

CONTEXT Student evaluations of teaching (SETs) inform faculty promotion decisions and course improvement, a process that is predicated on the assumption that students complete the evaluations with diligence. Anecdotal evidence suggests that this may not be so.

OBJECTIVES We sought to determine the degree to which medical students complete SETs deliberately in a classroom-style, multi-instructor course.

METHODS We inserted one fictitious lecturer into each of two pre-clinical courses. Students were required to submit their anonymous ratings of all lecturers, including the fictitious one, within 2 weeks after the course using a 5-point Likert scale, but could choose not to evaluate a lecturer. The following year, we repeated this but included a portrait of the

fictitious lecturer. The number of actual lecturers in each course ranged from 23 to 52.

RESULTS Response rates were 99% and 94%, respectively, in the 2 years of the study. Without a portrait, 66% (183 of 277) of students evaluated the fictitious lecturer, but fewer students (49%, 140 of 285) did so with a portrait (chi-squared test, $p < 0.0001$).

CONCLUSIONS These findings suggest that many medical students complete SETs mindlessly, even when a photograph is included, without careful consideration of whom they are evaluating and much less of how that faculty member performed. This hampers programme quality improvement and may harm the academic advancement of faculty members. We present a framework that suggests a fundamentally different approach to SET that involves students prospectively and proactively.

Medical Education 2015; 49: 928–932
doi: 10.1111/medu.12647

Discuss ideas arising from the article at
www.mededuc.com/discuss.



Center for Educational Development and Research, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA

Correspondence: Sebastian Uijtdehaage, 60-048 Center for the Health Sciences, David Geffen School of Medicine, University of California Los Angeles, Box 951722, Los Angeles, California 90095, USA. Tel: 00 1 310 794 9009; E-mail: bas@mednet.ucla.edu

INTRODUCTION

Student evaluations of teaching (SETs) are, as a rule, required in medical schools in the USA. Commonly, their purpose is two-fold: (i) they provide formative feedback to instructors seeking to improve their teaching, and (ii) they are included in faculty staff dossiers as summative feedback to promotion and tenure committees. A thoughtful and conscientious approach to SETs by medical students is critical because their evaluations may impact both the careers of faculty members and the quality of instruction.

Unfortunately, there is a burgeoning body of research from undergraduate and professional courses at North American campuses that casts serious doubt on the validity of SETs,¹ suggests they are rife with biases^{2,3} and even untruths,^{4,5} and indicates that students may complete evaluations in a mindless manner that further harms the validity of the process.⁶ Defenders of SETs in medical education may dismiss these findings as not likely to apply to our students. After all, aren't medical students carefully selected professionals who are expected to bring the same level of care and attention to course evaluations as they bring to all aspects of their professional training?

A telltale incident at our medical school suggested that SETs by medical students suffer from challenges similar to those elsewhere. In the autumn of 2011, one of the present authors received evaluations for a lecture that he had not taught. Flummoxed (particularly at the averageness of the evaluations), the author went on to discover that a clerical error had inserted his name as an instructor in a multi-instructor course. Multiple students had rated his 'performance' as an instructor, despite the fact that he had not once set foot in the classroom. The incident was caught and corrected, but raised the sticky question of whether medical students are completing SETs mindlessly, without due diligence.

With a simple intervention, our study sought to determine the degree to which medical students complete SETs deliberately rather than doing so in an automatic, mindless manner. Specifically, we examined if and how students would evaluate a fictitious lecturer inserted by ourselves into a list of actual instructors on a course, and whether or not the inclusion of a portrait of that instructor would make a difference. We assumed that if students complete SETs mindfully, they will not choose to evaluate a fictitious lecturer. Furthermore, we

hypothesised that students will be less likely to evaluate a fictitious lecturer when a photograph of the lecturer is provided.

METHODS

We inserted one fictitious lecturer into the evaluation forms for two 8-week, pre-clinical, classroom-style courses (for the Year 2 class of 2010 and the Year 1 class of 2011). We gave these 'lecturers' gender-ambiguous names (e.g. 'Pat Turner', 'Chris Miller') that were distinct from existing names, and added generic lecture titles (e.g. 'Introduction, Lung Disease') (Table 1). Students were required to submit their anonymous ratings of all lecturers, including the fictitious ones, within 2 weeks after the course using our online evaluation system CoursEval[®] (ConnectEDU, Inc., Boston, MA, USA). Students were asked to use a 5-point Likert scale (1 = not effective, 5 = very effective) to respond to the question: 'Please rate the effectiveness of this faculty member's teaching.' Optionally, students were able to add written comments on a lecturer's teaching. Students could choose not to evaluate a lecturer by marking the option 'Not Applicable'. The following year, we repeated this process (in the classes of 2011 and 2012), but also included a small portrait (150 × 150 pixels) of an attractive young model who, perhaps regretfully, did not resemble any of our faculty members. The number of actual lecturers in each course ranged from 23 to 52, most of whom were depicted in portraits of similar dimensions in our evaluation system. Our electronic system reminded students automatically to complete the evaluations (up to seven times, as necessary) as the deadline for evaluations approached without revealing their identity to administrators. Classroom attendance was voluntary and typically around 75%. Lectures (sound and slides) were made available through podcasts.

The University of California Los Angeles Office of Human Research Protection Programme deemed this study exempt from requirements for full institutional review board review as the study was designed not to meet the definition of human subject research as per federal regulations.

RESULTS

Response rates were 99% and 94%, respectively, in the 2 years of the study. Table 1 lists the numbers

Table 1 Numbers of responses from students evaluating fictitious lecturers who were listed with or without a portrait

Fictitious lecturer Title of lecture	Evaluation					Total	N/A	Total responses	Response rate
	Not effective	Less effective	Effective	More effective	Very effective				
Fictitious lecturer without portrait	Class of 2011, evaluations, <i>n</i>								
Pat Turner	0	2	33	35	18	88	55	143	99% (143/145)
Introduction to Infertility									
Fictitious lecturer without portrait	Class of 2010, evaluations, <i>n</i>								
Chris Miller	1	1	28	32	33	95	39	134	99% (134/136)
Introduction, Lung Disease									
Total	1	3	61	67	51	183	94	277	99% (277/281)
Fictitious lecturer with portrait	Class of 2012, evaluations, <i>n</i>								
Pat Turner	0	1	22	19	17	59	81	140	89% (140/158)
Introduction to Infertility									
Fictitious lecturer with portrait	Class of 2011, evaluations, <i>n</i>								
Kim Phillips	1	0	27	32	21	81	64	145	100% (145/145)
Gastrointestinal									
Total	1	1	49	51	38	140	145	285	94% (285/303)

N/A = not applicable (option selected appropriately by students)

of evaluations collected for the fictitious lecturers who were listed with or without a portrait. Using the full range of the Likert scale, 66% (183 of 277 students) evaluated the fictitious lecturer without a portrait; the remaining 34% (94 students) appropriately chose the option 'N/A'. The following year, when the lecturer's name was accompanied by a portrait, fewer students (49%, 140 of 285 students) evaluated the instructor's effectiveness and 51% chose not to. This was a significant drop compared with the previous year ($\chi^2 = 16.50$, $p < 0.0001$).

We found, however, an interaction between medical school class and portrait condition, suggesting that a cohort effect may, in part, explain this drop. The class of 2011 (which participated in both years of the study) showed no difference between the no-portrait and portrait conditions: 88 of 143 students

submitted evaluations when no portrait was provided, and 81 of 145 students did so when a portrait was provided.

A handful of students even went so far as to provide comments on the performance of the fictitious lecturers. Although three students explicitly stated that they did not recall the lectures but wished they had ('I don't think we had this lecture but it would have been useful!'), three other students confabulated: 'She provided a great context'; 'Lectures moved too fast for me', and 'More time for her lectures'.

DISCUSSION

About two-thirds of our sample of medical students evaluated a faculty member whom they

had never seen, and about half did so even when a photograph was included. These findings suggest that many medical students complete SETs mindlessly, without careful consideration of whom they are evaluating and much less of how that faculty member performed. These data add to the growing body of research that questions the validity of SETs and suggest that at least some of that research can be generalised to medical education.

Mindless evaluation is not a modern problem. Our findings echo those described by Reynolds in a landmark 1977 paper.⁷ Like us, he serendipitously found that a vast majority of undergraduate psychology students rated a movie on sexuality higher than a lecture on the history of psychology, although in fact neither event had taken place (due to cancellations). Where our scenario becomes more problematic than that described by Reynolds⁷ is in the unique structure of a medical curriculum, in which a multitude of instructors teach in the same course and are evaluated in bulk by students.

Dunegan and Hrivnak⁶ describe three risk factors that may encourage mindless evaluation practices: (i) the cognitively taxing nature of SETs; (ii) the lack of perceived impact of SETs on the curriculum, and (iii) the degree to which the evaluation task is experienced as just another routine 'chore'. Clearly, all of these risk factors may present themselves in a medical school environment.

With regard to the first risk factor, evaluating teachers is a cognitively demanding task when it is done conscientiously weeks after the fact. Students are required to recall a lecture given by one particular teacher and, from memory, to appraise the effectiveness of that teacher. Unless a teacher is literally 'outstanding' (in either a good or a bad way), this process is fraught with peril as it relies on episodic memory which is subject to rapid decay and depends largely on reconstruction and much less on actual recollection.⁸

The second risk factor that may encourage mindless evaluation by students is the perceived lack of impact of their evaluations. Chen and Hoshower⁹ use expectancy theory to show that students are less motivated to partake in an activity such as evaluation if they fail to see the likelihood that the activity will lead to a desired outcome (e.g. teacher change). This problem is compounded by the fact that medical students are not likely to encounter many lecturers more than once

and therefore will see little, if any, benefit from a positive change promoted by their evaluation.

Lastly, medical students can certainly be forgiven for finding evaluations to be painfully routine and burdensome. As SETs are part of the mainstay of teaching assessment, medical students fill out evaluations numerous times per year during all 4 years of their training. Given that they tend to rely on Likert scales and questions that do not vary among courses or among teachers, the only thing that makes any given evaluation experience unique is the name (and photograph, if included) of the instructor on the form. This routinisation is also a symptom of the one-size-fits-all approach to evaluation that most institutions take. If medical schools wish to break students out of this routine, they must find a way to make the evaluation experience valuable and unique.

Using the framework described by Dunegan and Hrivnak,⁶ we can conceive of an alternative approach to the SET that may mitigate the risk factors described here. For example, at the beginning of a course, a sample of students in a class could be charged to be prospective (not retrospective) course and faculty evaluators. As part of their professionalism training, these students could first be educated in the effective use of evaluation tools that can be employed *in situ* (e.g. with hand-held devices) and that do not rely on the activation of episodic memory. As a team, these students could practise providing constructive feedback when, upon completion of the course, they collaborate on a comprehensive report to the course chair and teachers involved in the course. Evaluation tools could be focused on predictive evaluations (e.g. by asking students to predict their peers' opinions of a teacher) rather than on opinion-based evaluations; predictive evaluations have been shown to require fewer responses to achieve the same result.¹⁰ Faculty members could be required to respond to these reports and explain how the feedback is to be used or not used so that students understand the impact of their efforts. In addition to the educational benefit to be derived from practising teamwork and providing constructive feedback, such an approach may engage students in a mindful way and, importantly, may yield information that provides a more robust foundation for programme improvement and promotion decisions.

Limitations of the study

Our study has some obvious limitations. Firstly, students are never asked to evaluate lecturers they have

not actually seen in lecture unless a clerical error is made. Our use of 'decoy' faculty staff that lured students into artificial evaluations does not represent common custom, and our findings do not necessarily speak of the quality of SETs for real lecturers. Secondly, our findings may generalise only to courses in which many teachers lecture only once or twice. Furthermore, for clinical teachers, reliable and valid evaluation instruments have been developed¹¹ and we suspect that evaluations of clinical teachers with whom students usually establish personal, one-to-one relationships do not suffer from the same pitfalls as SETs in large multi-instructor courses. Clearly not all SETs are mindless; it has been suggested⁶ that SETs are useful for those lecturers who are most in need of help with their teaching.

The present study should raise a red flag to medical schools in which students are asked to evaluate numerous lecturers after a time delay. It defies common sense (and a huge body of literature) to expect that such an evaluation approach procures a solid foundation on which decisions regarding faculty promotions and course improvement can be based. If we continue along this path, we may just as well follow Reynolds's tongue-in-cheek suggestion that 'as students become sufficiently skilled in evaluating [...] lectures without being there, [...] there would be no need [for them] to wait until the end of the semester to fill out evaluations'.⁷

Contributors: SU conceived and executed the research, and conducted the data analysis. CO'N participated in the analysis and interpretation of the data. Both authors drafted major sections of the manuscript, contributed to the critical revision of the paper, approved the final manuscript for submission and have agreed to be accountable for all aspects of the work.

Acknowledgements: none.

Funding: none.

Conflicts of interest: none.

Ethical approval: not required.

REFERENCES

- 1 Nilson LB. Time to raise questions about student evaluations. In: Groccia JE, Cruz L, eds. *To Improve the Academy: Resources for Faculty, Instructional, and Organizational Development*. Vol. 31. San Francisco, CA: Jossey-Bass 2012;213–228.
- 2 Sprinkle JE. Student perceptions of effectiveness: an examination of the influence of student biases. *Coll Student J* 2008;**42** (2):276–93.
- 3 Hamermesh DS, Parker A. Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity. *Econ Educ Rev* 2005;**24** (4):369–76.
- 4 Clayson DE, Haley DA. Are students telling us the truth? A critical look at the student evaluation of teaching. *Market Educ Rev* 2011;**21** (2):101–12.
- 5 Stanfel LE. Measuring the accuracy of student evaluations of teaching. *J Instr Psychol* 1995;**22** (2):117–25.
- 6 Dunegan KJ, Hrivnak MW. Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *J Manag Educ* 2003;**27** (3):280–303.
- 7 Reynolds DV. Students who haven't seen a film on sexuality and communication prefer it to a lecture on the history of psychology they haven't heard: some implications for the university. *Clin Psychol* 1977;**40**:474–80.
- 8 Tulving E. What is episodic memory? *Curr Dir Psychol Sci* 1993;**2** (3):67–70.
- 9 Chen Y, Hoshower LB. Student evaluation of teaching effectiveness: an assessment of student perception and motivation. *Assess Eval High Educ* 2003;**28** (1):71–88.
- 10 Schönrock-Adema J, Lubarsky S, Chalk C, Steinert Y, Cohen-Schotanus J. 'What would my classmates say?' An international study of the prediction-based method of course evaluation. *Med Educ* 2013;**47**:453–62.
- 11 Stalmeijer RE, Dolmans DH, Wolfhagen IH, Muijtjens AM, Scherpbier AJ. 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.

Received 11 July 2014; editorial comments to author 7 August 2014, accepted for publication 21 October 2014

Copyright of Medical Education is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.