

Effects of a teaching evaluation system: a case study

Shi-Hao Wen¹, Jing-Song Xu¹, Jan D. Carline², Fei Zhong¹, Yi-Jun Zhong¹, Sheng-Juan Shen¹

¹ Second Military Medical University, People's Republic of China

² University of Washington School of Medicine, Seattle, USA

Correspondence: Sheng-Juan Shen, Training department, Second Military Medical University, Shanghai, China

Email: smmuwsh@sina.com

Accepted: February 24, 2011

Abstract

Objectives: This study aims to identify the effects of evaluation on teaching and discusses improvements in the work of the evaluation office.

Methods: Teaching evaluation data from 2006 to 2009 was collected and analyzed. Additional surveys were conducted to collect the perceptions of students, faculty members, peer reviewers, deans and chairs about teaching evaluation.

Results: Evaluation scores for more than half of faculty members increased, significantly more for junior compared with senior faculty, over the period of the study. Student attendance and satisfaction with elective courses increased after interventions identified by teaching evaluations. All participants believed that teaching evaluation had positive effects on teaching quality and classroom behavior. Seventy-

three percent of faculty believed the evaluation helped to improve their teaching skills. Faculty perceptions of the helpfulness of teaching evaluation were related to the speed in which evaluations were reported, the quality of comments received, and the attitudes held by faculty towards evaluation. All the faculty members, chairs and deans read evaluation reports, and most of them believed the reports were helpful.

Conclusions: Teaching evaluation at SMMU was perceived to improve both the teaching quality and classroom behavior. Faster feedback and higher quality comments are perceived to provide more help to faculty members.

Keywords: Teaching evaluation, teaching quality, peer review, student rating, feedback

Introduction

In the last two decades, the importance of teaching evaluation has been emphasized in higher education. Many medical schools have searched for ways to effectively and constructively evaluate performances of their faculty members.¹⁻³ The most common sources of evaluation data have been students, peers, and teachers themselves.⁴⁻⁷ Teaching evaluation has been used to provide diagnostic information for teachers on specific aspects of their teaching to help them improve their performance (formative evaluation), and to serve as the basis for decision-making concerning hiring, contract renewal, incentives, and promotions (summative evaluation).^{1, 3-5, 8-10} Although many studies have been published in Chinese journals about teaching evaluation in medical education in China, most of them provide only descriptions of the evaluation system. A few

quantitative studies published focus on the reliability and validity of the evaluation form. There have been few studies on the effects of teaching evaluation.^{6, 7, 11, 12} This paper introduces the teaching evaluation system in Second Military Medical University (SMMU) and describes its effects on the improvement of teaching.

Introduction to teaching evaluation system in SMMU

Teaching evaluation has been conducted in SMMU since 1992. Its purpose has been to provide feedback to faculty members to help them improve their teaching skills, to provide information for promotion and merit, and to identify problems existing in teaching throughout the school. Teaching evaluation services were initially provided by the Education Administration Office.

In 2008 an independent Teaching Evaluation Office was established in the Department of Medical Education to conduct these services. Teaching evaluation units independent of central administration have become popular in China recently, and have contributed increased support to faculty development.^{6,7}

The teaching evaluation system in SMMU depends primarily on the student evaluation of teachers and courses, and peer faculty reviews. Eighty professors, one third of retired faculty and two thirds current faculty, act as reviewers. These professors are selected for their enthusiasm about teaching and extensive experience in teaching and evaluation. About 50 professors from this pool are invited every semester to form a team to observe the classroom sessions of required and elective courses, including lectures, laboratories, and practical skills sessions. Three peer reviewers will evaluate each individual faculty member. The following faculty members must be evaluated each year: new faculty members during the first five year of their appointment, faculty scheduled to be promoted within 2 years, and faculty whose performance was not good in the previous semester. Prior to July 2008, a single session for observation by the team of three peers was scheduled in advance and the faculty member was notified. Starting with the autumn 2008 semester, the observations were not scheduled in advance and each observer was free to select any session they wanted to observe, usually resulting in observation of three sessions.

After observation, the professor and students (20 sampled randomly from the class roster) complete a standard evaluation form to rate faculty performance and provide written comments. If possible, the peer observer provides verbal comments to the teacher immediately after the observation. A feedback report with the ratings and comments is written by the office and provided to the faculty within two weeks after the observation.

At the end of every semester, the office publishes a report including the ratings of each faculty, examples of excellent educational practices identified worth emulating in other courses, and general problems identified with suggestions for solutions. In 2007 we reported that the lesson plans of some faculty were inadequate, the content of some graduate continuing education courses was almost the same as that for undergraduate courses, and some faculty members did not interact well with students. In the following year these issues became a focus of the observation of all faculty.

In the first month of every semester, a summary meeting is held with the vice president of the university, deans of the Educational Administration Office, members of the Teaching Evaluation Office, and about 10 peer observers during which the evaluation results and general problems identified are reviewed. Additionally, teaching and evaluation results are reviewed by the vice president in an annual held meeting attended by the deans and vice deans of all the schools, chairs of departments, and course chairs. In addition

to the reporting the evaluations results, the office surveys students, faculty members, peer observers and some of the deans and chairs about the process of evaluation, generally in July of every year. This study aimed to determine the effect of the evaluation on teaching and improve upon the work of the evaluation office.

Methods

Subjects and Instruments

Teaching Evaluations: Teaching evaluation data from 2006 to 2009 was collected and analyzed. The standard evaluation form includes 10 items both for peer observers and students. A five-point scale is used, with a rating of 10 being best and 2 being worst, with possible ratings of 10, 8, 6, 4 and 2. The overall rating is reported as the sum of these items, with a possible range of 0 to 100. The final evaluation score is averaged across observations, with weightings of 60% for peer ratings and 40% for student ratings.

Survey on the Evaluation System: The teaching evaluation office periodically surveys students, faculty members, peer observers, deans and course chairs about the system of teaching evaluation. Students are selected randomly from all the students on campus. Faculty members surveyed are those who had been observed in the most recent 3 semesters. Peer observers surveyed are those who performed observations in the most recent 3 semesters.

In July 2009, questionnaires were distributed to 450 students, 230 faculty members who were observed in the most recent 3 semesters and 50 peer observers. Response rates were 90.44%, 94.78%, and 84.00% respectively. The questionnaire focused on the effects of teaching evaluation, including teaching quality of required and elective courses, and classroom behaviors in elective courses (student punctuality and attendance, and faculty maintenance of class assigned times). The faculty questionnaire also included items about their attitude towards evaluation, the speed with which feedback is provided (within two weeks, one month, or more than one month after observation), quality of feedback received (fair, good, and very good), and the effects of evaluation reports on their own teaching. The peer observer survey also included asked if they had asked students about problems experienced in the course and if the observer had reported these problems to appropriate administrative officers.

In July 2010, a questionnaire was sent to 500 students and 300 faculty members to survey their perceptions of the teaching evaluation system. The questionnaire included questions regarding the effects of evaluation on teaching quality. Response rates were 89.00% and 79.33% respectively.

In July 2009, twenty-five deans and chairs were visited to ask their perceptions of the system of teaching evaluation. **Elective Courses:** A questionnaire was sent to 1450 students in July 2008 to evaluate elective courses, with 1349

completed (response rate 93.03%); and to 785 students in December 2009, with 702 completed (response rate 89.43%).

Data analysis

Descriptive statistics were used to analyze the evaluation ratings, the effects of evaluation, the satisfaction of students in elective courses, and the attendance of students in elective courses. The Mann-Whitney Test was used to analyze the difference in satisfaction of students for elective courses between in July 2008 and December 2009. The Chi-Square Test was used to compare difference in effects of evaluation among the faculty members, their attitudes towards evaluation, the speed and quality of feedback they received, and improvement in evaluation scores by the faculty members with different professorial titles. All data were analyzed with the SPSS 17 statistical analysis software package.

Results

Evaluation scores: Four hundred and thirty four faculty members were observed for 1672 times from 2006 to 2009. Responses of all scaled items were limited to values of 6, 8 or 10. No rater chose value of 2 or 4 for any rating. One thousand two hundred and forty five observations were planned and 974 (78.23%) were completed during the five semesters from February 2006 to July 2008. Seven hundred and forty eight observations were planned and 698 (93.32%) were completed during the three semesters from September 2008 to December 2009. The mean score for the first five semesters was 93.32 ± 3.85 , and for the last three semesters 88.96 ± 4.73 .

Of 171 faculty members who were observed in more than three semesters from February 2006 to July 2008, 117 received higher ratings in the last semester than in first semester (Table 1).

Table 1. Changes in evaluation scores of faculty from Feb 2006 to Jul 2008 by faculty rank

Evaluation Change		Professorial Titles			Total
		Junior-level	Middle-level	Senior-level	
Evaluations Improved	N	28	55	34	117
	SR*	1.5	.7	-1.7	
Evaluations Did Not Improved	N	3	18	33	54
	SR	-2.2	-1.1	2.6	
Total		31	73	67	171

* Standardized residual

The improvements in evaluation differed by professorial title ($\chi^2=18.19$, $p < 0.01$). More junior-lever faculty members saw improvements than expected, and fewer senior-level faculty members than expected saw improvements as indicated by standard residual analyses. The mean score of these junior-lever faculty members for the first semester was 90.81 ± 2.82 , and for the last semester 94.52 ± 2.31 . The mean score of these middle-lever faculty members for the first

semester was 91.72 ± 3.02 , and for the last semester 94.81 ± 2.53 . The mean score of these senior-lever faculty members for the first semester was 91.84 ± 2.52 , and for the last semester 95.14 ± 1.89 .

Of 76 faculty members who were observed in all three semesters from September 2008 to December 2009, 48 received higher ratings in last semester than first semester (Table 2). No junior level faculty were observed in all the three semesters. The proportion of faculty who received increased ratings was greater for middle-level titles than for senior-level titles; a statistically significant difference ($\chi^2=14.35$, $p < 0.01$). The mean score of middle-lever faculty members for the first semester was 87.92 ± 5.31 , and for the last semester 90.33 ± 4.72 . The mean score of these senior-lever faculty for the first semester was 90.38 ± 3.23 , and for the last semester 91.39 ± 4.41 .

Table 2. Changes in evaluation scores of faculty from Sep 2008 to Dec 2009 by faculty rank

Evaluation Change		Professorial Titles		Total
		Middle-level	Senior-level	
Evaluations improved	N	38	10	48
	SR*	1.4	-1.8	
Evaluations did not improved	N	10	18	28
	SR	-1.8	2.4	
Total		48	28	76

* Standardized residual

Teaching problems: The peer observers were asked to pay more attention to three problems since spring semester in 2008: lesson plans, appropriateness of teaching content by levels of courses or programs, and interaction between teacher and students. The mean score of the item of lesson plans was 9.08 ± 1.03 in autumn semester 2008, 9.59 ± 0.81 in autumn semester 2009, a significant difference between these two years ($p < 0.01$), effect size 0.54. The mean score of the item of teaching contents suited to different programs was 9.09 ± 1.06 in autumn semester 2008, 9.43 ± 0.95 in autumn semester 2009, a significant difference ($p < 0.01$), effect size 0.34. Similarly, the mean score of the item of interaction between teacher and students was 8.36 ± 1.11 in autumn semester 2008, 8.71 ± 1.08 in autumn semester 2009, a significant difference ($p < 0.01$), effect size 0.32. All three evaluation items received at least moderately higher evaluations after attention was explicitly paid to them by evaluators.

Analyses were completed for faculty members who had received scores in the bottom third of the ratings in 2008. Twenty-eight who received relatively low scores on the lesson plans item in 2008 were observed 37 times in autumn semester 2008 and 75 times in autumn semester 2009. Their scores improved from 7.89 ± 0.46 in 2008 to 9.68 ± 0.73 in 2009, a significant difference between these two years ($p < 0.01$), with an effect size 2.74.

Twenty-nine faculty who received relatively low scores on the appropriate content item were observed 65 times in

autumn semester 2008 and 66 times in autumn semester 2009. Their scores improved from 8.74 ± 1.09 in 2008 to 9.55 ± 0.84 in 2009, a significant difference between these two years ($p < 0.01$), effect size 0.83.

Thirty-two faculty who received relatively low scores on the faculty/student interaction item were observed 55 times in autumn semester 2008 and 79 times in autumn semester 2009. Their scores improved from 7.67 ± 0.75 in 2008 to 8.58 ± 1.12 in 2009, a significant difference between these two years ($p < 0.01$), effect size 0.92. All faculty who performed in the lower third of evaluations showed large improvements on these three items.

Attendance and satisfaction of students in elective courses: The attendance rate in elective courses in spring semester 2008 was 67.21%, in contrast to 87.56% in autumn semester 2009. Levels of student satisfaction in elective courses increased significantly between Spring 2008 and Autumn 2009 semesters ($p < 0.01$) (Table 3).

Table 3. Level of satisfaction of students in elective courses by year

Year	High n (%)	Moderate n (%)	Low n (%)	Total n (%)
2008	216 (16)	796 (59)	337 (25)	1349 (100)
2009	260 (37)	358 (51)	84 (12)	702 (100)

Effects of evaluation on teaching: In both July 2009 and July 2010 surveys, students, peer reviewers and faculty members believed that evaluation had an effect on improvement of teaching (Table 4 and 5). Feedback delivered quickly, within two weeks after observation, was of greater help to faculty than delayed feedback ($\chi^2=57.40$, $P < 0.01$). The faculty believed that they received more help from more detailed and higher quality comments, ($\chi^2=63.77$, $P < 0.01$). Faculty who were open to evaluation believed they got more help from evaluation than faculty less interested in being evaluated ($\chi^2=57.40$, $P < 0.01$).

All 218 faculty members reported reading the evaluation report every semester and 85.58% stated that it was helpful. All 25 deans and chairs read the report every semester and believed that it was helpful. They said they learned from the examples of other departments listing in the reports, and utilized solutions for problems in their own departments. Comments and suggestions for the teaching evaluation system: Thirty two of the 42 peer reviewers said they often asked students about problems they experienced in their education and reported these problems to appropriate administrative officers. Ten observers did not actively ask students about problems, but did report these issues if informed. All peer reviewers welcomed being able to select the specific class to observe. Most faculty members (62.98%) believed that they would spend more time to prepare for every class if they were not informed when they were to be observed, and 52.40% believed that it added to their work burden.

Table 4. Level of perceived effect of evaluation on teaching by students, peers and faculty members

Year	Item	Effect	Student n (%)	Peer n (%)	Faculty n (%)
2009	Required courses	High	217 (53)	23 (55)	
		Moderate	163 (40)	19 (45)	
		Low	27 (7)	0 (0)	
	Quality of elective courses	High	152 (37)	17 (40)	
		Moderate	205 (50)	19 (45)	
		Low	50 (12)	6 (14)	
	Class-room behaviors in elective courses	High	152 (37)	14 (33)	96 (44)
		Moderate	199 (49)	23 (55)	107 (49)
		Low	56 (14)	5 (12)	15 (7)
Positive effect on teaching skill	High			160 (73)	
	Moderate			49 (22)	
	Low			9 (4)	
2010	Teaching	High	264 (59)		
		Moderate	152 (34)		
		Low	29 (7)		
	Positive effect on teaching skill	High			159 (67)
		Moderate			54 (23)
		Low			25 (11)

Discussion

The system of teaching evaluation contributed to the perceived improvement of teaching quality and faculty's teaching skills at SMMU. The faculty members received affirmations of their strengths and helpful suggestions for improvement.^{13, 14} Performance ratings increased both for when faculty had been informed which session would be observed and when they were not informed. More faculty with lower professorial title received increased score, perhaps because they had less experience in teaching and were less proficient in teaching content.¹⁵ They felt they could improve even more if they received detailed and effective comments from peer observers and students. Though there were many other factors that might have improved the faculty's teaching skills, such as faculty development program, help from colleges and self study, evaluation were perceived to have contributed to their improvement. The faculty with higher title did not receive significantly higher rating after initial evaluation, suggested that this system teaching evaluation was more help to newer faculty members.

We noticed that the scores for lesson plans, appropriateness of content to level of the course and program, and interaction between teacher and students were low in 2007. Some lesson plans were too simple or did not meet the standard syllabus requirements. Some content in graduate level continuing education courses was found to be at the level undergraduate courses, which dissatisfied students in those courses. Some faculty did not interact well with students in class. We asked the peer reviewers to pay more attention to these aspects and provide more comments and suggestions for change. Subsequently, ratings of these three

aspects improved. This suggested that teaching evaluation could identify common problems existing in teaching and help provide solutions for them. Most of peer reviewers asked students about problems they experienced in their education and reported these problems to appropriate administrative officers, particularly when students were reluctant to report the problem or when they did not know who to report the issues.

Table 5. Benefit from evaluation by speed at which it is delivered, the quality of comments provided, and their attitude towards evaluation

		High	Middle	Low	Total
Time feedback was received	Within two weeks	100	25	2	127
	Within one month	50	28	2	80
	More than one month	3	3	5	11
	Total	153	56	9	218
Perceived quality of comments	Very good	98	20	1	119
	Good	53	26	2	81
	Fair	2	10	6	18
	Total	153	56	9	218
Attitude of openness to evaluation	Most	130	38	3	171
	Middle	17	8	2	27
	Least	6	2	4	20
	Total	153	27	9	218

Students, faculty, professors, deans and chairs all believed that evaluation can improve the teaching quality, for both required and elective course. The elective courses, given in the evening in our university, traditionally receive less attention from the Education Administration Office, deans, chairs and faculty. The student satisfaction rating in elective courses increased after these courses were observed, and faculty members stated that instruction benefited from observation. Classroom behaviors improved and increased attendance rate of students also supported this finding. After the elective courses were observed, the teachers were stricter about the university attendance policies.

It was important to provide feedback to faculty as fast as possible. The faculty members who received feedback earlier believed that they got more help from it, and remembered more specific details about that session. They could more easily use the comments to affirm their strengths and improve weaknesses.^{7, 12, 17} Quick reporting of observation results requires efficiency in the Teaching Evaluation Office. Faculty members believed that specific and detailed comments gave them more help. Peer reviewers and students should be prompted to provide detailed comments in the evaluation. All the faculty members, chairs and deans read the summary reports of teaching evaluation and found them helpful as they described strengths and problems in teaching, and successful examples of teaching strategies. Some of these examples have led to major changes in departments. Open discussions and meetings about evaluation with senior administrators emphasize both the

importance of good teaching and the role of evaluation in improvement.

Sustaining a viable peer review system is time-consuming and resource-intensive. Some professors were reluctant to act as peer reviewers, because they were overwhelmed with existing job responsibilities.^{8, 14} We invited retired professors to observe and allowed them the flexibility to choose sessions to observe. Most peer reviewers welcomed this change. The faculty members did not know which session would be observed under this change, however more than half of them indicated a positive response to this change. The faculty members who had positive attitudes towards teaching evaluation believed that they received more help from evaluation than those with less positive attitudes. We should help faculty members understand that evaluation is primarily aimed at course and teaching improvement. While evaluation is done to provide information for decision making and merit, the more important function is to give feedback to faculty to help their improvement. The teaching evaluation system in SMMU worked, in part, because the peer reviewers were experienced and respected by their peers and they took their work seriously. Additionally, the presidents, deans, chairs and faculty members supported and respected evaluation, and the Teaching Evaluation Office improved its work continually.^{16, 18}

Limitation of the Study

The study is limited that it was retrospective. The survey data was from the teaching evaluation office and the questionnaires were different, preventing direct comparisons. This may lessen the strength of our conclusion.

Conclusion

The teaching evaluation system in SMMU is successful. The system of teaching evaluation helped improve teaching quality and classroom behaviors. While information from the systems was used to provide information for decision-making, it's more important functions were in faculty development and educational improvement. Quicker feedback and more detailed comments increase the value of evaluations to faculty. The support and respect of evaluation by presidents, deans, chairs and faculty members, as well as continual efforts to improve the evaluation system add to its success.

Acknowledgments

This study was conducted at the Department of Medical Education, Training Department, Second Military Medical University, Shanghai, People's Republic of China.

References

1. Gimbel RW, Cruess DF, Schor K, Hooper TI, Barbour GL. Faculty performance evaluation in accredited U.S.

- public health graduate schools and programs: a national study. *Acad Med.* 2008;83:962–968.
2. Rosenbaum ME, Ferguson KJ, Kreiter CD, Johnson CA. Using a peer evaluation system to assess faculty performance and competence. *Fam Med.* 2005;37:429-33.
 3. Bland CJ, Wersal L, VanLoy W, Jacott W. Evaluating faculty performance: a systematically designed and assessed approach. *Acad Med.* 2002;77:15–30.
 4. Paulsen MB. Evaluating teaching performance. *New Directions for Institutional Research.* 2002;114:5-18.
 5. Valle R, Alaminos I, Contreras E, Salas L E, Tomasini P, Varela M. Student questionnaire to evaluate basic medical science teaching. *Rev Med IMSS.* 2004;42:405-411.
 6. Wen SH, Jinag XB, Yang H. Problems in evaluating teaching quality in colleges and universities and related measures to deal with the problems. *Researches in medical education.* 2003;4:17-19.
 7. Li P, Wang ZW, Wang ZQ, Wang Y. Study on the situation and tendency of teaching evaluation in university in China. *Northwest Medical Education.* 2009;2:211-212.
 8. Jahangiri L, Mucciolo TW, Choi M, Spielman AI. Assessment of teaching effectiveness in U.S. dental schools and the value of triangulation. *J Den Educ.* 2008;6:707-718.
 9. Williams BC, Pillsbury MS, Stern DS, Grum CM. Comparison of resident and medical student evaluation of faculty teaching. *Eval Health Prof.* 2001;24:53-60.
 10. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med.* 2000;75:161–166.
 11. Fu LQ, Li JC, Zhang XM. Situation and countermeasures of the educational inspection for colleges and universities. *Researches in Medical Education.* 2008;11: 1136-1137.
 12. Wen SH, Jiang XB, Jiang P. Study on the problems in the teaching evaluation system and feedback in university. *Northwest Medical Education.* 2004;1:20-21.
 13. Banda SS. Student evaluation of teaching: a case study from School of Medicine (UNZA). *Medical Journal of Zambia.* 2008;35:94-99.
 14. Irby DM. Peer review of teaching in medicine. *Journal of Medical Education.* 1983;58:457-461.
 15. Liu TB. Study on the improvement of teaching skills of young faculty. *Journal of Qiqihar Medical College.* 2007;10:1229-1230.
 16. Horowitz S, Eyck SV, Albanese M. Successful peer review of courses: a case study. *Acad Med.* 1998;73:226-271.
 17. Daberkow DW, Hilton C, Sanders CV, Chauvi SW. Faculty evaluations by medicine residents using known versus anonymous systems. *Med Educ Online.* 2005;10:12.1-9.
 18. Kogana JR, Sheab JA. Course evaluation in medical education. *Teaching and Teacher Education.* 2007;23:251–264.