

The Utility of Student Evaluations for Medical Sciences Teachers and Administrators

Seyedeh Azadeh Safavi

Ph.D in Curriculum & Instruction
Educational Development Centre (EDC)
Hamedan University of Medical Sciences
6517619653, Hamedan, Iran.

Kamariah Abu Bakar

Professor of Science Education
Institute for Mathematical Research
Universiti Putra Malaysia (UPM)
43400 Selangor, Serdang, Malaysia.

Abstract

Since its introduction 93 years ago, student evaluation of teaching has swept the higher education around the world. The literature is replete with multisection studies, multitrait- multimethod validity studies, bias studies, laboratory validity studies, and dimensionality studies, all of which have identified some common themes; however, little research evidence is available on the efficacy of student evaluations and its consequences. Typically, student evaluations are utilized for formative and summative purposes. Given the mandatory nature of student evaluations in some medical sciences institutions, it is essential to establish that this process is, indeed, effective. The research reported in this paper explores the applicability and efficacy of student evaluations in medical sciences faculties and takes the additional step of delineating influenced aspects of teaching/administrative practices. The population comprised 320 administrators and lecturers from four medical sciences faculties at one of the leading research universities in Malaysia. The data collected were analyzed using descriptive and inferential statistics. The findings uncovered respondents' responsiveness to student evaluations, identified the perceived effect of student evaluations on different aspects of administrative practices, and explored specific aspect of teaching practices that medical sciences lecturers modified in response to information from student evaluations.

Keywords: student evaluations of teaching; medical sciences administrators, medical sciences lecturers

1. Introduction

Student evaluations of teaching (SET) have been in existence since the 1920s (d'Apollonia & Abrami, 1997). The continued use of SET in higher education for teaching or administrative purposes has been based on empirical research indicating that a well-constructed instrument can be considered a useful measure of teaching effectiveness (Penny, 2003; d'Apollonia & Abrami, 1997). Also, results from SET should not be interpreted as the sole measure of teaching effectiveness. In other words, SET in isolation will not necessarily lead to teaching improvement (Seldin, 1993).

In higher education institutions, student evaluations influence some administrative decisions and improve teaching practices (Onwuegbuzie et al., 2009; Murray, 2005; Kulik, 2001). A paucity of research about the influenced practices by SET limits what claims can be made about the value of such information for both formative and summative purposes. Indeed, without knowing if administrators/lecturers actually take advantage of the presumed benefits of SET, it is difficult to judge the effectiveness of SET for the intended purposes. Gall (2004) called for further research for determining the nature of changes that SET creates in teaching practices.

Although previous studies generally found that student evaluations lead to teaching improvement and have impact on personnel decisions (Murray, 2005, 1997), influenced aspects of teaching/administrative practices have rarely been researched, specifically in medical sciences institutions.

As such, the purpose of the present study was to gain a broader understanding of the applicability and efficacy of SET in medical sciences faculties by identifying influenced aspects of administrative and teaching practices as perceived by the Medical Sciences Administrators (MSAs) and the Medical sciences lecturers (MSLs), respectively.

2. Method

2.1. Population and Sampling

The present study was conducted at a public research university in Malaysia. The population consisted of 320 academic staff from four medical sciences faculties including Medicine and Health Science faculty, Veterinary Medicine faculty, Food Science - Technology faculty, and Biotechnology- Bimolecular Sciences faculty. The population comprised medical sciences lecturers with the academic ranks of professor, associate professor, senior lecturer, and lecturer as well as the medical sciences administrators who had more than 6 months of administrative experience. The sample size was determined at 212 academic staff based on the Cochran's sample size formula (1977) for continuous data. The sample drawn from each faculty was based on stratified proportionate random sampling. Overall, 97 questionnaires were returned resulting in a return rate of 53.8% for the MSAs and 43.1% for the MSLs. All of the returned questionnaires were used in the analysis.

2.2. Design and Instrumentation

This study employed two self-developed questionnaires, including the Medical Sciences Administrators' Questionnaire (MSAQ) (21 items) and the Medical Sciences Lecturers' Questionnaire (MSLQ) (25 items). The MSAQ and MSLQ were separately used to collect quantitative data from the administrator participants and the lecturer participants, respectively. The items of the questionnaires were developed based on the related literature in three parts, including the respondents' demographic characteristics (Part A), the efficacy of student evaluations (Part B), and the respondents' reactions to student evaluations (Part C). In part B, the respondents were asked to report how often the SET had influenced their administrative/teaching practices using a 4-point Likert scale that ranged from 'Never' to 'Always'. Also, the respondents' reactions were presented by their levels of agreement with the items in part C of the questionnaires, using a four-point Likert scale that ranged from 'Strongly Disagree' to 'Strongly Agree'.

In order to examine content validity of the instruments, a panel of five experts who had professional knowledge and strong background in design or usage of teaching evaluation tools was consulted. The comments from the panel resulted in modification of some items. Additionally, two separate pilot studies were carried out with 50 administrators as well as 50 lecturers (who were not involved in the actual research) using MSAQ and MSLQ, respectively. The results were used to assess the appropriateness of the operational definitions and to determine the degree of construct validity. To further verify strengths of the results, reliability of the questionnaires was investigated using the measures of internal consistency.

3. Finding and Discussion

3.1. Respondents' Demographic Characteristics

Two different groups of respondents participated in the present study including medical sciences administrators (MSA) and medical sciences lecturers (MSL). The MSAs' demographic characteristics included information about their gender, administrative rank, administrative experience, and their experience in using SET for summative purposes. The MSLs' demographic characteristics focused on their gender, academic rank, level of courses they taught, and experiences in medical sciences education.

The majority of MSAs were male while more than half of the MSLs were female. From four medical sciences faculties, almost three quarters of the MSAs were department heads (71.4%, n=15), and nearly one third of them were deputy deans and deans (28.5%, n=6). The MSAs' experiences in administrative work ranged from 3 to 24 years with a mean of about 7 years. On average, the MSAs who participated in the study had about 5 years of experience in using SET for administrative purposes.

Almost three quarters of the MSLs were lecturers and senior lecturers (71.05%, n=54), and more than one quarter of them were associate professors and professors (28.94%, n=22). Over half of the MSLs taught in both postgraduate and undergraduate levels (59.2%, n=45). The MSLs also had considerable experience (a mean of 10 years) in medical sciences education.

Hence, they were well qualified to identify the efficacy of SET for teaching improvement.

3.2. Response to SET in medical sciences faculties

The MSA and MSL respondents provided generally positive responses with the majority disagreeing with statements such as ‘Teaching performance should not be evaluated by students’, ‘I don’t take significant results from student evaluations of teaching’, and ‘The student evaluation of teaching is a meaningless activity’(Tables1&2). Some scholars argue that students can be considered the ideal source for evaluating teaching effectiveness since they are in a good position to observe lecturers directly and provide first hand information for the users (Wallace & Wallace, 1998). As can be seen in Tables 1 and 2, the majority of respondents disagreed with the statement ‘Teaching performance should not be evaluated by students’. This result is supported by several studies on the subject of students’ competency as evaluators of teaching effectiveness such as Nasser and Fresko (2002), Ory and Ryan (2001), Wachtel (1998), Feldman (1997), Wagenaar (1995), and Scriven (1995). These studies demonstrated that students can rate several aspects of teaching effectiveness. Review of the literature indicates that there is similarity between the attitudes expressed by the medical sciences respondents in this study and those described in previous studies. Beran et al. (2005) and Schmelkin et al. (1997) found that a majority of lecturers had generally positive views of student evaluations. For instance, 84% of the respondents in Schemelkin and colleagues’ study disagreed that lecturers should not be evaluated by students. In summary, the results from Table 1 and 2 revealed that the majority of MSAs and MSLs who participated in this study had positive reactions to student evaluations. Similarly, review of the literature shows that higher education administrators and lecturers possess generally positive reactions to student evaluations and regard results from student evaluations as useful information (Campbell & Bozeman, 2008; Beran et al., 2005).

A total of 71.1% (n=54) of the MSL respondents and a total of 76.2% (n=16) of the MSA respondents concurred that ‘the university uses student evaluations as the only measure of teaching effectiveness’. All of the MSAs who participated in this study strongly agreed or agreed that for improving judgments about teaching effectiveness other methods of teaching evaluation should be used to supplement SET information. Although results from SET can influence teaching practices, the caution is that it should not be interpreted as the sole measure of teaching effectiveness in universities. The body of literature in this area indicates that the contribution of student feedback to teaching improvement or personnel decisions is greatly augmented if they are accompanied by other methods of teaching evaluation such as peer observation and teaching portfolios (Murray, 2005; Algozzine et al., 2004).

In order to improve validity of SET, it is crucial to make sure that they are applied correctly and consistently. According to Gravestock and Gregor-Greenleaf (2008), “many threats to validity are introduced through inconsistent administration of [SET] (p.54).”In this study, almost three quarters of the MSLs (total of 72.4%, n=55) strongly agreed or agreed that administration of the current evaluation process needs revision. Hence, this finding recommends developing targeted interventions aimed at assessing the administration of the current evaluation procedure in order to address the revisions that need to be made.

3.3. The efficacy of student evaluations: Influenced practices

Each group of the respondents completed their relevant questionnaire that asked them to consider the influence of SET on the specified practices. The closed-ended questions included, for example, ‘Based on your administrative experiences, please indicate the influence of student feedback provided by the university on personnel decisions regarding reappointment’ or ‘Based on your experience in medical sciences education, please indicate the influence of student feedback on improving your lectures’. These questions were presented on a range of four-point Likert scale from ‘Never’ to ‘Always’. Table 3 shows the influenced aspects of theoretical teaching practices in medical sciences education as perceived by the MSLs. Also, Table 4 presents the influenced aspects of administrative practices in medical sciences faculties as perceived by the MSAs.

The results indicated that the SET influenced two aspects of teaching practices, namely ‘delivery’ (M= 2.94) and ‘professionalism’ (M= 2.68), while two aspects of teaching practices namely ‘student assessment’ (M=2.42) and ‘lesson planning’ (M= 2.25) were not influenced by SET. The effect of SET on ‘delivery’ aspect of theoretical teaching in medical sciences education was presented on selecting teaching methods, selecting presentation methods, lecturing, communicative skills, and clarity. Also, the ‘professionalism’ aspect of theoretical teaching in medical sciences education was presented by respect for student, interaction with students, and lecturers’ availability for help and consultation.

These results are supported with the idea of distinguished scholars that students can accurately rate effective teaching in several aspects (e.g. Theall & Franklin, 2001; d'Apollonia, 1997; Greenwald, 1997; and Mckeachie, 1997). For instance, Theall and Franklin (2001) argued that students accurately judge the teaching aspects such as interaction with students, professional behavior, and delivery (methods and skills). From the MSLs' perspectives, SET had no influence on an aspect of their teaching practices namely 'student assessment'. This aspect consisted of the practices such as selecting course text books, selecting course objectives, and selecting exam content. This finding corresponds with the finding of the study conducted by Beran and her colleagues in 2005. They found that results from SET were least often used by the instructors to make decisions about course textbooks, exams, and course assignments.

Whilst the MSA respondents reported that the SET influenced administrative decisions related to the 'annual reviews' of academic staff ($M=3.17$), two aspects of their practices, namely 'quality evaluation' ($M=2.46$) and 'educational planning' ($M= 1.57$) were not influenced by the SET (Table 4). The aspect of 'annual reviews' consisted of the personnel decisions regarding tenure, promotion, and reappointment. The aspect of 'quality evaluation' comprised the administrative practices such as documenting teaching quality, monitoring improvements in a specific course, assessing course quality, comparing teaching quality of departments, and monitoring teaching improvements in faculties/departments. Also, the 'educational planning' aspect of their administrative work reflected the practices pertaining to teaching development and course schedules development.

There are parallels between the MSAs' perspectives in this study and the administrators' opinion in other research studies. For instance, in a Canadian university Beran et al. (2005) found that SET is useful for making personnel decisions, monitoring progress, and evaluating teaching. Also, the findings of the present study regarding the influenced administrative practices is further supported by the opinion expressed by Murray (2005), Algozzine et al. (2004), Penny (2003), and Abrami (2001). These scholars generally accepted that SET is a major input for making the personnel decisions such as employment and promotion.

The results of this study revealed that the medical sciences administrators as well as the medical sciences lecturers possessed positive reactions toward student evaluations overall. In addition, to advance our understanding about the efficacy of student evaluations in medical sciences faculties, we asked the respondents to determine the effect of SET on their specified practices. The literature is replete with multisection studies, multitrait- multimethod validity studies, bias studies, laboratory validity studies, and dimensionality studies, all of which have identified some common themes; however, little research evidence has been devoted to the efficacy of SET and its consequences. Ory and Ryan (2001) argue that "to improve the validity of our student ratings, we need to conduct research on their use and consequences" (p.40). Although previous researchers found that student evaluations are useful for administrative purposes and teaching improvement, research has not precisely specified influenced practices (Wachtel, 1998; Stratton & Myers, 1994; Spencer & Flyr, 1992). Hence, the results of the present study corroborate the common conclusion in the literature by defining the influenced aspects of administrative and teaching practices in the medical sciences faculties.

4. Conclusion

In medical sciences faculties, students are requested to rate theoretical teaching quality of their lecturers. The research reported in this paper explored the efficacy of student evaluations of theoretical teaching in the medical sciences faculties and defined the influenced aspects of teaching/ administrative practices in these faculties. More broadly, the effectiveness of SET needs to be explored in greater depth to include clinical/laboratory aspect of teaching in medical sciences institutions. Further studies are still needed to explore the applicability of student evaluations for assessing clinical /laboratory aspects of teaching quality in medical sciences education.

References

- Abrami, P. C. (2001). Influencing judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research*, 109, 59–87.
- Algozinne, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Mohanty, G., & Spooner, F. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching*, 52 (4), 134-141.
- Beran, T., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty, and administrators: A consequential validity study. *The Canadian Journal of Higher Education*, XXXV(2), 49-70.
- Campbell, J.P., & Bozeman, W.C. (2008). The value of student ratings: Perceptions of students, teachers and administrators. *Community College Journal of Research and Practice*, 32(1), 13-24.
- Cochran, W.G. (1977). *Sampling techniques*. New York: John Wiley and Sons Publisher.
- d' Apollonia, S. (1997). *The dimensionality and validity of student ratings of instruction: Two meta-analyses*. Unpublished doctoral dissertation, Concordia University, Canada.
- d' Apollonia, S & Abrami, P.C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208.
- Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In *Effective Teaching in Higher Education: Research and Practice* (pp.241-320). New York: Agathon Press.
- Gall, A. R. (2004). Faculty perceptions of the effects of student evaluations of teaching on higher education instructional practices and instructor morale. Doctoral dissertation, Marshall University, Huntington, USA.
- Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student course evaluations: research, models and trends*. Toronto: Higher Education Quality Council of Ontario.
- Greenwald, A.G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychological Association*, 52(11), 1182-1186.
- Kulik, J.A. (2001). Student ratings: validity, utility, and controversy. *New Directions for Institutional Research*, 109, 9–25. doi: 10.1002/ir.1
- Mckeachie, W.J. (1997). Student ratings: The validity of use. *American psychologist*, 52(11), 1218-1225.
- Murray, H. G. (2005). Student evaluation of teaching: Has it made a difference? Paper presented at the Annual Meeting of the Society for Teaching and Learning in Higher Education, Charlottetown. Prince Edward Island
- Murray, H. G. (1997). Does evaluation of teaching lead to improvement of teaching? *International Journal of Academic Development*, 2(1), 8–2.
- Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, 27(2), 187-198.
- Onwuegbuzie, A.J., Daniel, L.G., & Collins, K.M.T. (2009). A meta-validation model for assessing the score-validity of student teaching evaluations. *Quality & Quantity*, 43(2), 197-209.
- Ory, J.C., & Ryan, K.E. (2001). How do student ratings measure up to a new validity framework? *New directions for institutional Research*, 27(5), 27-44.
- Penny, A.R. (2003). Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education*, 8(3), 399-411.
- Scriven, M. (1995). Student ratings offer useful input to teacher evaluations. *Practical Assessment, Research & Evaluation*, 4(7), 1-4.
- Schmelkin, L.P., Spencer, K.J., & Gillman, E.S. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education*, 38, 575–592.
- Seldin, P. (1993). The Use and Abuse of Student Ratings of Professors. *The Chronicle of Higher Education*, 39(46), A40.
- Spencer, P. A., & Flyr, M. L. (1992). The formal evaluation as an impetus to classroom change: myth or reality? (report no. JC 920 441). Riverside: Riverside Community College.
- Stratton, R. W., & Myers, S. C. (1994). Faculty behavior, grades, and student evaluations. *Journal of Economic Education*, 25(1), 5–16.
- Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction. *New directions for institutional research*, 109, 45-56. doi: 10.1002/ir.3
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: a brief review. *Assessment & Evaluation in Higher Education*, 29(2), 191–121.
- Wagenaar, T.C. (1995). Student evaluation of teaching: Some cautions and suggestions. *Teaching Sociology*, 23 (1), 64-68.
- Wallace, J. J., & Wallace, W. A. (1998). Why the costs of student evaluations have long since exceeded their value. *Issues in Accounting Education*, 13(2), 443-448.

Table 1: Medical Sciences Administrators' Reactions toward SET

Items	Mean	Percentage of Agreement			
		SD	D	A	SA
Teaching performance should not be evaluated by student.	2.52	.00	57.1	4.8	38.1
I don't take student evaluations of teaching seriously.	2.85	9.5	71.4	14.3	4.8
The student evaluation of teaching is a meaningless activity.	2.71	14.3	47.6	33.3	4.8
The university uses student feedback as the only measure of teaching effectiveness.	2.76	9.5	14.3	66.7	9.5
To supplement feedback information, other methods of teaching evaluation should be utilized.	3.47	.00	.00	52.4	47.6
By this time, I have not used the feedback information for my administrative purposes.	2.71	4.8	66.7	23.8	4.8

Table 2: Medical sciences lecturers' Reactions toward SET

Items	Mean	Percentage of Agreement			
		SD	D	A	SA
Teaching performance should not be evaluated by students.	2.73	22.4	42.1	22.4	13.2
I don't take student evaluations of teaching seriously.	2.78	22.4	40.8	30.3	6.6
The student evaluation of teaching is a meaningless activity.	2.82	57.9	14.5	3.9	23.7
The university uses student evaluations as the only measure of teaching effectiveness.	2.80	14.5	14.5	47.4	23.7
Administration of teaching assessment process needs revision.	2.92	6.6	21.1	46.1	26.3

Table 3: Influenced Aspects of Theoretical Teaching Practices in Medical Sciences Education

Teaching practices	*Influenced aspect			
	Delivery	Professionalism	Student assessment	Lesson Planning
Preparations for teaching	.75			
Selection of teaching methods	.59			
Selection of presentation methods	.70			
lectures	.88			
Communicative skills	.85			
Clarity	.84			
Selection of course textbooks				.59
Setting course objectives				.77
Refinement of teaching objectives				.58
Exam modification			.86	
Selection of assessment method			.88	
Selection of exam content			.89	
Planning assignments			.69	
Attitude toward students		.91		
Respect for students		.88		
Interaction with students		.88		
Availability for help/consultation		.68		
Eigen value	8.05	2.60	1.84	1.03
% of variance	47.40	15.32	10.82	6.10
Coefficient alpha	.92	.92	.89	.82
Inter item correlation	.66	.71	.68	.70

Perceived influenced aspects' mean score: (Less through 2.5) had no influence on teaching practices, (2.51 through highest) had influence on teaching practices.

Table 4: Influenced Aspects of Administrative Practices in Medical Sciences Faculties

Administrative practices	Influenced aspects		
	Annual Reviews	Quality Evaluation	Educational Planning
Personnel decisions regarding tenure.	.94		
Personnel decisions regarding promotion.	.97		
Personnel decisions regarding reappointment.	.93		
Documenting teaching quality.		.73	
Monitoring improvements in a specific course.		.92	
Assessing course quality.		.83	
Comparing teaching quality of departments.		.48	
Monitoring teaching improvements in faculties/departments.		.86	
Teaching schedule development.			.95
Course schedule development.			.95
Eigen value	4.34	2.99	1.81
% of variance	39.51	27.20	16.46
Coefficient alpha	.96	.84	1.00
Inter item correlation	.89	.52	1.00

Perceived influenced aspects' mean score: (Less through 2.5) had no influence on administrative practices, (2.51 through highest) had influence on administrative practices.