The Role of Student Ratings of Instruction from Perspectives of the Higher Education Administrators

Seyedeh Azadeh Safavi
PhD Candidate of Curriculum & Instruction
Faculty of Educational Studies
Universiti Putra Malaysia (UPM)
43400 Selangor, Serdang
Malaysia.

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

Rohani Ahmad Tarmizi
Associate Professor of Mathematics Education
Institute for Mathematical Research
Universiti Putra Malaysia (UPM)
43400 Selangor, Serdang
Malaysia.

Nor Hayati Alwi
Senior Lecturer of Curriculum
Faculty of Educational Studies
Universiti Putra Malaysia (UPM)
43400 Selangor, Serdang
Malaysia.

Abstract

Student ratings of higher education instructors are the most common evaluative procedure in universities around the world. Nevertheless, there are certain aspects of student ratings that require further research. This study investigated the role that student ratings actually played in administrative practices regarding instructional performance of academic staff within the context of a major research university. The population comprised administrators from 15 faculties which had a total enrolment of 25,628 students. The responses were subjected to descriptive statistics, multiple group analysis, one way ANOVA, and independent t-test. The results revealed that student ratings played an important role in administrative decision-making with regard to the instructional performance of academic staff. The findings also took cognisance of the respondents' recommendations such as supplementing the ratings information with other measures of instructional quality, providing opportunities for less effective instructors to receive instructional consultation services, and providing an array of items so that different faculties could select items suitable for their needs.

Keywords: student ratings of instruction; administrative practices; higher education administrators

1. Introduction

Student ratings of higher education instructors had its beginnings in North American universities in the mid-1920s (d'Apollonia and Abrami, 1997). Today, student ratings are widely practiced by universities around the world (Palmer, 2011), including most Malaysian universities, to help in administrative decision-making. In 2007, Wan Salmuni Wan Mustaffa and Hariri Kamis conducted a study at 17 public Malaysian universities involving 2,580 decision-makers who were responsible for academic staff promotion.

In undertaking this task, student ratings were taken into consideration in the annual reviews of academic staff performance. There is a general consensus in the literature that student ratings provide useful information to higher education administrators. An enormous volume of research supports validity and reliability of student ratings of instruction (e.g. Abrami, 2001; McKeachie, 1997; Greenwald and Gillmore, 1997; Marsh and Roche, 1997; d'Apollonia and Abrami, 1997). Most scholars generally attest to validity of student ratings for making administrative decisions regarding instructional performance of academic staff. Nevertheless, the role that such ratings actually play in various administrative practices is still less clear. Moreover, there are not many studies dealing with how student ratings can be used more effectively (Penny, 2003).

This paper explores the role that student ratings actually have in different administrative practices from perspectives of the administrators. It also includes recommendations by the administrators for a more effective use of student ratings. Student ratings are carried out routinely at the end of each semester. The results are analyzed by the deputy dean's office in each of the 15 faculties of the university. Copies of the results are maintained in the file by the office of the Deputy Dean of Academics to be used in annual reviews of academic staff performance. Another set is sent to department heads for inclusion in the personnel file of the academic staff. The academic staff are also given their own copies for the purpose of enhancing their instructional performance.

2. Method

2.1. Population and Sampling

The present study was conducted at a public research university which had an enrolment of 25,628 students. The population of this study consisted of 191 administrators from 15 faculties including deans, deputy deans, and department heads. The sample size was determined at 110 administrators based on the rule of thumb proposed by Hair et al. (1999) that would be a ratio of five participants to one item. The sample drawn from each faculty was based on stratified proportionate random sampling. Descriptive statistics (frequency, percentage, mean, and standard deviation), Principal Component Analysis (PCA), one way ANOVA, and independent t-test were utilized to analyze the response data.

2.2. Design and Instrumentation

A self-developed instrument consisting of 20 items was used to collect the data. The items, inspired and generated by research literature, were grouped into three parts of A, B, and C. The instrument addressed the administrators' demographic characteristics (Part A), the perceived role of student ratings in administrative practices (Part B), and the administrators' recommendations (Part C). The administrators' demographic characteristics yielded information about their gender, administrative rank, administrative experience, and their evaluation experience in using student ratings for administrative purposes. In order to explore perceived role of the student ratings in administrative practices, the respondents were asked to determine the influence of student ratings on each administrative practice (11 items) using a 4-point Likert scale that ranged from 'Never' to 'Always'. Also, the administrators' recommendations were presented by their levels of agreement with five items, using a four- point Likert scale that ranged from 'Strongly Disagree' to 'Strongly Agree'.

2.3. Validity and Reliability of the Instrument

The levels of content and construct validity were established by a panel of experts (who had professional knowledge in validation and a strong background in design or usage of student ratings of instruction). After confirming the relevance of the items to the content domain by the judgmental procedure, the statistical procedure was conducted in order to examine construct validity of the instrument. A pilot study was conducted with 50 administrators who were not involved in the actual data collection. The result of statistical analysis addressed construct validity of the instrument and provided further support for its content validity. To further verify the strengths of the results, the reliability of the instrument was estimated with the measures of internal consistency including Cronbach's alpha, inter-item correlations, and item to total correlations. The results showed that the reliability of the components was generally high, exceeding standard cut-offs for internal consistency recommended in the literature.

3. Results

3.1. Respondents' Demographic Characteristics

The demographic breakdown of this study found that almost three quarters of the respondents were male (70.5%; n= 62). Also, 10.23% (n= 9) of the respondents were Deans, 29.54% (n= 26) were Deputy Deans, and 60.23% (n= 53) were Department Heads. The administrators had considerable experience in administrative work in higher education and were experienced in using the student ratings for their administrative purposes as well. For instance, they had a mean of about 7 years' experience in administrative duties in higher education and a mean of about 5 years' evaluation experience in using the student ratings for administrative purposes. Hence, they were well qualified to identify the role that student ratings actually played in their administrative practices and were also in a position to make recommendations for more effective use of the student rating of instruction.

3.2. Perceived Role of Student Ratings in Administrative Practices

This study explored the role that student ratings actually play in administrative practices by examining the actual influence of the student ratings on selected administrative practices as perceived by the administrators. A Principal Component Analysis with Varimax rotation was utilized to explore the perceived influence of student ratings on the underlying dimensions of administrative practices. The values of the Bartlett's test of sphericity (.00) and the Kaiser-Meyer-Olkin (.74) addressed the critical assumptions for appropriateness of the PCA for the response data. Both the methods, namely Kaiser Criterion and Scree plot, illustrated the perceived influence of student ratings in three dimensions of administrative practices with eigenvalues greater than 1.0 accounting for 77.83 % of the variance. A meaningful overall communality (.77) was obtained to interpret three extracted dimensions of administrative practices. Table 1 illustrates loadings, individual means, dimension mean, eigenvalues, and percent of explained variance for three explored dimensions of administrative practices regarding the perceived role of student ratings.

The perceived role of student ratings was specified in three administrative dimensions, namely Personnel Decisions (PD), Monitoring-Documenting (MD), and Teaching-Course schedule Development (TCD). The first dimension contained four administrative practices with loading sizes of .94, .94, .94, and .92; the second dimension contained five administrative practices with loading sizes of .80, .85, .80, .59, and .74; and the third dimension comprised two administrative practices with loading sizes of .92 and .92 (Table 1).

The mean for each dimension was produced by averaging the means of the individual administrative practices in the dimension, and was the basis for interpreting the perceived role of the student ratings in each specified dimension. The results indicated that from the perspective of the administrators, student ratings did exert an influence on Personnel Decisions (M=2.82) including administrative practices such as making personnel decisions in annual reviews, tenure, promotion, and reappointment of the academic staff.

Also, from the perspective of the administrators, two other dimensions of administrative practices, namely Monitoring- Documenting (M=2.46) and Teaching- Course schedule Development (M= 1.67) were, however, not influenced by the results of student ratings. The dimension of Monitoring-Documenting (MD) contained 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 dimension of Teaching-Course schedule Development (TCD) reflected the administrative practices pertaining to the development of teaching and course schedules (Table 1).

The finding that student ratings played an important role in administrative decision-making is supported by two leading experts, namely Algozzine et al. (2004) and Abrami (2001) who note that student ratings are valid measures for making personnel decisions. Hence, this finding provides the university with reliable evidence for justifying the use of student ratings of instruction in administrative decision-making. Nevertheless, further studies are still needed to address the ideal/expected role that student ratings should play in various administrative practices from perspectives of administrators.

In addition to the abovementioned analyses, independent t-test and one way ANOVA were utilized to determine significant differences in the perceived role of student ratings in administrative dimensions based on administrators' characteristics constituting the dependent and the independent variables. Tables 2, 3, 4, and 5 illustrate the results of the independent t-test and the one way analyses of variance respectively. The results showed that both male and female administrators agreed that student ratings played an important role in personnel decision-making with regard to the instructional performance of academic staff.

Nevertheless, the female administrators regarded the student ratings on personnel decisions as having a higher influence than did their male colleagues. The findings also indicated that there was no significant difference in the views of male and female administrators regarding the role that student ratings played in two other administrative dimensions. They were in agreement that student ratings did not have a very strong role in administrative practices regarding monitoring/documenting teaching quality and developing teaching/course schedule (Table 2). Three one-way ANOVAs were also performed to assess differences in the perceived role of student ratings in administrative dimensions based on the administrative rank, administrative experience and evaluation experience in using the student ratings. Overall, the results from Tables 3, 4, and 5 showed that regardless of differences in the respondents' administrative position, administrative experience, and experience in evaluating subordinates using student ratings, they held similar views on the role of student ratings in the administrative dimensions.

3.3. Recommendations for the Effective Use of Student Ratings

This study recorded administrators' recommendations for a more effective use of student ratings of their lecturers. A total of 95.5% (n=84) of the participants strongly agreed or agreed that the role that student ratings play in enhancing instructional performance of academic staff should be investigated. Also, a total of 86.3% (n=75) strongly agreed or agreed that research evidence was needed on how academic staff could utilize the results from student ratings to make improvements in their instructional performance. Indeed, the lack of concrete evidence on the role that student ratings play in enhancing the quality of instruction leaves a gap in justifying the use of student ratings for the purpose of instructional improvement.

The majority of the administrators recommended that information from student ratings should be supplemented by other measures of instructional quality. Their recommendations were consistent with the body of literature in this area (e.g. Murray, 2005; Abrami, 2001; Sproule, 2000). For instance, Murray (2005) believes that to rely on student ratings as a sufficient and sole measure of teaching effectiveness is deeply flawed as there are many essential aspects of teaching in universities.

There was also strong consensus (40.9% strongly agreed, 58% agreed; total of 98.9%; n=87) that an available pool of items should be provided so that different faculties/departments could select items suited to their needs. Franklin (2001) recommends that, depending on course objectives and instructional methods, departments and individual instructors should be encouraged to use items relevant to their classes. Hence, an array of items similar to the Cafeteria system should be provided to help academic staff from different faculties make instructional improvement. The Cafeteria, a system for appraising instructors and courses in higher education, currently contains about 200 items within 18 dimensions of instructional practices.

The results from the present study also revealed that more than three quarters of the respondents (76.1 %; n=67) concurred that instructors who had instructional problems should make use of instructional consultation services. Indeed, instructional consultation as a support mechanism can help instructors view student ratings more positively, and this will encourage them to improve their instructional practices (Hodges and Stanton, 2007).

4. Conclusion

This research showed that student ratings played a significant role in university administrative practices. Nevertheless, the ideal/expected role that student ratings should play in administrative practices from the perspective of administrators needs to be explored in greater depth. More broadly, further studies are still needed to address the role that student ratings play in enhancing instructional performance of higher education instructors.

References

Abrami, P.C. (2001). Improving 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.

d'Apollonia, S & Abrami, P. C. (1997). Navigating student ratings of instruction. American Psychologist, 52(11), 1198-1208.

Franklin, J. (2001). Interpreting the numbers: Using a narrative to help others read student evaluations of your teaching accurately. In K. Lewis (Ed.), Techniques and strategies for interpreting student evaluation, (pp. 85-100): New Directions for Teaching and Learning. San Francisco: Jossey-Bass.

Greenwald, A. & Gillmore, G. (1997) Grading leniency is a removable contaminant of student ratings. American psychologist, 52(11), 1209-1217.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, C. (1999). Multivariate Data Analysis. New Jersey: Prentice Hall.

Hodges, L. C., & Stanton, K. (2007). Translating comments on student evaluations into the language of learning. Innovative Higher Education, 31(5), 279-286

Marsh, H. W., & Roche, L.A.. (1997). Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. American Psychologist, 52(11), 1187–1197.

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.

Palmer, S. (2011). The performance of a student evaluation of teaching system. Assessment & Evaluation in Higher Education, 36(0), 1-11.

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.

Sproule, R. (2000). Student evaluation of teaching: A methodological critique of conventional practices. Education Policy Analysis Archives, 8 (50). Available online at http://epaa.asu.edu/epaa.

Wan Salmuni, W. M, & Hariri, K. (2007). Prioritizing academic staff performance criteria in higher education institutions to global standards, Proceedings of the 13th Asia Pacific Management Conference, Melbourne, Australia, pp.1281-1288.

Table 1: Principal Component Analysis of Administrative Practices In Relation to the Perceived Role of Student ratings

Administrative Practices	**Mean**	**Administrative Dimension***		
		PD	**MD**	**TCD**
Personnel decisions regarding annual reviews of academic staff.	2.78	.94		
Personnel decisions regarding tenure.	2.75	.94		
Personnel decisions regarding promotion.	2.93	.94		
Personnel decisions regarding reappointment.	2.81	.92		
Documenting teaching quality.	2.86		.80	
Monitoring improvements in a specific course.	2.55		.85	
Assessing course quality.	2.32		.80	
Comparing teaching quality of departments.	2.06		.59	
Monitoring teaching improvements in faculties/departments.	2.50		.74	
Teaching schedule development.	1.73			.92
Course schedule development.	1.61			.92
Component mean		2.82	2.46	1.67
Eignvalue		3.65	3.19	1.71
% of variance		33.21	29.04	15.58

* Administrative Dimension: Personnel Decisions (PD), Monitoring/ Documenting (MD), Teaching/ Course schedule Development (TCD).

Table 2: Perceived Role of the student ratings in dimensions of administrative practices: Evaluation of gender difference (Independent Sample t-test)

Dimension of Administrative Practices					
Gender	**N=88**	**Mean**	**SD**	**t**	**Sig-t (two tailed)**
Personnel Decisions (PD)					
Male	62	2.68	1.01	2.07	.04
female	26	3.15	.87	2.20	.03
Monitoring-Documenting (MD)					
Male	62	2.41	.66	.97	.33
female	26	2.56	.62	1.00	.32
Teaching- Course schedule Development (TCD)					
Male	62	1.64	.60	.72	.47
female	26	1.75	.65	.70	.48

Table 3: Perceived Role of the student ratings in dimensions of administrative practices: Evaluation of administrative rank differences (One way analysis of variance)

Dimension of Administrative Practices					
Administrative Rank	**N=88**	**Mean**	**SD**	**F**	**Sig-F**
Personnel Decisions (PD)				.50	.60
Department Head	53	2.73	.96		
Deputy Dean	26	2.93	1.05		
Dean	9	3.00	1.03		
Monitoring-Documenting (MD)				2.29	.10
Department Head	53	2.46	.60		
Deputy Dean	26	2.33	.57		
Dean	9	2.86	1.03		
Teaching- Course schedule Development (TCD)				2.17	.12
Department Head	53	1.69	.63		
Deputy Dean	26	1.51	.51		
Dean	9	2.00	.66		

Table 4: Perceived Role of the student ratings in dimensions of administrative practices: Evaluation of administrative experience differences (One way analysis of variance)

Dimension of Administrative Practices					
Administrative Experience	**N=88**	**Mean**	**SD**	**F**	**Sig-F**
Personnel Decisions (PD)				.30	.87
≤ 3	35	2.87	1.00		
4-6	21	2.91	.95		
7-9	9	2.75	.95		
10-12	8	2.50	1.48		
≥13	15	2.76	.80		
Monitoring-Documenting (MD)				2.18	.07
≤ 3	35	2.48	.52		
4-6	21	2.20	.59		
7-9	9	2.71	.73		
10-12	8	2.90	.62		
≥13	15	2.41	.86		
Teaching- Course schedule Development (TCD)				.63	.64
≤ 3	35	1.78	.48		
4-6	21	1.64	.80		
7-9	9	1.44	.46		
10-12	8	1.62	.69		
≥13	15	1.63	.63		

Table 5: Perceived Role of the student ratings in dimensions of administrative practices: Evaluation of administrators' evaluation experience differences (One way analysis of variance)

Dimension of Administrative Practices					
Evaluation Experience	**N=88**	**Mean**	**SD**	**F**	**Sig-F**
Personnel Decisions (PD)				.29	.87
≤ 3	45	2.91	1.02		
4-6	26	2.65	.96		
7-9	3	2.83	.87		
10-12	4	2.93	1.35		
≥ 13	10	2.77	.91		
Monitoring-Documenting (MD)				1.52	.20
≤ 3	45	2.39	.57		
4-6	26	2.51	.52		
7-9	3	2.80	1.31		
10-12	4	3.10	.47		
≥13	10	2.28	1.02		
Teaching- Course schedule Development (TCD)				.30	.87
≤ 3	45	1.68	.51		
4-6	26	1.69	.77		
7-9	3	1.66	.57		
10-12	4	1.87	.62		
≥13	10	1.50	.66		