University. Retrieved from http://www.princeton.edu/mcgraw/library/for-faculty/midcourseevals/

32. Murray, H., Rushton, J., & Paunonen, S. (1990). Teacher personality traits and student instructional ratings in sex types of university courses. Journal of educational psychology. 82, 250-261.

33. Rahman, K. (2006). Learning from your business lectures: using stepwise regression to understand course evaluation data. Journal of American Academy of Business, Cambridge. 19(2), 272–279.

34. Student Ratings of Instruction – Report and Recommendation. Office of Institutional Research. Suffolk County Community College. Retrieved from http://instsrv.sunysuffolk.edu/strate.htm

35. Suggested guidelines for effective use of student evaluations of instruction. (2006). Center for Scholarship in Teaching and Learning. Southeast Missouri State University.

36. Retrieved from http://cstl.semo.edu/cstl/idea/StudentEvalGuidelines.pdf

37. Weinberg, B., Hashimoto, M., & Fleisher, B. (2009). Evaluating teaching in higher education. Journal of Economic Education. 40(3), 227–261.

15. Dunegan, K. & Hrivnak, M. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. Journal of Management Education. 27(3), 280–303.

16. Early Course Evaluations. Eberly Center for teaching Excellence. Retrieved from http://www.cmu.edu/teaching/designteach/teach/early-course-evaluations.html

17. Feldman, K. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multi section validity studies. Research in Higher Education. 30. 583-645.

18. Field, A (2005). Discovering statistics Using SPSS. 2nd ed. London, Thousand Oaks, New Delhi: SAGE Publications.

19. Gilbaugh, J. (1982). Renner Substantiated. Phi Delta Kappan. 63, 428.

20. Greenwald, A. & Gerald M. (1997). Grading Leniency Is a Removable Contaminant of Student Ratings. American Psychologist. 11, 1209-17.

21. Huemer, M. Student Evaluations: A Critical Review. Retrieved from http://home.sprynet.com/~owl1/sef.htm

22. Kim, C., Damewood, E., & Hodge, N. (2000). Professor attitude: its effect on teaching evaluations. Journal of Management Education. 24(4), 458–473.

23. Krautmann, A., & Sander, W. (1999). Grades and student evaluations of teachers. Economics of Education Review. 18(1), 59–63.

24. Langbein, L. (2008). Management by results: student evaluation of faculty teaching and the mis-measurement of performance. Economics of Education Review. 27(4), 417–428.

25. Marsh, H. (1987). Student Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research. International Journal of Educational Research. 11, 253-388.

26. Marsh, H., & Dunkin, M. (1992). Students' evaluations of university teaching; A multidimensional perspective. In J. C. Smart (Ed.), Higher Education: Handbook of theory and research (Vol. 8, pp. 143-233). New York: Agathon.

27. Marsh, H. & Roche, L. (1997). Making students' evaluation teaching effectiveness effective: The critical issues of validity, bias and utility. American Psychologist. 52, 1187-1197.

28. Martin, C., Hanney, S., Henkel, M. & Kogan, M. (1997). The Use of Performance Indicators in Higher Education: The Challenge of the Quality Movement, 3rd ed. London: Jessica Kingsley Publishers.

29. McKeachie, W. (1994). Teaching tips: Strategies, research, and theory for college and university teachers. (9th ed.). Lexington, MA: D. C. Heath.

30. McPherson, M., Jewell, R., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. Eastern Economic Journal. 35(1), 37–51.

31. Mid-semester course evaluations. (2011). McGraw Center. Princeton

effective steps toward improving the teaching while the course is still in progress.

As this study was conducted in a health college (college of applied medical sciences), it is also recommended to reproduce it in a literary college and with students in final level of the program.

References

1. Abrami, C., Levanthal, L. & Perry, R. (1982). Educational Seduction. Review of Educational Research. 52, 446-64.

2. Abrami, P., & D"Apollonia, S. (1990). The dimensionality of ratings and their use in personnel decisions. In M. Theall, and J. Franklin (Eds.), Student ratings of instruction: Issues for improving practice: New Directions for Teaching and Learning, No. 43 (pp. 97-111). San Francisco: Jossey-Bass.

3. Aleamoni, L., & Hexner, P. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. Instructional Science. 9, 67-84.

4. Aleamoni, L. (1987). Techniques for evaluation and improving instruction. New directions for teaching and learning. San Francisco: Jossey-Bass.

5. Ambady, N. & Rosenthal R. (1993). Half a Minute: Predicting Teacher Evaluations from Thin Slices of Nonverbal Behavior and Physical Attractiveness. Journal of Personality and Social Psychology. 64, 431-41.

6. Anastasi, A. & Urbina, S. (1997). Psychological Testing, 7th edition. Upper Saddle River, NJ: Prentice Hall.

7. Arreola, R. (1995). Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system. Bolton, MA: Anker Publishing Co.

8. Braskamp, L. & Ory, J. (1994). Assessing faculty work: Enhancing individual and institutional performance. San Francisco: Jossey-Bass.

9. Cashin, E. (1990). Students do rate different academic fields differently. In M. Theall and J. Franklin (eds.), Student ratings of instruction: Issues for improving practice. New Directions for Teaching and Learning, no. 43. San Francisco: Jossey-Bass Publishers.

10. Cashin, W. (1995). Student ratings of teaching: The research revisited. IDEA Paper No. 32. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

11. Centra, J. (1993). Reflective Faculty Evaluation. San Francisco: Jossey-Bass Publishers.

12. D'Apollonia, S. & Abrami, C. (1997). Navigating Student Ratings of Instruction. American Psychologist. 52 (11), 1198-1208.

13. Davis, B. (1993) Tools for Teaching. San Francisco: Jossey-Bass Publishers.

14. Dershowitz, A. (1992). Contrary to Popular Opinion. New York: Pharos Books Publisher.

evaluations of instruction, 2006). A different timing is suggested by the centre for scholarship in teaching and learning at Southeast Missouri State University. In the centre's guidelines for effective use of student evaluation it was mentioned that the best time to administer teaching evaluations is during the penultimate week of the semester (the week prior to the last week of teaching). It was stated that if administered too early in the semester, students will not have had the time to experience enough of the class to determine if various objectives had been met. While if administered too late in the semester, students could likely suffer from evaluation fatigue and end-of-semester stress, and take their responsibility too lightly (Suggested guidelines for effective use of student evaluations of instruction, 2006). On the other hand, Huemer stated that academic staff who received results of a mid-semester evaluation tended to have higher ratings on end-of-semester evaluations than those who did not, suggesting that student evaluation cause changes in teaching behaviours which result in higher ratings. Furthermore, the improvement was greatest when the academic staff received professional consultation on the interpretation of the evaluations and the student evaluation forms included specific items as opposed to vague items. Agree with the mid-semester student evaluation is the McGraw Centre for teaching and learning, Princeton University which reported that more and more academic staff find that mid-semester evaluations are valuable resources for teaching. And that unlike the evaluations completed at the conclusion of a course, feedback from students midway through the semester provides information that can lead to more adjustments, yielding a better experience for both academic staff and students. Moreover, most academic staff find that students whose views are solicited via mid-semester feedback committed to the course, engage more actively in it, and consider themselves partners and this ensure successful course outcomes (Mid-semester course evaluations, 2011).

The current study researcher believes that student evaluation should be dealt as an instrument for enhancing teaching effectiveness and improving the quality of educational process. That is why she supported the application of student evaluation scale after the midterm exam so that the students will be able to evaluate all items included in the scale and the academic staff will have the time to adjust and modify the course for students' benefits.

Conclusion

The student evaluation scale designed by Saudi National Commission of Assessment and Academic Accreditation and used by College of Applied Medical Sciences has temporal stability and is highly reliable. Administration of the scale in different occasions would not change the students' responses.

Implication on teaching quality

This study indicated that the student evaluation scale used by College of Applied Medical Sciences is reliable one. It can be safely used as one measure to evaluate the teaching effectiveness. Results of this scale would be handled as a valid base for improvement strategies and actions that may be taken by the academic programs.

Recommendation

It is recommended that the best time of this scale administration is one to two weeks after the midterm exam. This time will allow the academic staff to take

the first two responses were of high scores too and if it is a grading leniency; the higher scores would appear at the third response when students got their midterm exam's marks. The previously discussed hypothesis of that the students are already had positive background regarding the evaluated academic staff was the cause of high rating scores from the first to the fifth response. The halo effect may be another explanation. Halo effect refers to the tendency of ratters to be unduly influenced by a favourable or unfavourable general opinion of the person being rated, and then to let that opinion colour all specific ratings. The halo effect causes ratters to make less differentiation between the specific strengths and weaknesses of academic staff or courses than warranted (Anastasi & Urbina, 1997). Any way these explanations and hypothesis should not ignore that students are able to effectively rate their academic staff and that student evaluation scale is an effective and objective tool to evaluate the academic staff teaching performance. The study of Murray and Paunonen (1990) indicated that students' global ratings of academic staff were accurate when compared with trained observers' ratings of specific behaviours of academic staff.

The high Cronbach's α showed in the current study can fairly guarantee the reliability and temporal stability of the student evaluation scale used by KSU. Cronbach's α was high for the four hard copies of the scale administered through the course at four different occasions and for the electronic copy as well. Although reliabilities vary according to the number of ratters and this study included only 13 students, the results of this study are valid as Cashin (1995) recommended a minimum of ten ratters for an acceptable reliability of 0 .70 or better.

Researcher of current study suggests administration of the student evaluation scale one-two weeks after the midterm exam. This would give chance to the students to develop their thoughts regarding the course academic staff based on full experience with the academic staff in all aspects; organization of the course, teaching effectiveness, grading exams and student's interaction. By that time, students would be more capable to answer all items of the scale more objectively. Meanwhile, two weeks after the midterm exam gives a chance to the academic staff to get benefit from the scale results and apply the required modification to increase the teaching effectiveness. Let us not forget that student evaluation scale is not for overwhelming students or to frighten the academic staff but it is a measure to improve the quality of education, a condition both student and academic staff can benefit from.

Researchers showed diversity in their opinion regarding the optimal time for student evaluation scale administration. D'Apollonia and Abrami (1997) found that the association between student rating and student learning was significantly higher when academic staff evaluation was carried out after, rather than before, the final examination. They believed that students may be rewarding academic staff who had given them high grades, or they may be using grading or their performance on the final examination as one of the indicators of academic staff effectiveness. Controversy, the Eberly centre for teaching excellence at Carnegie Mellon University stated that the best time to conduct an evaluation of a course in progress is in the first three to six weeks of a full semester course. Because this time frame gives students a reasonable sample of how academic staff teaches and how their learning is evaluated to make substantive comments (Early Course Evaluations). On the other hand it allows academic staff time to make adjustments and see their impact (Suggested guidelines for effective use of student

should be highly assured. This is especially important hence the results of this scale are used to develop the program improvement action plans as earlier discussed and they guide the courses' academic staff' efforts to professionally improve themselves. That is why this study aimed to investigate the temporal stability of student evaluation scale administered by King Saud University through answering the following question: If the student evaluation scale was introduced to students in five different occasions, would their responses vary?

Reviewing literature showed that exam marks are the most important factor that may bias the students' ratings to the student evaluation scale. Weinberg et al., (2009) found that student evaluation scores in first-year economics courses at Ohio State University are positively related to the grades academic staff assign. Others have also found a positive relationship between grades and student evaluation scores (Langbein, 2008 & McPherson et al., 2009). A paper by Krautmann and Sander (1999) documented that the grades students expect to receive in a course are positively related to student evaluation scores. Consequently, the researcher arranged the collection of the four different responses of the scale in consideration to the exams' times, mid and final term exams. The first response two weeks before the midterm, the second response during the midterm exam, the third response two weeks before the final exam and the fourth response during the final term exam. The fifth response is an electronic one and is collected within 24-48 hours after the final term exam when the course academic staff finishes marking the exam sheet and posts the results online to be accessible to all students. Results of this study, repeated measures ANOVA test, showed that there were changes in the mean ratings of the students between the five responses but these differences did not reach to the level to be statistically considered. Moreover, it was remarkable that the ratings were high in all items from the first to the last response as shown in table one. Another observation is that some of the scale items are asking about the grading and exam system. These items are not encountered early in the beginning of the course except for the weekly assignment. So, these items suppose to have variation from the first response to the next responses as the course goes on and the students more and more evaluate the academic staff performance regarding the grading and exam system. This could be explained by the fact that those students are of advanced level, level eight, and already had previous courses with the academic staff so they had saved perception which they could not eliminate while they were rating the current course student evaluation scale. Some researchers have suggested that student motivation and course level may bias student ratings of academic staff (D'Apollonia & Abrami, 1997) and those higher level courses, especially graduate level courses, tend to receive higher ratings (Aleamoni & Hexner, 1980, Braskamp & Ory, 1994 and Marsh & Roche, 1997).

Revising the students' grades showed that eight out of the 16 students enrolled in the course had A+ grade i.e. 50 percent of the students had the highest course's grade which is considered a high percentage. This raises the question of: Is the grading leniency is applied here? Or it is as optimists have suggested that the correlation between high students' grades and high academic staff rating scores might be due to the fact that greater teaching effectiveness on the part of the academic staff leads to both higher grades and higher ratings of the academic staff. Thus, the effect might actually be a sign of the validity of student ratings (Gilbaugh, 1982). Researcher believes that it is not a grading leniency because

	A	B	A	B	A	B	A	B	A	B
11	0.655	0.966	0.688	0.688	0.762	0.957	0.537	0.944	0.962	0.985
12	0.891	0.965	0.701	0.701	0.862	0.956	0.716	0.942	0.962	0.985
13	0.820	0.965	0.701	0.701	0.773	0.957	0.573	0.944	0.962	0.985
14	0.577	0.967	0.408	0.408	0.711	0.958	0.714	0.942	0.962	0.985
15	0.852	0.965	0.879	0.879	0.717	0.956	0.707	0.942	0.974	0.985
16	0.876	0.964	0.721	0.721	0.662	0.958	0.748	0.941	0.974	0.985
17	0.973	0.964	0.611	0.611	0.938	0.955	0.824	0.941	0.962	0.985
18	0.877	0.964	0.001	0.001	0.933	0.955	0.776	0.941	0.892	0.986
19	0.909	0.964	0.767	0.767	0.394	0.960	0.681	0.942	0.725	0.986
20	0.547	0.967	0.405	0.405	0.559	0.959	0.697	0.942	0.885	0.986
21	0.547	0.967	0.405	0.405	0.286	0.961	0.607	0.943	0.962	0.985
22	0.620	0.967	0.456	0.456	0.448	0.959	0.509	0.945	0.962	0.985
23	0.587	0.967	0.255	0.255	0.447	0.959	0.524	0.944	0.767	0.986
24	0.896	0.965	0.560	0.560	0.867	0.956	0.702	0.942	0.885	0.986
25	0.923	0.964	0.688	0.688	0.892	0.955	0.805	0.941	0.709	0.987
26	0.827	0.965	0.674	0.674	0.914	0.955	0.701	0.942	0.709	0.987
C	0.967		0.928		0.960		0.945		0.986	

A= Correlated item-Total correlation, B= Cronbach's Alpha if item deleted, C= Cronbach's Alpha

Discussion

The vice deanship of quality and development at College of Applied Medical Sciences underwent remarkable steps toward application of King Saud University quality deanship's recommendation about the application of student evaluation scale to collect feedback regarding the effectiveness of teaching. It developed quality management system through which the student evaluation scale was posted on line with easy accessibility and clear instructions to students to answer the items of the scale. Students' ratings are collected, statistically treated and report of rating means of each item is prepared and sent to the quality committee at each department and program within College of Applied Medical Sciences to be interpreted and used for improvement action plans. Moreover, the student's ratings to the scale items are used to calculate an overall score of the scale with maximum of five points. Course academic staff who receives less than three out of five in this overall scale score is interviewed by the department chair to clarify the reason for low score and to explain the action (s) he/ she will follow to improve his/ her teaching effectiveness.

The student evaluation scale adopted is the one designed by the Saudi National Commission of Assessment and Academic Accreditation and it is recommended to be applied during the mid semester.

Since the scale is self-answered by students and represents their subjective perceptions to their academic staff, the reliability and validity of the used scale

25	13	1	4.2 (0.83)	2.25	2.64	0.85	1.49	0.240
		2	4.5 (0.52)					
		3	4.4 (0.77)					
		4	4.2 (0.56)					
	16	5	4.0(0.41)					
26	13	1	4.2 (0.99)	2.86	2.53	1.13	1.39	0.266
		2	4.6 (0.51)					
		3	4.4 (0.96)					
		4	4.2 (0.56)					
	16	5	4.0 (0.41)					

Results also showed that the test-retest reliability of the studied student evaluation scale is markedly high. The Cronbach's a of the scale was higher than 0.8, which is the value of a considered to be good, in the five responses as shown in table two.

Table Two: Cronbach's a of the student evaluation of academic staff scale in five responses.

Item	Response 1		Response 2		Response 3		Response 4		Response 5	
	A	B	A	B	A	B	A	B	A	B
1	0.734	0.966	0.416	0.416	0.862	0.956	0.636	0.943	0.932	0.985
2	0.711	0.966	0.674	0.674	0.933	0.955	0.602	0.943	0.962	0.985
3	0.722	0.966	0.486	0.486	0.911	0.956	0.736	0.942	0.962	0.985
4	0.575	0.967	0.464	0.464	0.680	0.958	0.703	0.942	0.932	0.985
5	0.641	0.967	0.590	0.590	0.078	0.963	0.604	0.943	0.892	0.986
6	0.669	0.966	0.601	0.601	0.866	0.957	0.557	0.944	0.932	0.985
7	0.676	0.966	0.352	0.352	0.474	0.959	0.477	0.944	0.904	0.986
8	0.672	0.966	0.841	0.841	0.500	0.959	0.351	0.948	0.816	0.986
9	0.590	0.967	0.659	0.659	0.708	0.957	0.748	0.942	0.280	0.989
10	0.558	0.968	0.514	0.514	0.556	0.959	0.502	0.947	0.931	0.987

20	13	1	4.8 (0.44)	3.29	2.62	1.26	2.84	0.060
		2	4.7 (0.48)					
		3	4.4 (0.65)					
		4	4.2 (0.69)					
	16	5	4.4 (0.65)					
21	13	1	4.8 (0.44)	2.70	2.92	0.81	2.53	0.074
		2	4.7 (0.48)					
		3	4.6 (0.51)					
		4	4.5 (0.52)					
	16	5	4.2 (0.60)					
22	13	1	4.5 (0.66)	2.37	3.42	0.70	1.39	0.257
		2	4.4 (0.65)					
		3	4.5 (0.52)					
		4	4.0 (0.82)					
	16	5	4.2 (0.60)					
23	13	1	4.5 (0.52)	1.14	3.11	0.37	1.31	0.287
		2	4.9 (0.23)					
		3	4.7 (0.44)					
		4	4.6 (0.51)					
	16	5	4.7 (0.63)					
24	13	1	4.4 (0.65)	2.00	2.93	0.68	1.43	0.251
		2	4.6 (0.51)					
		3	4.5 (0.88)					
		4	4.1 (0.76)					
	16	5	4.4 (0.60)					

15	13	1	4.1 (0.86)	1.54	2.90	0.54	0.79	0.500
		2	4.2 (0.80)					
		3	4.4 (0.51)					
		4	3.9 (0.76)					
	16	5	4.2 (0.60)					
16	13	1	4.1 (0.86)	2.25	2.87	0.78	0.97	0.414
		2	4.5 (0.52)					
		3	4.4 (0.87)					
		4	4.0 (0.91)					
	16	5	4.1 (0.86)					
17	13	1	4.3 (0.63)	0.71	3.20	0.22	0.50	0.690
		2	4.5 (0.66)					
		3	4.5 (0.78)					
		4	4.2 (0.60)					
	16	5	4.2 (0.60)					
18	13	1	4.2 (0.90)	1.63	3.06	0.53	0.74	0.536
		2	4.4 (0.87)					
		3	4.5 (0.66)					
		4	4.2 (0.56)					
	16	5	4.2 (0.60)					
19	13	1	4.0 (1.00)	0.77	2.84	0.27	0.33	0.797
		2	4.2 (0.93)					
		3	4.2 (0.56)					
		4	4.1 (0.76)					
	16	5	3.9 (0.49)					

10	13	1	4.2 (0.89)	2.62	3.49	0.75	0.63	0.623
		2	4.2 (0.93)					
		3	4.1 (0.86)					
		4	3.7 (1.11)					
	16	5	3.9 (1.21)					
11	13	1	4.4 (0.65)	0.99	3.15	0.312	0.81	0.502
		2	4.5 (0.52)					
		3	4.5 (0.66)					
		4	4.2 (0.44)					
	16	5	4.2 (0.60)					
12	13	1	4.4 (0.51)	1.79	3.47	0.51	1.55	0.211
		2	4.7 (0.48)					
		3	4.5 (0.66)					
		4	4.3 (0.63)					
	16	5	4.2 (0.60)					
13	13	1	4.4 (0.77)	0.71	2.66	0.27	0.43	0.709
		2	4.5 (0.66)					
		3	4.5 (0.66)					
		4	4.2 (0.44)					
	16	5	4.2 (0.60)					
14	13	1	4.7 (0.48)	1.69	3.37	0.50	2.44	0.071
		2	4.6 (0.51)					
		3	4.7 (0.48)					
		4	4.4 (0.51)					
	16	5	4.3 (0.48)					

5	13	1	4.0 (0.82)	2.71	3.31	0.82	1.50	0.227
		2	4.3 (0.75)					
		3	4.3 (0.63)					
		4	3.8 (0.73)					
	16	5	4.2 (0.56)					
6	13	1	4.5 (0.66)	1.63	3.04	0.54	1.40	0.258
		2	4.7 (0.48)					
		3	4.7 (0.48)					
		4	4.4 (0.51)					
	16	5	4.3 (0.63)					
7	13	1	4.8 (0.44)	0.69	1.00	0.69	1.32	0.273
		2	4.8 (0.44)					
		3	4.5 (0.66)					
		4	4.6 (0.51)					
	16	5	4.2 (0.73)					
8	13	1	3.9 (0.89)	1.69	3.31	0.51	0.70	0.572
		2	3.9 (0.80)					
		3	4.2 (0.59)					
		4	3.8 (0.93)					
	16	5	3.9 (0.76)					
9	13	1	4.3 (0.63)	3.84	3.10	1.26	2.25	0.098
		2	4.5 (0.66)					
		3	4.4 (0.65)					
		4	4.1 (0.49)					
	16	5	3.9 (0.80)					

Table One: One-way repeated measures ANOVAs of the student evaluation of academic staff 26 items scale.

Item	N	Response	Mean (SD)	Sum of Squares	df	Mean Squares	F	Sig
1	13	1	4.4 (0.65)	2	3.13	0.64	1.81	0.158
		2	4.8 (0.44)					
		3	4.5 (0.52)					
		4	4.3 (0.48)					
	16	5	4.3 (0.63)					
2	13	1	4.3 (0.86)	1.69	3.22	0.53	0.91	0.451
		2	4.6 (0.51)					
		3	4.5 (0.66)					
		4	4.2 (0.73)					
	16	5	4.2 (0.56)					
3	13	1	4.5 (0.52)	1.69	4.00	0.42	1.65	0.177
		2	4.6 (0.51)					
		3	4.6 (0.51)					
		4	4.3 (0.48)					
	16	5	4.2 (0.56)					
4	13	1	4.6 (0.51)	1.32	2.76	0.48	1.75	0.179
		2	4.7 (0.48)					
		3	4.5 (0.52)					
		4	4.3 (0.48)					
	16	5	4.4 (0.51)					

National Commission of Assessment and Academic Accreditation. It is composed of 26 items. These items rate the students' perception to course organization and planning (five items), clarity and communication skills (six items), teacher-student interaction (six items), grading and examination (six items), and student-self-rated learning (two items). The last item was about the preference of students to study another course with the same academic staff. Students were required to rate each item according to their perceptions about the instructor performance at that item using five-point rating numerical scale with one representing 'not at all descriptive' and five 'very descriptive'. Four hard copies of the student evaluation scale were used to have the students' responses in four different occasions. The first response was two weeks before the midterm exam, the second response was after collecting the answer sheets of midterm exam, the third response was two weeks before the final term exam and the fourth response was after collecting the answer sheets of final term exam. In addition, a fifth response in which an electronic copy of the form was used, this electronic copy is a regular routine in King Saud University as students are not allowed to access their course marks, at the end of the semester, unless they answer the form on line.

Data collection

Data were analysed using SPSS statistical software, version 13. For each item of the scale, One-way repeated measures ANOVA was used to compare the means of students' ratings for the 5 responses. P value less than 0.05 was considered significant.

For each scale response, the Cronbach's Alpha (α) was used to investigate the test-retest reliability of the scale. The values of Cronbach's α for each item of the scale were measured and a value of 0.8 was considered as good value (Marsh, 1987).

Results

Results of this study showed that, for each item of the scale; the students' ratings were consistent throughout their five responses. The statistical difference between the mean ratings of the item from response one to five was not significant; this was applicable to the 26 items of the scale as shown in table one.

They identified six factors commonly found in student rating scales: 1- Course organisation and planning, 2- Clarity and communications skills, 3- Academic staff student interaction, rapport, 4- Course difficulty, workload, 5- Grading and examinations, and 6- Student self-rated learning (Abrami & D"Apollonia, 1990, Braskamp & Ory, 1994, Feldman, 1989, Marsh & Dunkin, 1992).

Student evaluation is a tool to produce useful feedback which the academic staff and program can use to improve their quality of education. The process of (a) gathering information about the impact of learning and of teaching practice on student learning, (b) analyzing and interpreting this information, and (c) responding to and acting on the results, is valuable for several reasons (Rahman, 2006). It is beneficial because academic staff can review how others interpret their teaching methods, thereby improving their instruction. The information can be also used by academic staff committees (Davis, 1993) and administrators, along with other input, to make summative decisions (e.g., decisions about promotion, tenure, salary increases, etc.) and make formative recommendations (e.g., identify areas where a faculty member needs to improve) (Dunegan & Hrivnak, 2003).

In spite of the above, student evaluation have came under fire on several fronts. The most common criticism of student evaluation seems to be that it is biased; students tend to give higher ratings when they expect higher grades in the course, this correlation is well-established (D'Apollonia & Abrami, 1997) and many authors believe that this causes rampant grade inflation (Dershowitz, 1992). In one survey, 70 percent of students admitted that their rating of an academic staff was influenced by the grade they expected to get (Greenwald & Gerald, 1997). In another survey, 38 percent of professors admitted to make their courses easier in response to student evaluation (Marsh, 1987).

In the last five years King Saud University showed improvements in many of its practices. The deanship of quality adopted the use of many surveys and scales to collect feedback regarding the effectiveness of teaching. The deanship recommended that these scales to be analyzed and to be used as a base for improvement strategies within the colleges. Among these scales is the student evaluation scale.

This study aimed to investigate the temporal stability of student evaluation scale administered by the College of Applied Medical Sciences, King Saud University through answering the following question: If the student evaluation scale was introduced to students in five different occasions, would their responses vary?

Method

The current study was conducted at King Saud University- College of Applied Medical Sciences- Department of Rehabilitation Health Sciences. It included an application of student evaluation scale to evaluate one course academic staff who was teaching a course for level eight students during spring semester of the academic year 2009-2010. Sixteen students were enrolled in the course and 13 students participated in the study. The researcher informed the students that the student evaluation scale, which is usually used by King Saud University at the end of the semester, will be introduced on four other occasions during the semester. The King Saud University form for student evaluation was designed by the Saudi

experiences in the class over the term and also because a substantial body of research has concluded that administering questionnaires to students is valid and reliable (Davis, 1993). According to a review of the literature conducted by Aleamoni (1987) and Arreola (1995), well-developed, tested, student rating forms of teaching effectiveness exhibit both reliability and validity. Moreover, student ratings are the single most valid source of data on teaching effectiveness (McKeachie, 1994). A test is said to be 'reliable' if it tends to give the same result when repeated. A test is said to be 'valid' if it is measuring what it is intended to measure. Student evaluation are highly reliable, in that students tend to agree with each other in their ratings of an academic staff, and that they are at least moderately valid, in that student ratings of course quality correlate positively with other measures of teaching effectiveness (Huemer, 2011). Student evaluation also tends to be correlated well with retrospective evaluations by alumni; in other words, former students rarely change their evaluations of their academic staff as the years pass (Martin et al., 1997). Furthermore, other methods of evaluating teaching effectiveness do not appear to be valid. Ratings by colleagues and trained observers are not even reliable (a necessary condition for validity) that is colleagues and observers do not even substantially agree with each other in academic staff ratings (Centra, 1993). Reliability varies according to the number of raters, such that Cashin (1995) recommended a minimum of ten raters for an acceptable reliability of 0.70 or better. Students can judge how well prepared academic staff are, how effectively they make use of class time, how well they explain things, with what level of enthusiasm, and how responsive they are to difficulties the students may be having in the course. Students can also comment on whether the academic staff promotes original thinking and critical evaluation of ideas (Davis, 1993). They can also evaluate variables such as communication skills, organisational skills, flexibility, attitude toward the student, teacher-student interaction, encouragement of the student, knowledge of the subject, clarity of presentation, course difficulty, fairness of grading and exams, and global student rating (Kim, 2000).

It was advised by Cashin (1990) to provide a numerical rating scale, for at least some of the key items of student evaluation. The use of quantifiable items enables the academic staff to calculate a class's average response and to note the distribution of responses, both of which are useful in interpreting the results of the evaluation. Either a five-point or seven-point scale should be used, with one representing 'not at all descriptive' and five (or seven) 'very descriptive'. It is also appropriate to provide a 'Don't know or doesn't apply' option that students can tick. In addition at least one quantitative measure on the overall effectiveness of the academic staff should be included and at least one open-ended item that asks about the overall effectiveness of the academic staff. For example, 'Please identify what you perceive to be the greatest strengths and weaknesses of this academic staff's teaching'. On the other hand, Huemer (2011) believes that an objective student evaluation which may includes items such as 'Did the professor come to class on time?', 'Did he read student work and return it within a reasonable time frame?', and so on, is more preferable than subjective items such as 'How would you rate this instructor?' or 'How fair was the grading?' because the former types of question would probably be less subject to the effects of bias than the latter. Researchers agree that teaching effectiveness is a multidimensional construction. Thus student ratings forms need to, and in fact do, measure a variety of different aspects of teaching.

Temporal Stability of the Students Scale used for Academic Staff Evaluation

Salwa B. El-Sobkey

Abstract

Student evaluation of academic staff is one of the most common practices for evaluating academic staff performance. This study aimed to investigate the temporal stability of a student evaluation scale administered by College of Applied Medical Sciences, King Saud University. A 26-item student evaluation scale designed by the Saudi National Commission of Assessment and Academic Accreditation was used. Thirteen students participated in this study. They responded to the student evaluation scale during the course on four separate occasions in addition to a fifth response to the electronic version of the scale at the end of the semester. The statistical difference between the mean ratings of the item from responses one to five was not significant; this was applicable to the 26 items of the scale. Results also showed that the test-retest reliability of the studied student evaluation scale is markedly high. The Cronbach's α of the scale was higher than 0.8 in the five responses. It was concluded that the studied student evaluation scale has temporal stability and is highly reliable. Study recommended that the best time of this scale administration is one to two weeks after the midterm exam.

Key words:

student evaluation of academic staff- scales reliability- guidelines of student scale administration.

Introduction

Colleges and universities seeking to improve teaching quality may take one or more of the following measures: academic staff members could be offered courses or workshops on improving teaching effectiveness; teaching can be evaluated in part by examination of syllabi and other course materials; or forms could be designed to emphasise objective matters for student evaluation of academic staff (D'Apollonia & Abrami, 1997; Huemer, 2011). Informal student evaluation was started in the 1960s (Abrami et al., 1982). Since then, its use has spread so that now it is administered in almost all colleges and universities and is probably the main source of information used for evaluating academic staff teaching performance (Ambady & Rosenthal, 1993). Student rating forms or scales are paper and pencil instruments on which students indicate their response to items on some numerically based scale (D'Apollonia & Abrami, 1997). Many multidimensional student scales contain items that assess the instructor's believes, knowledge, attitudes, organization and preparation, ability to communicate clearly and to stimulate students, interaction with students, feedback, workload, and fairness in grading. So, academic staff are actually evaluated in terms of three roles: presenter, facilitator, and manager ((D'Apollonia & Abrami, 1997). Such forms have become commonplace because it makes sense to survey students to find out what they think about their

Temporal Stability of the Students Scale used for Academic Staff Evaluation

Salwa B. El-Sobkey
King Saud University