

What do Student Ratings Reveal? A Case Study of a Saudi Arabian University

Moubarak Al Kahtani and Khalid Alotaibi*

College of Education, Prince Sattam bin Abdulaziz University, Saudi Arabia.

Received: January 7, 2016

Accepted: March 22, 2016

ABSTRACT

Evaluation of teachers by students has become a much practiced component of evaluating a teacher's performance. This study evaluates students' perceptions of their teachers from a survey of 1120 male and female students conducted at the College of Education at Prince Sattam Bin Abdulaziz University. The perceptions were assessed from responses to four major aspects of teaching: an instructor's ethics; their competence; use of appropriate and effective methods of students' evaluation; and satisfaction of students towards a teacher's overall performance. The findings of the study are that professional ethics followed by competency are the most affective traits that affect students' perceptions. Also gender wise analysis reveals that male and female students have significant difference in their mean score of responses regarding professional ethics and teachers' competency with higher mean by male students showing that male students have better perceptions regarding faculty. Also analysis by level of seniority shows that senior students had higher mean score of their responses than junior students in professional ethics and teachers' competency. The ANOVA reveals that gender and seniority level have interaction affect in the traits of professional ethics and teachers' competency. The principal policy recommendation is that ethics and subject competencies should be stressed in professional development of teachers.

KEY TERMS: Student evaluations; faculty performance; teaching effectiveness; classroom observations.

INTRODUCTION

Perceptions of students about their teachers have become a common measure of student satisfaction about their institutions and their faculty. A good reason for conducting such surveys is that the quality of education cannot be merely assessed from determining a good syllabus or hiring competent faculty at an educational institution. At the end of the day a client's (student's) satisfaction is paramount in determining the quality of the educational service received by them. Also a student's learning is affected by the 'perceived' quality of teaching received from teachers. Thus a teacher's relationship with students and their acceptance of the teacher becomes a factor in creating a conducive learning and teaching environment.

Ratings of teachers by students are only one of the methods through which teacher quality can be evaluated. The other methods of judging the quality of a teacher are through evaluation by peers or superiors and through research output. Despite these alternative methods of assessing the quality of a teacher the use of students' ratings for evaluating teacher effectiveness is one of the most researched issues in higher education (Doyle, 2004). In Saudi Arabia too, teacher evaluations are now required at many universities and have been introduced by the quality assurance departments as an indicator of teaching quality and faculty teaching performance.

At the outset we wish to state that concerns over validity of student ratings shall always remain in some form, however, we wish to summarize some findings on this important issue. McKeachie (1997) concluded that although students' ratings were a valid measure of gauging teacher effectiveness it was an imperfect measure at best; and while these ratings can be used in judgements by superiors their use needs to be improved. d'Apollonia and Abrami (1997) observed that the ratings are a valid measure of teacher effectiveness as they could not find evidence of biasedness based on student related characteristics. They opined that such ratings should be interpreted with care and not given undue importance to the neglect of other factors related to teaching quality. Marsh and Roche (1997) also came to the same conclusions. However, Greenwald and Gilmore (1997) observe that student ratings are biased by student grades and lead to grade inflation.

Notwithstanding any unresolved issues of validity of ratings due to bias, such ratings can still be instrumental in improving educational quality through constructive feedback from students for improvement of teaching. Much can be learnt about effects of feedback from teacher evaluations conducted in primary and secondary schools and

* **Corresponding Author:** Moubarak Al Kahtani, College of Education, Prince Sattam bin Abdulaziz University, Saudi Arabia.

then the same feedback can be tied to teacher retention and promotion. In Washington DC which has a high number of schools with low achieving students, teacher effectiveness in public schools was tied to an increase in student achievement in the national student achievement test. The teachers are assessed on multiple criterion including contributions to community, contributions to student learning as measured by their achievement test scores and other professional services and the good teachers were retained while others were relieved of their services (Ortega and Crutchfield, 2014). At this point in time no such high-stakes system is in place in Saudi Arabian schools. However, teacher evaluations by students are more common in universities as required by the NCAAA accreditation requirements. Al-Kuwaiti (2014) reports that in a sample of 7 public universities, both students and teachers agreed that there was a positive impact of student evaluations of teaching effectiveness (SETE). However the students assessed the impact to be lower than what was assessed by the teachers.

Considerable research has been conducted to find out the determinants of good teaching. In a comprehensive study for all 50 states in USA, Hammond (2000) found that teacher qualifications were correlated with student achievement. The policy conclusion was that investments in teacher training and certification would boost student achievement. Although student perception surveys have been shown to be effective in evaluating instructors' performance, their use in performance reviews is not widespread (MET Project, 2012). However Clifford et al. (2010) advocate that performance of instructors should be evaluated by students' surveys in a reliable way. Researchers emphasize that the validity of student perceptions is dependent on the survey itself and that a well-designed survey with correct instruments and a representative sample could lead to valid results (Goe and Little, 2008).

Research has shown that some of the most important student characteristics that affect student ratings of teachers are the expected grades of students but it was commonly believed that there were limits on an instructor's ability to make their students give them high ratings. However, Clifford (2007) observed that instructors lower their grading standards and in this way buy higher student ratings. Other factors that affect student ratings are classroom characteristics. These include such factors as time of day, class size, subject matter and class level (Millea and Grimes 2002). When conducting faculty's evaluations, it would be appropriate for administrators to control for differences not directly under faculty members' direct influence. This is important because substantial evidence suggests that students evaluate faculty based on a myriad of factors, not all of which are controlled by the faculty (Isley and Singh, 2005). Also student evaluations of teacher effectiveness (SETEs) are affected by a number of variables that may have little or nothing to do with evaluating the teachers' ability to teach (Simmons, 1998). A teacher's gender also plays a role in how students evaluate their teachers' performance. Sprague and Massoni (2005) show that the gender bias in the evaluations may not be detected from quantitative scales rather qualitative questions can reveal more. They found that there was greater hostility towards female teachers who did not meet students' expectations. These results would not be completely applicable in the case of Saudi Arabia where gender segregation is practiced in all educational institutions. Nevertheless, it would be interesting to see whether the same conclusions are tenable for the male instructors when the professor and the females cannot see their male instructors as is the case in Saudi Arabia.

Though the debate over the structure and usability of student course evaluations of instruction continues, some researchers assert the usefulness of these evaluations as a means of transforming teaching and learning. Bubb et al. (2013) strongly assert the usability of formative student input with regard to instructional improvement. They conclude that while response rates of students have reduced due to electronic collection of ratings, they demonstrate the usefulness of a formative mid-course evaluation by students. This research builds on earlier studies by Stark-Wroblewski, et al. (2007).

Several studies demonstrate students' reliability and effectiveness in terms of evaluating teaching behaviors such as presentation, clarity, organization, and active learning techniques (Nasser and Fresko, 2002). Hodson (2012) notes that teaching effectiveness can be evaluated based on students' feedback, external examiner inputs, analysis of data e.g. enrollment, progression, and graduation, review of course planning, teaching and assessment compatibility, teaching staff personal review and observation of a teachers classroom conduct. Additionally, the effectiveness of planned teaching strategies in achieving different types of learning outcomes should be regularly assessed and adjustments should be made in response to evidence regarding their effectiveness. In this context, the quality cycle is an important tool of evaluation of teaching effectiveness. It encompasses quality across the educational institution which includes quality of student performance, institution's concern for quality and teacher effectiveness. Students take active role that everything that happens at the college or university should meet certain minimum quality standards.

Additionally (Berk 2005) identifies twelve potential sources of evidence of teaching effectiveness: (1) student ratings, (2) peer ratings, (3) self-evaluation, (4) videos, (5) student interviews, (6) alumni ratings, (7) employer ratings, (8) administrator ratings, (9) teaching scholarship, (10) teaching awards, (11) learning outcome measures, and (12) teaching portfolio. National standards are presented to guide the definition and measurement of effective teaching. A unified conceptualization of teaching effectiveness is suggested to use multiple sources of evidence,

such as student ratings, self-evaluation, and peer ratings, to offer an accurate and reliable base for formative and summative decisions. Multiple sources build on the strengths of all sources, while compensating for the weaknesses in any single source. This triangulation of sources is recommended in view of the complexity of measuring the act of teaching and the variety of direct and indirect sources and tools used to produce the evidence.

The Center for Research on Learning and Teaching (CRLT, 2015) at the University of Michigan stresses the necessity to consider a number of factors in teaching evaluations and involving multiple sources of data as well as involvement of faculty in the development of students rating questionnaires. Also effective teaching evaluation must be individualized. A uniform system may discriminate against some individuals, so a plan sensitive to individual variation should be developed and as teaching includes activities broader than classroom instruction, the evaluation of teaching should assess more than classroom performance e.g. curriculum development, mentoring of graduate students, students advising and one on one consultation with students.

Murray (2005) indicates that student ratings are appropriate in terms of reliability, in that ratings of a given instructor are reasonably stable or consistent across courses, years, rating forms, and groups of students involved in rating. Other research shows that student evaluations are valid or accurate in that they are relatively free of bias, and agree with evaluations made by others, such as colleagues and alumni. He substantiates the validity of student ratings through classroom observation of specific behavior of teachers. The observers reports provide evidence that students infact correctly predict teacher behavior. Further evidence to support the validity of student ratings comes through multi-section validity studies.

Conceptual frame work

The conceptual framework rests on the premise of the earlier literature that student ratings are valid and provide a useful observation to assess a teacher's abilities and conduct in the classroom. In our study teachers were evaluated through the core attributes of Professional Ethics, Competence, Teaching Styles and the Assessment Styles of teachers. These attributes are considered as a measure of an instructor's skills which are assessed by students through a survey response. The paper is influenced from the proposed model of Cowie and Bell (1999) that distinguishes between planned and interactive formative assessment. In the paper ethics is related to a teacher's efforts in creating a learning environment that exemplifies the highest ethical standards and a treatment of equity towards all students. This set of ethics is developed by the American Association of Educators and the competence standards for instructors are adopted from American Federation of Teachers who developed the criteria to evaluate teachers in USA. These constructs were labeled in the factor analysis of this paper as p1..p8 for professional ethics, c1..c4 for competence t1..t6 for teaching style and a1..a7 for assessment style (See Table 1).

RESEARCH METHODOLOGY

While there are a multitude of factors that can be used to measure teaching effectiveness, this study gauges effectiveness from a survey analysis of students. A random sample of 1120 male and female students was taken at different levels of educational attainment and these are referred as senior and junior in our sample. The sample comprised of 42% male and 58% female students. Students' perceptions regarding the instructor's ethics, instructor competence, instructor teaching styles and strategies and their assessment styles were sought using a five point Likert scale. In their response the students expressed their opinion by the options of strongly disagree, disagree, neutral, agree, or strongly agree.

Validity of the instrument

For content validity expert opinions were sought in the relevant field. Item suitability regarding each construct, suitability of wording and face validity was ensured in the light of their opinions.

Table 1: Factor Analysis of Students' Questionnaire

	Rotated Component Matrix ^a				Component
	1	2	3	4	
p1	.59				
p2	.55				
p3	.45		.39		
p4	.55				
p5	.55				
p6	.49				.42
c1			.49		

c2			.42	
c3			.50	
c4			.48	
t1				.46
t2	.30			.46
t3				.49
t4		.382		.39
t5				.45
t6			.462	.46
a1	.32	.46		
a2		.30		
a3		.37		
a4		.36		
a5		.53		.41
a6		.51		.42
a7		.52		

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Table 1 shows that most of the items were found loading on to their respective components. Some of the items were found cross loading as well onto other constructs but as a whole they were found loading sufficiently to their respective component which confirms the theoretical model in this work. However two items p7 and p8 were found not sufficiently loading on the model so were deleted for the final analysis.

Analysis

The data was analyzed for correlation to see mutual relationship between all four constructs and also to see how far is each construct related with overall perceptions of the students regarding their faculty. Table 2 presents the results.

Table 2: Correlation Coefficients of Components and Total Perceptions

		Competency	Teaching style	Assessment style	Total
Ethics	Pearson Correlation	.43*	.42*	.48*	.549**
	Sig. (2-tailed)	.000	.000	.000	.000
Competency	Pearson Correlation		.08	.070	.450**
	Sig. (2-tailed)		.432	.155	.000
Teaching style	Pearson Correlation			.32	.473**
	Sig. (2-tailed)			.041	.000
Assessment style	Pearson Correlation				.436**
	Sig. (2-tailed)				.000

Table 2 shows that professional ethics of faculty is the most important contributing factor that builds their image among students and this is followed by their teaching style and competency. The least correlated components were the assessment style, clearly indicating that a faculty member with command in their subject may not guaranty a good assessment from students. Regressions were done to see which factor most affects students' responses regarding faculty perceptions. Table 3 shows the results.

Table 3: Regression Analysis of Students' Perceptions

Model	Unstandardized Coefficients		Standardized Coefficients	Sig.
	B	Std. Error	Beta	
(Constant)	16	4.3	--	1.000
Professional Ethics	12	5.3	.41	.000
Competency	11	4.2	.30	.000
Teaching style	11	4.3	.18	.000
Assessment style	1.2	5.1	.05	.000

Table 3 reveals that professional ethics explain 40% of the variance of overall perceptions followed by teachers' competency which explains 30% of the variance. Assessment style is the least explaining factor which explains only 5% perceptions.

Furthermore t test was applied to see any difference between the students' perception grade wise to see whether the maturity and seniority of students determines regarding their judgment. T-test was applied to analyze the responses grade wise as students were studying in different grades and their time with teachers and their own maturity level might have affected their response pattern. Table 4 shows the t-test analysis.

Table 4: t Test Analysis of Students' Perceptions

	Mean		Std. deviation		T	Sig (2-tailed)	Effect size
	Junior	Senior	Junior	Senior			
Total	65.57	75.47	7	4	12.03	.000**	1.7

Table 4 shows that mean values of junior and senior level students varies with greater mean for the junior level students showing that they have better response level than their senior counterparts. Also standard deviation for junior students was higher which shows higher variability in their responses while senior level students were relatively stable with low standard deviation. t values show that this difference in mean is significant with effect size of 1.7 indicating a large difference.

Table 5: t-Test Analysis of Components of the Model

	mean		Std. deviation		T	Sig (2-tailed)	Effect size
	Junior	Senior	Junior	Senior			
Prof. Ethics	45	54	6	4	12.0	.000**	1.6
Competency	34	40	5	3	13.0	.021**	1.4
Teaching style	32	35	4	4	14	.251	0.75
Assessment style	33	36	4	4	13	.142	0.75

Table 5 indicates that seniors have better perceptions about their faculty with relatively greater mean scores than the juniors. However, the t test analysis reveals that this difference in mean score is only significant for professional ethics and competency only. The eta square values for professional ethics and competency shows that the mean difference has a large effect. These results confirm the overall findings of this work which consistently reveals that these two factor i.e professional ethics and competency are the most defining components of the model. The other two also have role in faculty perceptions but are not significant.

Table 6: t Test Analysis of Students Perceptions about Teacher's Personality (gender wise)

	Mean		Std. deviation		T	Sig (2-tailed)	Effect size
	male	female	male	female			
Total	63	55	7	3	11	.000**	1.4

Table 6 above reveals that mean values of male and female students varies with greater mean for the male students showing that they have better response level than their male counterparts. Also standard deviation for male students was higher thereby showing variability in their responses while female students were relatively stable with low standard deviation. Also t value shows that this difference in mean is significant.

Table 7: t-Test Analysis of Components of the Model

	Mean		Std. deviation		T	Sig (2-tailed)	Effect size
	male	female	Male	female			
Prof. Ethics	52	43	6	4	10.0	.000**	1.4
Competency	42	33	5	4	13.0	.021**	1.3
Teaching style	34	33	3	4	14	.251	0.70
Assessment style	35	32	3	4	13	.142	0.70

Table 7 indicates that males have better perceptions about their faculty with relatively greater mean scores than the females. However, t test analysis reveals that this difference in mean scores is only significant for professional ethics and competency only. The gender based results reveal that the two factors of professional ethics and competency are the most defining components of the model. The other two components also have role in student ratings but they are not significant.

Table 8: Interaction Effect with Educational Level and Gender Independent Variables

	Level Eta ²	Gender Eta ²	Level and gender Eta ²
Total	.027*	.051*	.006*
Prof. Ethics	.042*	.032*	.023*
Competency	.000	.106	.014*
Teaching style	.120	.160	.511
Assessment style	.106	.140	.621

* Interaction is significant at $P < .05$

**Interaction is significant at $p < .01$

Analysis of variance shows that there was significant interaction effect among two aspects of teachers' standards i.e. professional ethics and competency. Interaction was also significant for total perception between level of seniority of the students and gender. The results show that students at different seniority levels responded differently in relation to their gender.

DISCUSSION

The findings of this study confirm the proposed model regarding faculty perceptions of students. This is seen from the factor analysis which reveals that items used in faculty assessment are loading on their respective components. The correlations show that professional ethics is very important for a faculty member to be considered effective in the eyes of the students. Similarly, faculty competency is also significantly important. Regression results also confirm these findings which explain a major percentage of the image of faculty. These results are expected particularly in Arab culture where ethics is valued in society as a whole and for teachers in particular. Similarly competency being an important component also reflects local culture due to respect for knowledge in that society. A comparative picture of the junior and senior students reflects that senior students have more positive perception that differs significantly from their juniors; this may be due to their maturity level and possibility of spending more time with their faculty members.

Similarly a comparative picture of the male and female students reflects that male students have more positive perception that differs significantly from their female counterparts; this may be due to their interaction level as male students in the segregated Saudi Arabian culture have more interaction chances which give them more awareness about their teacher's personality. The ANOVA shows that compared to female students the males are influenced more by the level of seniority in the program and their perception varies more than their female counterparts.

The findings suggest that professional ethics should be given more importance in the disciplinary scenario of the universities in Saudi Arabia. Further, professional training is suggested for teachers to enhance their knowledge level and their teaching and assessment styles. This study was limited to a particular college therefore it can be reflected in other universities with the same model or improved version of the model as well that may include more constructs like classroom management skills, interpersonal skills etc.

Conclusion

The findings of this study lead to the recommendation that professional ethics should be given more importance in the disciplinary scenario of the universities in Saudi Arabia. Moreover, teachers should be provided an enabling environment to enhance their knowledge and professional training to improve their teaching and assessment style as well. This study was a pilot study of a representative college therefore it can be tested in other universities by including more constructs like classroom management skills and interpersonal skills as well as further cultural factors.

REFERENCES

- Al-Kuwaiti A Ahmed. 'Students evaluating teaching effectiveness process in Saudi Arabian medical colleges: A comparative study of students' and faculty members perception, Saudi Journal of Medicine and Medical Sciences, 2014 Vol. 2, Issue 3, 166-172.
- Bubb, D., Schraw, G., James, D., Brents, B., Kaalberg, K., Marchand, G., Amy, P., & Cammett, A. (2013). Making the case for formative assessment: How it improves student engagement and faculty summative course evaluations, *Assessment Update*, 25(3), 8-12.
- Berk. R.A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17, Number 1, 48-62.

- Clifford Nowell (2007). The Impact of Relative Grade Expectations on Student Evaluation of Teaching International Review of Economic Education, vol. 6, issue 2, 42-56
- Clifford Nowell, Lewis R. Gale and Bruce Handley (2010). Assessment & Evaluation in Higher Education Vol. 35, No. 4, 463–475.
- Cowie, B. and Bell, B. (1999). A Model of Formative Assessment in Science Education. Assessment in Education: Principles, Policy and Practice, Volume 6, Number 1, 101-116.
- CRLT (2015). Guidelines for evaluating teaching. The University of Michigan. <http://www.crlt.umich.edu/tstrategies/guidelines>
- d'Apollonia, S. & Abrami, P. C. (1997). Navigating student ratings of instruction, American Psychologist, 52(11), 1198–1208.
- Doyle, T. (2004). Evaluating teacher effectiveness-research summary. Learner Centered Teaching, Ferris State University.
- Goe, L., Bell, C., and Little, O. (2008). "Approaches to Evaluating Teacher Effectiveness: A Research Synthesis", National Comprehensive Center for Teacher Quality. June, p. 13.
- Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. American Psychologist, 52, 1209-1217.
- Hammond L. D. (2000). Teacher quality and educational achievement. Education Policy Archives; Vol 8. No. 1.
- Hodson, P. (2012). Learning and teaching workshop. Salman Bin Abdulaziz University, Saudi Arabia.
- Isley, P., and H. Singh. (2005). Do higher grades lead to favorable student evaluations? Journal of Economic Education 36, no. 2: 29–42.
- Marsh H.W. and Roche L.A. (1997). Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility. American Psychologist, Vol. 52, No.11, 1187-1197.
- McKeachie, W. J. (1997). Student ratings: The validity of use. American Psychologist, 52, 12 18 - 1225.
- MET Project, (2012). Asking Students about Teaching: Student Perception Surveys and Their Implementation, Bill & Melinda Gates Foundation. September, 2012, p. 8.
- Millea, M. and Grimes, P.W. (2002) Grade Expectations and Student Evaluation of Teaching. College Student Journal, 36, 582-590.
- Murray, H. G. (2005). Student Evaluation of Teaching: Has It Made a Difference? Department of Psychology. University of Western Ontario
- Nasser. F. & Fresko, B. (2002). Faculty view of student evaluation of college teaching, Assessment & Evaluation in Higher Education, 27(2), 187-198.
- Ortega G, T and T' Nia Crutchfield. (2014), Can Evaluating Teachers Improve Teacher Effectiveness? IMPACT Evaluations in Washington, DC.
- Simmons, T.L. (1998). Student evaluation of teachers, professional practice or punitive policy? Shiken: JALT testing & evaluation, SIG Newsletter, 1 (1).
- Sprague Joey and Kelley Massoni (2005). Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us. Sex Roles. Volume 53, Issue 11, 779-793.
- Stark-Wroblewski, K., Ahlering, R. F., & Brill, F. M. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: Supplementing student evaluations of teaching with pre-post learning measures. Assessment & Evaluation in Higher Education, 32, 403–415.