academicJournals

Vol. 8(8), pp. 378-389, 23 April, 2013
DOI: 10.5897/ERR2013.1154
ISSN 1996-0816 © 2013 Academic Journals
http://www.academicjournals.org/ERR

Educational Research and Reviews

Full Length Research Paper

Student evaluation of lecture and teaching effectiveness in higher education

Kenan Özcan

Department of Educational Sciences, Adiyaman University Education Faculty, Altinşehir 02040, Adiyaman, Turkey.

Accepted 25 March, 2013

As good teachers may have great influence on positive outcomes of students, educational systems should provide feedback about their professional performances in any way. Otherwise, not only do teachers fail but the system fails. It is claimed that students have some reasons while they are evaluating the lecture and teaching. This study was conducted to discover what students' reasons are while evaluating the lecture and teaching. Results reveal that students from colleges of education at developing universities evaluated lecture and teaching more favorably compared to students at developing and newly-established universities. Another result shows that students with higher Grade Point Averages (GPAs) tend to evaluate lecture and teachers more favorably. Finally, significance was discovered between students' views in terms of newly-established universities and developing universities, classroom size, their grades and students' GPAs.

Key words: College of education, teacher, student, course evaluation, teacher evaluation, higher education.

INTRODUCTION

If educational systems are unable to provide feedback to their teachers, not only do they fail teachers, in the end they also fail students. As good teachers may have great influence on positive outcomes of students, supporting ongoing growth and development should be a priority for educational systems. Without appropriate assessments to identify problems about lecture and teaching in educational processes, resources spent on teacher's education become a waste of time for educational systems. In this respect, university educators place great influence on the affective skill in the admission and education of their students. However, faculty may not sometimes perform their professional responsibilities properly by conducting their classes with no or little prior preparation, showing some unethical attitudes towards students, offending their students, threatening their students to give lower grades and discriminating them according to gender, race and in some other ways. At the same time, they may cause some problems by not creating a free atmosphere for students to learn and assessing them unfairly. For that reason, it is believed that student evaluation of lecture and teaching effectiveness process may prevent students from all these problems and keep the quality of education high. It also helps increase it to the desired level.

Research that student evaluations of teaching effectiveness are positively related to grades goes back to at least the 1950s (Anikeeff, 1953), but heightened interest in the validity of teaching evaluations began in earnest in the 1970s, with researchers systematically manipulating grades to see the effect on evaluations (Holmes, 1972) and conducting several correlational studies on the topic (Gessner, 1973). Over the last several decades, research on student evaluation of teaching effectiveness has considered a variety of variables that play a role in influencing the process of student evaluation (Addison and Best, 2006). In general, these variables are classroom size or required/elective status (Kulik and Kulik,

E-mail: kozcan04@hotmail.com.

1974; McKeachie, 1997), instructor variables such as expertise (Marsh, 1980; Marsh and Roche, 1997), personality characteristics (Best and Addison, 2000), or nonverbal behavior (Babad et al., 2004), and finally, student variables like age, gender, grades and grade point averages (GPAs).

University tradition may play an important role in this process. Among academic institutions, newly-established universities whose establishment dates are between 1 to 5 years. Some are called developing universities whose establishment dates goes back between 6 to 10 years, and finally developed ones which were established more than 20 years ago. In this regard, while newly established universities scarce qualified academics, established ones have many qualified ones. Their university tradition, physical conditions, technology or other environmental issues can also affect quality of instruction. It is thought that they all determine students' lecture and teaching evaluation decisions.

Despite the fact that colleges and universities place importance on student evaluations of teaching, it is well established that students who receive higher grades rate professors more favorably (Cohen, 1981; Feldman, 1976; Goldberg and Callahan, 1991). This is of some concern, because colleges evaluate faculty members on these evaluations and because research indicates that students use such evaluations when they are selecting courses (Wilhelm, 2004). Perhaps the most appealing reason for giving higher grades is because students learn more. Thus, it is not surprising that faculty would receive higher course evaluations (Marsh and Roche, 1997; Marsh et al., 1997).

It is assumed that students who receive grades they perceive as fair will rate instructors more positively (Cherry et al., 2003; Tata, 1999). For example, in some studies, it was found that students who receive lower than expected grades rated the instructors less favorably (Holme, 1972; Vasta and Sarmiento, 1979; Blunt, 1991; Aleamoni, 1999; Wachtel, 1998; Seiver, 1983). Moreover, consistent with this finding, research indicates that class attendance is positively related to higher course evaluations. Here it is claimed that students who attend class regularly are likely to learn more, especially in higher education (Davidovitch and Soen, 2006). It was found in some studies that at the undergraduate level, those who have the highest on measures of achievement striving were less likely to give higher evaluations to teachers (Bacon and Novotny, 2002; Marsh, 1987). Another possibility is that students who are highly motivated to achieve in a course may work harder and achieve higher grades. Their success may lead them to view the faculty member as very effective. It was reported in some studies that student interest in the course is related to evaluation of the value of the course and stimulation felt by a course affected teaching evaluations (Addison and Best, 2006; Heckert et al., 2006a; Remedios and Lieberman, 2008). Students may also attribute their success and high grades to instructional effectiveness. Attributional theories suggest that students' grades may provide students with information about how effective the course is and how much they like it (Gessner, 1973; Addison and Best, 2006). Conversely, those who are pessimistic about their future grades are more likely to negatively evaluate the course (Millea and Grimes, 2002). This result is in keeping with the self-serving effect that students externalize responsibility for negative outcomes. In other words, poor grades must be the fault of the course or the faculty member (Marsh, 1986).

Similarly, in a study students were randomly assigned to receive feedback that they did good, satisfactory or poor on their performance on a test. Although students were told that these grades would have no bearing on their class grades, the manipulated grades reported to the students were predictive of instructor ratings (Worthington and Wong, 1979). However, another study did not find course difficulty to be a predictor of student evaluations (Millea and Grimes, 2002). Furthermore, it was found that both students' efforts and perceptions of the appropriateness of the difficulty of the course were positively related to lecture evaluations (Heckert et al., 2006b). There is also support for non-teaching-related variables to be related to effectiveness. Both the youth and extraversion of faculty have been identified as positively related to course evaluations (Radmacher and Martin, 2001).

Finally, students may simply like easy courses and punish those that they perceive as difficult by giving low evaluations. In a survey of college students about why they gave poor teaching evaluations, 8% of students reported giving low evaluations for revenge. In the same survey, being unfair in grading or hard grading was the second most common reason given for poor evaluations, behind only poor teaching style or methods (Boysen, 2008; Clayson, 2004; Clayson et al., 2006).

Moreover, the student's understanding of the evaluation process could influence evaluations as well. It was found that students view formative uses of evaluation as being far more important than summative uses of evaluation (Chen and Hoshower, 2003). Furthermore, it was discovered that students adopt different cognitive schemas, with resulting differences in evaluations, when these different purposes are used as the basis for evaluation (Young et al., 1999).

In a further extension of this argument, it was indicated that students compare the grades they receive in one class with grades received in other courses to evaluate the leniency-or lack thereof-in a particular course (Chambers and Schmitt, 2002). College teachers commonly believe grade expectation to be a very strong influence in student evaluation (Baldwin and Blattner, 2003). Such beliefs might underlie the relatively wide spread distrust of student evaluations of faculty by some of their intended users, namely the faculty themselves

(Nasser and Fresko, 2002).

Taking all these into consideration, it is evident that students evaluate lecture and teaching with some reasons. In this respect, they commonly evaluate lectures with some personal and academic reasons. However, it is a matter of question how they evaluate lecture and teaching at higher education.

The purpose

Educational institutions and teachers carry out responsibility for educating and training of the individuals of any society according to the needs of the new age. For this reason, teachers have critical roles in creating manpower which will determine competitiveness in the global market based on information and knowledge. Therefore, teacher education becomes more important.

Colleges of education and teachers have significant responsibilities in this respect. However, faculty professional behaviors (communication skills, motivating students, knowledge of subject matter, contemporary teaching strategies, classroom management skills, technology literacy etc.) and the classes they conduct (appropriateness to the goals, drawing students attention etc.) are important in teacher education. This study purposes to discover student evaluation to the course. For this purpose the answers of the following questions were researched:

When student evaluation of lecture and teaching is concerned, are there any differences between students from the newly-established and developing colleges? In Turkey, The Higher Education Council (YÖK) determined newly-established and developing colleges. In this regard, colleges that have been educating teachers less than 10 years while those which are in practice more than 10 are accepted as developing ones.

1. Are there any differences between students' views in terms of classroom size and student evaluation of lecture and teaching?
2. Are there any differences between students' views in terms of students' academic success and student evaluation of lecture and teaching?
3. Are there any differences between students' views in terms of students' grades and student evaluation of lecture and teaching?

By conducting this research, it is purposed to increase awareness of administrators, planners, coordinators of higher education in Turkey in terms of student evaluation of lecture and teaching. The results and recommendation to be provided here may help these people and institutions to revise the process.

METHODOLOGY

This study proposes to find out how students evaluate lecture and teaching effectiveness. Therefore, it employed a survey method. The data were collected through a scale called "Students' Perceptions about Lecture and Teaching Evaluation Scale" with a written permission of the developers (Norvilitis and Zhang, 2009).

Study group

The sampling of this study was determined through a purposive sampling method.

This type of sampling is particularly relevant when you are concerned with exploring the universe and understanding the audience and when the desired population for the study is rare or very difficult to locate and recruit for a study. In this method, researchers can use their prior knowledge to choose respondents (Bailey, 1994). This study was conducted in 2012 Spring Term with three student groups.

In the first group, in the quantitative part, 290 students participated in the reliability and validity analysis process. Here, the data were obtained and analyzed for exploratory factor analysis (EFA). They were from Colleges of Education from Adiyaman University (26.2%), Cumhuriyet University (32.4%), Sakarya University (13.3%) and ArtvinÇoruh University (28.1%). Of these participants, 47.8% were female and 52.2% were male. In addition, 22.6% of the participants were freshmen, 34.3% second, 19.2% third and 23.9% senior grades.

In the second group, 228 students participated in the validity, reliability, and adaptation process of the scale. They were from Colleges of Education from Adiyaman University (34.5%), Sakarya University (34.1%), and Inönü University (31.4%). Of these participants, 49.7% were female and 50.3% were male. In addition, 23.8% of the participants were freshmen, 36.2% second, 18.3% third and 21.7% senior grades.

In the third group, 1440 students participated in the research. Participants were from Colleges of Education from Adiyaman University (18.1%), MuşAlpaslan University (17.3%), Cumhuriyet University (15.1%), Artvin Çoruh University (12.8%), Mersin University (13.4%), Sakarya University (12.7%) and Inönü University (10.6%). Of the participating students, 20.1% were from Primary School Teaching, 16.9% Science and Technology Teaching, 15.9% Psychological Counseling and Guidance, 11.6% Pre-School Teaching, 11.5% Social Sciences Teaching, 9.7% Turkish Teaching, 8.7% Department of Religious Studies and Ethics Education, 4.2% Mathematics Teaching, and 2.1% from other departments. Of these participants, 58.9% were female and 41.1% were male, 32.2% of the participants were freshmen, 27.2% second, 18.2% third and 22.5% senior grades. When classroom size is concerned, 3.5% were from classes that had less than 30 students, 19.0% from a class of 31 to 40 students, 32.8% from a class that had between 41 to 50 students, and 44.7% from a class that had over 51 students. Concerning their GPAs, 4.2% had E/F, 20.5% D GPAs, 34.7% C, 31.7% D, and 4.8% A.

Data collection instrument

Scale adaptation process

In order to discover how students evaluate lecture and teaching effectiveness, we used the "Students' Perceptions about Lecture and Teaching Evaluation Scale" (Norvilitis and Zhang, 2009). Items are rated on a five-point Likert scale ranging from "*Strongly Disagree"* to *"Strongly Agree*" with a neutral midpoint of 3. The scale has two dimensions: Lecture Evaluation (13 items) and Teaching Evaluation (14 items). In the original scale, Cronbach's Alpha Coefficient was 0.94 for the Lecture Evaluation and 0.93 for the Teaching Evaluation subscales. Higher scores show positive behaviors; lower scores indicate negative ones (Norvilitis and

Zhang, 2009).

In the adaptation process, this study used the back-translation method, which is the preferred method although it can be time consuming and expensive. In this method, a questionnaire is translated into the target language by one translator and then translated back into the source language by an independent translator who is blinded to the original questionnaire. The two source-language versions are then compared (Bailey, 1994; Sperber, 2004; Looman and Farrag, 2009). It was proofread by 60 students. The language was understandable; thus, the scale was valid and usable in Turkish.

Exploratory factor analysis (EFA)

Factor analysis, a complex, multi-step and broadly applied statistical technique (Costello and Osborne, 2005), is an approach for expressing hypothetical constructs in the language of mathematics by using a variety of observable indicators that can be directly measured. The analysis is considered exploratory when determining how many constructs (factors) is needed to explain the relationships among the observed indicators and confirmatory when a preexisting model of the relationship among the indicators directs the search (Raykov and Marcoulides, 2000). Exploratory analysis allows the exploration of empirical data for characteristic features and interesting relationships without imposing any definite model on the data (Jörreskog and Sörbom, 1993).

Exploratory factor analysis (EFA) was used to test validity. Sample size plays an important role in examining statistical techniques (Raykov and Marcoulides, 2000). Sample groups (290 students respectively) were used for EFA analysis. An acceptable sample proportion for AFA is 4:1 for each item (Floyd and Widaman, 1995). While there is some debate about optimum sample size (Gorsuch, 1983; Kline, 1979; Guilford, 1954; Cattell, 1978; Comrey and Lee, 1992 cited in: MacCallum and Widaman, 1999; Hoyle, 1995, Schermelleh-Engel et al., 2003), this study used a sample scale of 10:1 for EFA for each item.

In order to test the compatibility of the data for factor analysis, Kaiser-Meyer-Olkin (KMO) and Barlett Sphericity tests were used. The KMO statistic variables are accepted to be greater than 0.50. Furthermore, values between 0.50 and 0.70 are mediocre, between 0.70 and 0.80 are good, between 0.80 and 0.90 are great and above 0.90 are superb (Hutcheson and Sofroniou, 1999). For these data the values are about EFA .92, which fall into the great range. Bartlett's (1954) test of sphericity is a notoriously sensitive test of the hypothesis that the correlations in a correlation matrix are zero. The test is available in SPSS factor, but because of its sensitivity and its dependence on sample size the test is likely to be significant with samples of substantial size even if correlations are very low. Therefore, use of the test is recommended only if there are fewer than, say, five cases per variable (Tabachnick and Fidell, 2007). For these data, Bartlett's test is highly significant for EFA ($X^2_{(253)}$ =2846.63; p<.01). The data showed the multivariate normal distribution.

In order to reveal the Factor design of the scale, Principal Components analysis and Varimax Rotated Component Matrix was chosen as the factor analysis. In applied social science research, orthogonal rotation is used most often, perhaps because it is the default in major statistical programs such as SPSS (varimax rotation), and the perception that orthogonally rotated solutions are more easily interpreted because the factor loadings represent correlations between the indicators and the latent factors (e.g., squaring the factor loading provides the proportion of variance in the indicator that the factor solution explains) (Brown, 2006). Analysis showed that 23 items with the eigenvalue above value 1, has the same factor distribution as the original scale.

As for the items factor loadings, 0.50 is an acceptable level. The magnitude of the factor loading must be at least 0.30 (Barnes et al, 2001). As a rule of thumb, only variables with loadings of 0.32 and above are interpreted. The greater the loading, the more the variable is a pure measure of the factor. Comrey and Lee (1992) suggest that loadings in excess of 0.71 (50% overlapping variance) are considered excellent, 0.63 (40% overlapping variance) very good, .55 (30% overlapping variance) good, .45 (20% overlapping variance) fair, and 0.32 (10% overlapping variance) poor. Choice of the cutoff for size of loading to be interpreted is a matter of researcher preference. Sometimes there is a gap in loadings across the factors and, if the cutoff is in the gap, it is easy to specify which variables load and which do not. Other times the cutoff is selected because one can interpret factors with that cutoff but not with a lower cutoff (Comrey and Lee, 1992; Cited in: Tabachnick and Fidell; 2007). Analysis with regards to Factor design, item factor loadings, total factor variance and item analysis are shown in Table 1.

Results analysis showed that items appear under the dimensions defined in the theoretical context. Factor loadings for each sub scale is as follows: the teaching evaluation (TE) sub-dimension between0 .51 and 0.64, and the lecture evaluation (LE) sub-dimension between 0.55 and 0.72 (Table 1).

There is no relation with the sample and the factor loadings. Guadagnoli and Velicer (1988) challenged such rules and argued that no sound theoretical or empirical basis exists for across-the-board participant-to-variable ratio recommendations. Instead, their Monte Carlo study suggests that variable saturation with the factors, indicated by the size of the factor loadings along with the total sample size and the number of indicators per factor was important in determining the stability of factor solutions. Most notably, with factor loadings of 0.80, solutions were highly stable across replicated samples regardless of the number of indicators, even with as few as 50 participants. When factor loadings were in the 0.60 range, stable solutions were obtained with sample sizes greater than 150, or with still smaller samples when each component contained at least four variables loading at 0.60. In general, larger samples of 300 to 400 were needed when the factor loadings were only 0.40 (Floyd and Widaman, 1995).

Reliability and validity analysis of "Students' Perceptions about Lecture and Teaching Evaluation Scale" showed that sub dimensions were divided under two sub dimensions with 23 items. The distribution of the sub scales are as follows: ten items TE, thirteen items in LE. The latest version of the scale showed that the lowest value is 23 and the highest value is 115.

During the adaptation process, error variances of the items 9, 11 and 12 were high in "Lecture Evaluation" and item 27 in "Teaching Evaluation" sub-dimensions, decreasing reliability coefficient values. Thus, they were removed from the scale.

Students' Perceptions about Teaching and Lecture Evaluation Scale sub dimension correlation is between 0.45 and 0.78, total correlation among the sub scales is between 0.27 and 0.76. and the correlation among sub-scales is 0.60. In empirical studies, a correlation value among the factors equal to 0.85 or less is desirable (Brown, 2006). Analysis shows that 23 items fall into two dimensions (eigenvalue=1), explaining the 45.27% of the variance. In social sciences, a range of 40 to 60% change in variants rates is accepted as sufficient (Scherer et al., 1988), and an explained variance of 60%, and sometimes less, is acceptable (Vieira, 2011). Eigenvalue and variance scores the first factor is determined as 6.35 to 27.61%, and the factor two is accepted as 4.06 to 17.67%.

Confirmatory factor analysis (CFA)

CFA is not concerned with discovering a factor structure, but with confirming the existence of a specific factor structure. In this respect, CFA is considered to be a general modeling approach that is designed to test hypotheses about a factor structure whose number and interpretation are given in advance (Raykov and

Table 1. Confirmatory factor analysis of students' perceptions about lecture and teaching evaluation scale in higher education. Figure of factor (vertical rotated -varimax) and items analysis.

Items	Factor Design and Item Loadings		Factor Common Variance	Item Analysis								
				Correlation		Item Discrimination Feature					Reliability	
				Item Fcator	Item Total	Sub %27 (n=78)		Top %27 (n=78)			Total (α= 0.92)	
											Items (α)	Sub Scale (α)
	TE	LE	(h^2)	r*	r*	$\bar{X}$	Sd	$\bar{X}$	Sd	t*		
D1	0.14	0.59	0.37	0.68	0.62	2.18	1.05	4.01	0.71	-12.74	0.79	
D2	0.19	0.59	0.38	0.69	0.65	2.15	0.85	3.90	0.69	-13.99	0.79	
D3	0.33	0.55	0.41	0.65	0.59	2.27	0.96	3.99	0.65	-13.03	0.79	
D4	0.20	0.64	0.45	0.70	0.62	2.18	0.92	3.96	0.78	-12.38	0.79	
D5	0.08	0.52	0.28	0.54	0.42	2.12	1.16	3.32	1.13	-6.56	0.81	0.82
D6	-0.12	0.51	0.27	0.45	0.27	2.40	1.28	3.10	1.24	-3.49	0.82	
D7	0.17	0.59	0.37	0.59	0.50	2.35	1.16	3.67	0.78	-8.33	0.80	
D8	0.15	0.60	0.38	0.60	0.50	2.33	1.15	3.64	0.88	-7.98	0.80	
D9	0.11	0.56	0.33	0.60	0.52	2.19	1.07	3.60	0.90	-8.90	0.80	
D10	0.21	0.62	0.42	0.69	0.63	2.10	1.03	3.91	0.78	-12.40	0.79	
D11	0.66	0.29	0.52	0.73	0.71	2.05	0.97	3.97	0.76	-13.85	0.79	
D12	0.70	0.26	0.56	0.75	0.71	2.01	0.86	4.04	0.80	-15.26	0.79	
D13	0.71	0.02	0.50	0.67	0.57	2.72	1.18	4.14	0.75	-8.97	0.79	
D14	0.72	0.05	0.51	0.69	0.60	2.60	1.12	4.27	0.66	-11.33	0.79	
D15	0.73	0.22	0.58	0.74	0.70	2.45	1.10	4.35	0.55	-15.36	0.79	
D16	0.64	0.21	0.45	0.71	0.69	1.96	1.02	4.04	0.61	-15.36	0.81	0.91
D17	0.71	0.25	0.57	0.78	0.76	2.05	0.92	4.08	0.62	-16.08	0.82	
D18	0.61	0.19	0.40	0.65	0.59	2.54	1.12	4.18	0.62	-11.29	0.80	
D19	0.64	0.27	0.48	0.69	0.66	2.55	1.06	4.17	0.61	-11.62	0.80	
D20	0.56	0.12	0.33	0.59	0.53	2.36	1.02	3.72	0.84	-9.11	0.80	
D21	0.63	0.23	0.45	0.72	0.69	2.05	0.98	4.03	0.62	-15.02	0.79	
D22	0.71	0.13	0.52	0.74	0.69	2.19	0.97	4.09	0.74	-13.74	0.79	
D23	0.55	0.12	0.32	0.58	0.53	2.45	1.20	4.10	0.75	-10.31	0.79	

*p<.01. Note: TE: Teaching Evaluation Sub Dimensions, LE: Lecture Evaluation Sub Dimensions

Marcoulides, 2000). In CFA one builds a model assumed to describe, explain, or account for the empirical data in terms of relatively few parameters (Jörreskog and Sörbom, 1993). The technique of CFA analyzes a priori measurement models in which both the number of factors and their correspondence with the indicators are explicitly specified (Kline, 2011). The goodness fit indexes of the items according to CFA are presented in Table 2.

As seen in Table 2, the scale was analyzed with CFA for two correlated dimensions, corresponding to the two subscales in the measure. The most frequently used statistics regarding CFA and model-data are chi-square (χ^2), RMSEA, NFI, NNFI, CFI, GFI, and AGFI. According to those indices, Chi-square test is significant (χ^2= 436.34, df= 228) and the ratio chi-square/degrees of freedom is below 2 (χ^2/df = 1.91). The overall model fit statistics in LISREL are within the generally accepted thresholds and suggest an acceptable goodness-of-fit (Cote et al., 2001; Vieira, 2011; Hooper et al., 2008; Brown, 2006; Schreiber et al., 2006; Schermelleh-Engel et al., 2003; MacCallum et al., 1996; Hu and Bentler, 1999; Baumgartner and Homburg, 1996). Individual item loadings of the students' perceptions about lecture and teaching evaluation scale are shown in Figure 1.

The correlation among the sub dimensions was 0.72. This moderate correlation among sub-scales shows that each subscale exists in the scale construct as well. With multi-collinearity, the variables are very highly correlated (say, 0.90 and above) (Tabachnick and Fidell, 2007, p.88). In order to decrease error variance and increase fit indexes between items 5and 6, a modification was done in the lecture evaluation sub-dimension.

Reliability

After conducting CFA analysis, the reliability coefficient of the new scale is 0.83 for the lecture evaluation sub-dimension and 0.91 for the teaching evaluation sub-dimension the reliability coefficient of the scale is as .89. Generally, reliability coefficients values around 0.70 were "adequate," 0.80 were "very good," and over 0.90 "excellent" (Kline, 2011, p. 70). Therefore, it can be said that the

Table 2. Fit Indices of the items in the Students' Perceptions about Lecture and Teaching Evaluation

Fit index	Acceptable fit index	Suggested Model (n=290)
χ^2/df	0<χ^2/df<3	436.34 /228= 1.91
RMSEA	.00 ≤ RMSEA ≤ .10	.06
RMR	.00 ≤ RMR ≤ .10	.07
SMR	.00 ≤ SMR ≤ .10	.05
NFI	.90 ≤ NFI≤ 1.00	.97
NNFI	.95 ≤ NNFI ≤ 1.00	.97
CFI	.90 ≤ CFI ≤ 1.00	.97
GFI	.90 ≤ GFI ≤ 1.00	.90
AGFI	.80≤ AGFI ≤ .90	.82
PGFI	.00 ≤ PGFI ≤ .95	.71

reliability coefficient is very good for this study.

Data analysis

In order to analyze the data, SPSS 15.0 and Lisrel 8.80 were used. For demographic variables of students like class, department, class size, and GPA, percentage (%), frequency (f), and Crosstabs analysis were done. For significance statistics and dual comparisons, t-test, for multi-comparisons, one-way ANOVA was made and significance level was accepted as *p*<.05 for both tests. In all one-way ANOVAs, "Test of Homogeneity of Variances" prerequisite was provided (*p*>.05). For significance of the findings among groups, Tukey multiple comparison test was used.

FINDINGS

In this part, the findings obtained with "Lecture Evaluation and Teaching Evaluation Scale" were given here in terms of some variables. Frequency, percent, mean and standard deviation distributions are presented in Table 3.

As seen in Table 3, according to lecture evaluation subscale, students strongly disagree and disagree that course objectives are clear (41.6%) and course is well organized (42.2%), assigned workload is appropriate for credit hours (44.9%) and exams are good measures of their knowledge and understanding or ability to perform (44.9%). However, students strongly agree and agree that their responsibilities are clearly defined (40.6 %) and course content is relevant and useful (43.8%).

As far as teaching evaluation is concerned, students strongly disagree and disagree that teachers are enthusiastic about the class (44.8%), available to students during regular and reasonable office hours (39.8%) and motivate them by their examples to learn about the subject (39.9%).

However, students strongly agree and agree that their teachers are well prepared for the class (40.0%), they make good use of class time (41.9%), give clear examples and explanations (43.4%), and they are genuinely interested in helping their students understand the subject (44.2%). On the other hand, students disagree that teachers respond respectfully to their students' questions and viewpoints (43.5%) and start/ dismiss class at scheduled times (43.9%). The findings obtained from the students from developed and developing universities on lecture and teaching evaluation are presented in Table 4.

There are significant differences among the students' views of lecture evaluation and also teacher evaluation in terms of establishment dates of colleges. According to this, students from developed colleges ($\bar{X}$ = 2.95) consider that their lectures are better conducted than those from newly-established colleges or developing ones ($\bar{X}$ = 2.86). As far as lecture evaluation is concerned, students from developed colleges also evaluate lecture better ($\bar{X}$ = 3.05) than those from newly-established ones ($\bar{X}$ = 2.92). The findings obtained about students views on classroom size and teacher and lecture evaluation are presented in Table 5.

In Table 5 significant difference was discovered between students' views and lecture evaluation [$F_{(3-1436)}$ = 4.19, p<.01] in terms of class size. In order to determine the source of difference, Tukey multi-comparison test was used and according to that, those who are from classes with 30 or lower students evaluate lecture better ($\bar{X}$ =3.20) than 41 to 50 students ($\bar{X}$ = 2.86) and 51 and over students ($\bar{X}$ = 2.91) (p<.05). There is also significant difference between students' views and teaching evaluation [$F_{(3-1436)}$ = 4.01, p<.05] in terms of class size. Tukey multi-comparison test was used to determine the source of difference and it was discovered that students from classes with 30 students or lower evaluate teaching better ($\bar{X}$ = 3.23), comparing those from classes with 51 and over students ($\bar{X}$ = 2.97) (p<.05). The findings related to GPAs (Grand Point Average) and teaching and lecture evaluation were presented in Table 6.

In the Table 6 significant difference was also found between students' views and lecture evaluation [$F_{(4-1435)}$ = 4.63, p<.01] in terms of their GPAs. According to that, students with 86 and over GPAs ($\bar{X}$ = 3.10) evaluate their lecture more favorably comparing those with 66-75 GPAs ($\bar{X}$ = 2.86), 51 to 65 GPAs ($\bar{X}$ =2.83) and 50 and below GPAs ($\bar{X}$ = 2.73) (p<.05).

Significant difference was also discovered between students' views and teaching evaluation [$F_{(4-1435)}$ = 2.96, p<.05] in terms of their GPAs. According to that, students with 86 and over GPAs ($\bar{X}$ =3.09), evaluate teaching more favorably comparing those who with 51 to 65 GPAs ($\bar{X}$ =3.75) and 50 and below GPAs ($\bar{X}$ =3.09) (p<.05).The students' perceptions related to teaching and lecture evaluations were given in Table 7.

In the Table 7, there is also significant difference between students' views and lecture evaluation [$F_{(3-1436)}$ = 4.97, p<.01] in terms of their grades. According to that

Table 3. Frequency, percent, mean and standard deviation distributions of Students' Perceptions about lecture and teaching evaluation

Items Rank	Students' Perception about Lecture/course and Teaching Evaluation at Higher Education (N=1440)	Strongly Disagree and Disagree		Agree and Strongly Agree			
		f	%	*f*	%	$\overline{X}$	SD
Lecture Evaluation							
D1	Course objectives are clear.	598	41.6	535	37.2	2.91	1.20
D 2	Course is well organized.	607	42.2	506	35.2	2.90	1.16
D 3	Student responsibilities are clearly defined.	534	37.1	585	40.6	3.02	1.20
D 4	Course content is relevant and useful.	508	35.2	631	43.8	3.06	1.19
D 5	Assigned workload is appropriate for credit hours.	647	44.9	525	36.5	2.83	1.27
D 6	Assigned homework is NOT just busywork.	682	47.3	504	35	2.78	1.31
D 7	Test(s) and other materials have helped me understand course topics.	542	37.7	593	41.2	2.98	1.24
D 8	Exams concentrate on important points of the course.	570	39.6	553	38.4	2.91	1.22
D9	Exams are good measures of my knowledge, understanding or ability to perform	646	44.9	460	31.9	2.78	1.23
D10	Course as a whole has produced new knowledge, skills, and awareness in me.	556	38.7	536	37.2	2.93	1.18
Teaching Evaluation							
D11	Has an excellent knowledge of the subject matter.	594	41.3	446	31	2.84	1.16
D12	Is enthusiastic about the class.	645	44.8	434	30.1	2.77	1.17
D13	Is well prepared for the class.	483	33.6	576	40.0	3.03	1.14
D14	Makes good use of class time.	485	33.7	604	41.9	3.06	1.17
D15	Gives clear examples and explanations.	456	31.7	625	43.4	3.12	1.16
D16	Makes helpful evaluations of my work (e.g., papers, exams).	558	38.7	523	36.4	2.92	1.19
D17	Clearly explains difficult concepts, ideas, or theories.	533	37	553	38.4	2.98	1.15
D18	Responds respectfully to student questions and viewpoints.	494	34.3	626	43.5	3.07	1.22
D19	Is genuinely interested in helping me understand the subject.	456	31.7	636	44.2	3.11	1.14
D20	Is available to students during regular and reasonable office hours.	573	39.8	508	35.3	2.89	1.21
D21	Motivates me by his/her example to what to learn about the subject	575	39.9	475	33.0	2.87	1.18
D22	Has produced new knowledge, skills, and awareness in me.	547	38	544	37.8	2.96	1.16
D23	Starts/dismisses class at scheduled times.	538	37.4	632	43.9	3.06	1.26

freshmen ($\overline{X}$ = 2.95) evaluate their lectures more favorably comparing second class ($\overline{X}$ = 2.92), the senior class ($\overline{X}$ = 2.78) (p<.5). Significant difference was also discovered between students' views and teaching evaluation [$F_{(3\text{-}1436)}$ = 8.45, p<.001] regarding their grades. This shows that freshmen evaluate teaching more favorably ($\overline{X}$ = 3.03) comparing students from second class ($\overline{X}$ = 3.01), third class ($\overline{X}$ = 2.98) and the senior class ($\overline{X}$ = 2.81) (p<.5).

DISCUSSION, CONCLUSION AND SUGGESTIONS

This study was conducted to discover how students evaluate courses and teaching effectiveness and a number of results were obtained. According to the results, students from colleges of education at developed universities evaluate their lectures and teachers more favorably comparing those from newly-established colleges. It is considered that the experience of a college has an influence on meeting students' expectations in

Table 4. T-test results of Establishment Date and Lecture/Teaching Evaluation Variables

Sub-dimensions	Establishment Dates of Colleges	*n*	$\bar{X}$	*SD*	*t*	Df	*p*
Lecture Evaluation	Developing	677	2.86	0.65	-2.65	1438	.008*
	Developed	763	2.95	0.68			
Teacher Evaluation	Developing	677	2.92	0.95	-3.29	1438	.001**
	Developed	763	3.05	0.96			

Table 5. One way ANOVA analysis Results of Classroom Size and Teaching and Lecture Evaluation

Sub-dimensions	Class Size	n	$\bar{X}$	SD	F	p
Lecture Evaluation	Lowerthan 30	47	3.20	0.72	4.20	.006*
	Between 31-40	274	2.95	0.73		
	Between 41-50	473	2.86	0.65		
	51 andover	646	2.91	0.66		
Teacher Evaluation	Lowerthan 30	47	3.23	0.90	3.04	.029*
	Between 31-40	274	3.06	0.80		
	Between 41-50	473	2.95	0.77		
	51 andover	646	2.97	0.72		

Table 6. One Way ANOVA Results concerning students' GPAs and Teaching and Lecture Evaluation

Sub-dimensions	GPAs	n	$\bar{X}$	SD	F	p
Lecture Evaluation	A	127	3.10	0.70	4.63	.001*
	B	458	2.94	0.65		
	C	500	2.86	0.67		
	D	295	2.90	0.70		
	E/F	60	2.73	0.61		
Teacher Evaluation	A	127	3.09	0.54	2.96	.019*
	B	458	3.00	0.50		
	C	500	2.95	0.51		
	D	295	3.04	0.54		
	E/F	60	2.73	0.47		

Table 7. One Way ANOVA Analysis Results Concerning Class and Teaching and Lecture Evaluation

Sub-dimensions	Grades	n	$\bar{X}$	SD	F	p
Lecture Evaluation	Freshmen	463	2.95	0.68	4.20	.002*
	Second class	391	2.92	0.65		
	Third class	262	2.90	0.67		
	Senior class	324	2.78	0.68		
Teacher Evaluation	Freshmen	463	3.01	0.73	3.04	.001**
	Second class	391	3.03	0.76		
	Third class	262	2.98	0.78		
	Senior class	324	2.81	0.76		

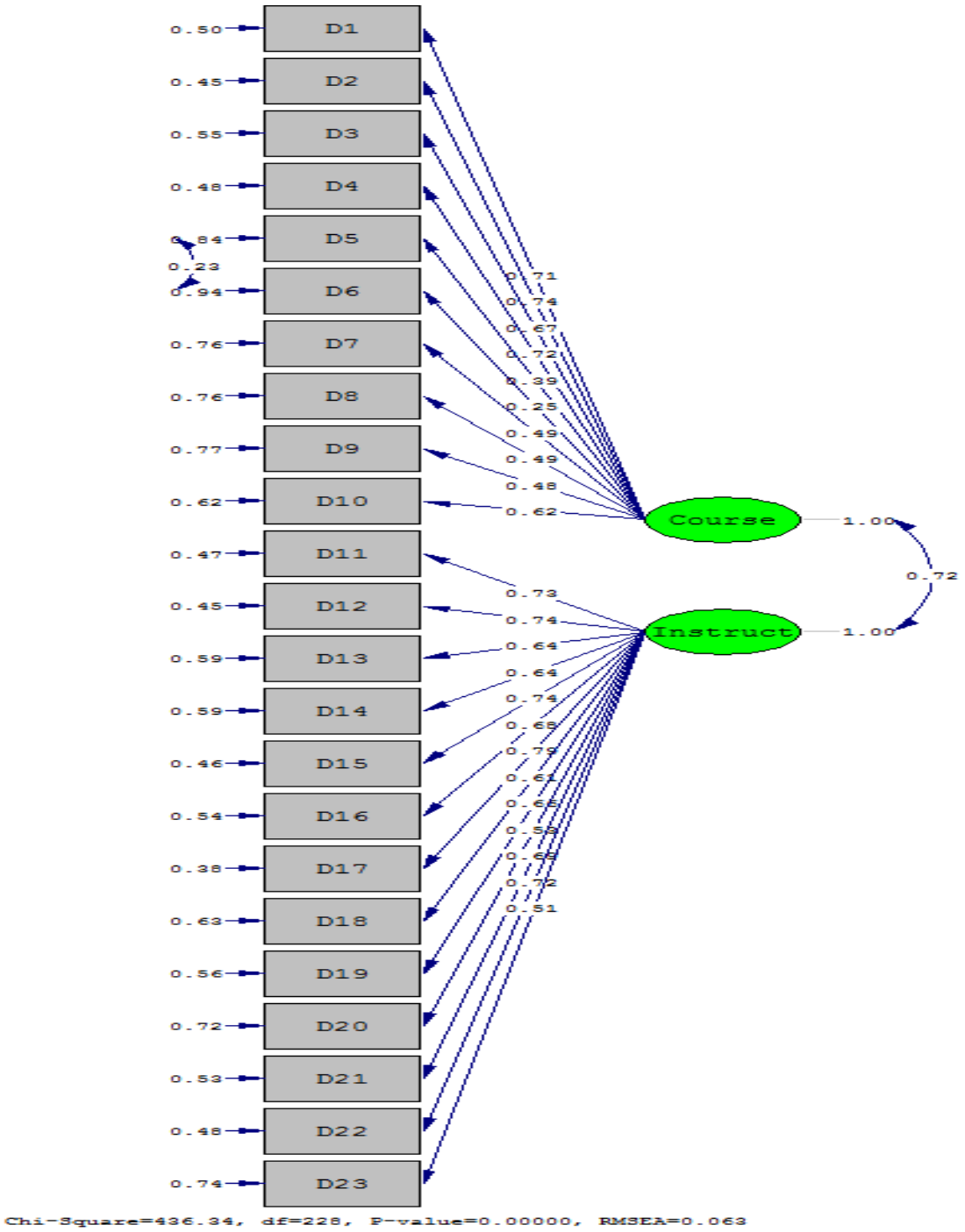

Figure 1. Significance Level Rates of Two-Dimension Model Variables of Students' Perceptions about Lecture and Teaching Evaluation Scale

establishing institutional culture, training and employing qualified academicians according to the needs of organization. However, in newly established colleges, it is known that are some problems like employing qualified academic staff, which is met by instructors most of the time, providing basic physical conditions, library and technological infrastructure. What is more, academicians also have a heavy workload, which is approximately 20 teaching hours therefore; they may not have enough time to conduct researches and prepare supplementary material for their students. Another reason may be weak institutional culture and institutionalism process. They may all have negative effects on instructional practices. Apart from these, number of students in classrooms may

also reduce the quality of instruction because in such classrooms there may have some limitations for teachers to apply variety of teaching methods and strategies. In this case, they mostly use traditional teacher-centered teaching method known as didactic approach. Here, as known, in this approach, while teachers are active, students are passive. They can rarely ask questions, which may cause poor teacher-student, student-student interactions. Also, sometimes, classrooms are not appropriate physically for collaborative educational facilities. Therefore, constructivist approach and problem based instruction cannot be implemented, which may also cause negative student perceptions in terms of lecture and teaching evaluation.

It was understood that students with higher GPAs evaluate lecture and teaching more favorably. This may stem from their attitudes towards learning. They probably listen to lectures better, participate in lessons more often, ask questions and have better relations with academicians by visiting them in their offices. Similar results were discovered in some studies (Davidovitch and Soen, 2006; Worthington and Wong 1979). According to them student attendance and participation in classroom facilities affect their evaluations. They tend to evaluate lecture and teaching more favorably. In this respect, while freshmen evaluate lecture and teaching more favorably, as they go upper classes, the evaluation gets worse. Finally, students evaluate the worst at the senior classes. It can be commented that the more they improve their knowledge and skills, the more their expectations increase and they increase their analytic thinking skills at upper classes and they can have a sense of objectivity to evaluate if something is qualified or disqualified, good or bad, correct or incorrect. Moreover, students take different lectures and they meet several academicians during their educational process at colleges. Therefore, they can compare lectures and teaching to the others as well and this may also affect their evaluation decisions. In some other studies, it was also found that successful students rate lecture and teaching more favorably (Addison and Best, 2006; Bacon and Novotny, 2002; Marsh, 1987; Heckert et al., 2006b; Remedios and Lieberman, 2008; Greenwald and Gillmore, 1997).

On the other hand, it was found in some studies that students who get lower grades rate lecture and teaching more negatively. Here, it is considered that teachers should develop themselves professionally. According to a research (Tok, 2011) this professional development equips them with new techniques, methods or approaches to put the new developments into practice in their classrooms. If academic staff do not conduct original researches and share the results of their studies with their students in their lessons and interact with students, it is inevitable to be evaluated less favorably. Because, these researches may provide a deep look in the field and students may benefit if (Holmes, 1972; Vasta and Sarmiento, 1979; Blunt, 1991; Marsh, 1986; Clayson, 2004; Clayson et al., 2006; Chambers and Schmitt, 2002; Norvilitis and Zhang, 2009; Kane et al., 2001; Blackhart et al., 2006). In another study it was found that teachers do not allocate extra time for slow-learners; they do not spend a lot of time in explaining a certain subject and do not provide one-to-one instruction to students (Celik and Topra, 2012).

It can be concluded from this research that different factors affect students' evaluation of lecture and teaching. In order to increase educational quality, lecture and teaching evaluation is substantial in educational institutions. According to some researchers (Mathers et al., 2008), when used appropriately, lecture and teacher evaluation provides instructional quality in terms of strategies, teachers' professional development and behaviors, and delivery of content knowledge that affect student learning. It is stated in a research (Calık, 2011) that teachers play an important role in implementing curriculum and, therefore their effectiveness is crucial in learning environments. The findings of the test items revealed that the PETs had strong understanding about the evaluation phase, moderate understanding about the elaboration phase, but weak understanding about the entrance, exploration, and the explanation phases (Ozsevgec, 2012). The recommendations reached through the results obtained in this study are follows:

- Number of students in classes should be reduced.
- Practical sides of the courses should be increased.
- Excellence Learning Centers should be established at universities in order for academic staff to share their experience and knowledge.
- Colleague mentorship should be improved.
- Research and publication of academic staff should be promoted better.
- Academic staff should benefit from Erasmus projects more.
- Library funds should be increased to reach more qualified and electronic resources.

NOTE

A part of this research was presented as an oral presentationat the First International Applied Education Congress, September 13-15 2012, Middle East Technical University, Turkey. I would also like to thank Dr. Aydın Balyer for his contribution to this study during the congress.

REFERENCES

Addison WE, Best J (2006).Students' perceptions of course difficulty and their ratings of the Instructor. College Student J. 40(2):409-416.
Aleamoni LM (1999). Student rating myths versus research facts from 1924 to 1998. J. Pers. Eval. Educ. 13(2):153-169.
Anikeeff AM (1953). Factors affecting student evaluation of college faculty. J Applied Psych. 7(3): 458-460.

Babad E, Avni-Babad D, Rosenthal R (2004). Prediction of students' evaluations from brief instances of professors' nonverbal behavior in defined instructional situations. Soc. Psychol. Educ. 7:3-33.
Bacon DR, Novotny J (2002). Exploring achievement striving as a moderator of the grading leniency effect. J. Mark. Educ. 24(1):4-14.
Bailey KD (1994). Methods of Social Research, Fourth Ed., A Division of Macmillan. Inc. New York. N.Y.
Baldwin T, Blattner N (2003). Guarding against potential bias in student evaluations: What every faculty member needs to know. College Teach. 51(1):27-32.
Barnes J, Cote J, Cudeck R, Malthouse E (2001). Factor analysis-checking assumptions of normality before conducting factor analysis. J. Consum. Psychol. 10(1,2):79-81.
Bartlett MS (1954). A note on the multiplying factors for various chi square approximations. J. Royal Statistic Soc. 16(Series B):296-298.
Baumgartner H, Homburg C (1996). Applications of structural equation modeling in marketing and consumer research: a review. Int. J. Res. Mark. 13(2):139-161.
Best JB, Addison WE (2000). A preliminary study of perceived warmth of professor and student evaluations.Teach. Psychol. 27(1):60-62.
Blackhart GC, Peruche BM, DeWall CN, Joiner TE (2006). Factors influencing teaching evaluations in higher education.Teach. Psychol. 33:37-39.
Boysen GA (2008) Revenge and student evaluations of teaching. Teach. Psychol. 35(3):218-222.
Brown TA (2006). Confirmatory factor analysis for applied research.The Guildford Press. New York. London.
Blunt A (1991). The effects of anonymity and manipulated grades on student ratings of instructors. Commun. College Rev. 18(4):48-53.
Calık M (2011) How did creating a construvist learning environment influence my graduate students' views. Energy Educ. Sci. Technol. Part B. 3(1):1-13.
Comrey AL, Lee HB (1992). A first course in factor analysis. Hillsdale, NJ: Erlbaum.
Cattell RB (1978). The scientific use of factor analysis. New York: Plenum.
Celik V, Toprak M (2012). Primary school teachers' perceptions regarding the effectiveness of theirschools: (Case of AdiyamanProvince), Energy Educ. Sci. Technol. Part B. 4(4):2197-2208.
Chambers BA, Schmitt N (2002). Inequity in the performance evaluation process: how you rate me affects how I rate you. J. Pers. Eval. Educ. 16(2):103-112.
Chen Y, Hoshower LB (2003). Student evaluation of teaching effectiveness: an assessment of student perception and motivation. Asses. Eval. Higher Educ. 28:71-87.
Cherry B, Ordonez LD, Gilliland SW (2003). Grade expectations: the effects of expectations on fairness and satisfaction outcomes. J. Behav. Decis. Mak. 16(5):375-395.
Clayson DE (2004). A test of the reciprocity effect in the student evaluation of instructors in marketing classes. Mark. Educ. Rev. 14:11-21.
Clayson DE, Frost TF, Sheffet MJ (2006). Grades and the student evaluation of instruction: a test of the reciprocity effect. Acad. Manag. Learn. Educ. 5(1):52-65.
Cohen PA (1981). Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies. Rev. of Educ. Res. 51: 281-309.
Cote JR, Netemeyer R, Bentler P (2001). Structural equation modeling-improving model fit by correlating errors. J. Consum. Psychol. 10(1, 2):87-88.
Costello AB, Osborne JW (2005). Best Practices in Exploratory Factor Analysis: Four. Recommendations for Getting the Most from Your Analysis. Pract. Assess. Res. Eval. 10 (7):1-9.
Davidovitch N, Soen D (2006). Class attendance and students' evaluation of their college instructors. College Student J. 40(3):691-703.
Feldman KA (1976). Grades and college students, evaluations of their courses and teachers. Res. Higher Educ. 4:69-111.
Floyd FJ, Widaman KF (1995). Factor Analysis in the Development and Refinement of Clinical Assessment Instruments.Psychol. Assess. 7(3):286-299.
Gessner PK (1973). Evaluation of instruction. Science 180(4086):566-570.
Goldberg G, Callahan J (1991). Objectivity of student evaluations of instructors. J. Educ. Bus. 66:377-379.
Guadagnoli E, Velicer WF (1988). Relation of sample size to the stability of component patterns. Psychol. Bull. 103(2):265-275.
Gorsuch RL (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.
Greenwald AG, Gillmore GM (1997). Grading leniency is a removable contaminant of student ratings. Am. Psychol. 52:1209-1217.
Guilford JP (1954). Psychometric methods (2nd ed.). New York: McGraw-Hill.
Heckert TM, Latier A, Ringwald-Burton A, Drazen C (2006a). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to 'buy' better evaluations through lenient grading? College Student J. 40(3):588-596.
Heckert TM, Latier A, Ringwald A, Silvey B (2006b). Relationship of course, instructor, and student characteristics to dimensions of student ratings of teaching effectiveness. College Student J. 40:195-203.
Holmes DS (1972). Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructors. J. Educ. Psychol. 63(2):130-133.
Hooper D, Coughlan J, Mullen MR (2008). Structural equation modelling: guidelines for determining model fit. Elect. J. Bus. Res. Methods 6(1):53-60.
Hoyle RH (1995). Structural Equation Modelling: Concepts, Issues, and Aplications. London: Sage Publication Inc.
Hu LT, Bentler PM (1999). Cutoff criteria for fit ındexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6 (1): 1-55.
Hutcheson G, Sofroniou N (1999). The multivariate social scientist: Introductory statistics using generalized linear models. London and Thousand Oaks: Calif. 224-225.
Jörreskog KG, Sörbom D (1993). Structural Equation Modelling with the Simplis Command Language. Scientific Software International Inc. Lincolnwood, IL 60712-1704, USA.
Kane T, Taylor ES, Tyler JH, Wooten AL (2011). Evaluating Teacher Effectiveness. Education Next. (Summer-2011):55-60.
Kline RB (2011). Principles and practice of structural equation modeling.3rd edition, the Guilford Press, New York, NY 10012.
Kline P (1979). Psychometrics and psychology. London: Academic Press.
Kulik JA, Kulik CC (1974). Student ratings of instruction.Teach. Psychol. 1(2):51-57.
Looman WS, Farrag S (2009). Psychometric properties and cross-cultural equivalence of the Arabic Social Capital Scale: instrument development study. Int. J. Nurs. Stud. 46(1):45-54.
MacCallum RC, Browne MW, Sugawara HM (1996). Power analysis and determination of sample size for covariance structure modelling. Psychol. Methods 1(2):130-149.
MacCallum RC, Widaman KF (1999). Sample Size in Factor Analysis. Psychol. Methods 4(1):84-99.
Marsh HW (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. Am. Educ. Res. J. 17(2):219-237.
Marsh HW (1986). The self-serving effect (bias?) in academic attributions: its relations to academic achievement and self-concept. J. Educ. Psychol. 78(3):190-200.
Marsh HW, Hau KT, Chung CM, Siu TLP (1997). Students'evaluations of university teaching: Chinese version of the students' evaluations of educational quality instrument. J. Educ. Psychol. 89(3):568-572.
Marsh HW, Roche LA (1997). Making students' evaluations of teaching effectivenesseffective: the critical issues of validity, bias, and utility. Am. Psychol. 52(11):1187-1191.
Mathers C, Oliva M, Sabrina MV (2008). Improving instruction through effective teacher evaluation: Options for states and districts, Research & Policy Brief, National Comprehensive Center for TecherQulity.US.
McKeachie WJ (1997). Student ratings: The validity of use. Am. Psychol. 52(11):1218-1225.
Millea, M, Grimes PW (2002). Grade expectations and student

evaluation of teaching. College Student J. 36(4):582-590.
Nasser F, Fresko B (2002).Facultyviews of studentevaluation of collegeteaching. Asses. Eval. Higher Educ. 27(2):187-198.
Norvilitis JM, Zhang J (2009). Theeffect of perceivedclassmean on theevaluation of İnstruction. Educ. Asses. Eval. Account. 21:299-311.
Ozsevgec CL (2012). Exploringpre-service elementaryteachers' understanding of 5E instructional model byusingnewlydeveloped 5E instrument, Energy Educ. Sci. Technol. Part B. 4(4):2189-2196.
Radmacher SA, Martin DJ (2001). Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. J. Psychol. 135(3):259-268.
Raykov T, Marcoulides GA (2000). A First Course in Structural Equation Modelling.Lawrence Erlbaum Associates. Inc.
Remedios R, Lieberman D A (2008). I liked your course because you taught me well: the influence of grades, workload, expectations, and goals on students' evaluations of teaching. Br. Educ. Res. J. 34(1):91-115.
Schreiber JB, Stage FK, King G, Nora A, Barlow EA (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. J. Educ. Res. 99(6):323-337.
Schermelleh-Engel K, Moosbrugger H, Müller H (2003).Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Method Psychol. Res. Online 8(2):23-74.
Scherer RF, Wiebe FA, Luther DC, Adams JS (1988). Dimensionality of coping: factor stability using the ways of coping questionnaire. Psychol. Rep. 62:763-770.
Seiver DA (1983). Evaluations and grades: a simultaneous framework. J. Econ. Educ. 14(3):33-38.
Sperber AD (2004). Translation and Validation of Study Instruments for Cross-Cultural Research. Gastroenterology 126(1-1):124-128.
Tabachnick BG, Fidell LS (2007) Using Multivariate Statistics, (Fıfth Edition), Pearson Education. Inc.
Tata J (1999). Grade distributions, grading procedures, and students' evaluations of instructors: a justice perspective. J. Psychol. 133(3):263-271.
Tok H (2011). Classroom instructional responsibilities and learner autonomy: A casestudy in Turkey. Energy Educ. Sci. Technol. Part B 3(3):211-220.
Vasta R, Sarmiento RF (1979). Liberal grading improves evaluations but not performance. J. Educ. Psychol. 71(12):207-211.
Vieira AL (2011). Interactive LISREL in practice, getting started with a simplisapproach.Springer Heidelberg Dordrecht London New York.
Wachtel HK (1998). Student evaluation of college teaching effectiveness: A brief review. Asses. Eval. Higher Educ. 23(2):191-211.
Wilhelm WB (2004). The relative influence of published teaching evaluations and other instructor attributes on course choice. J. Mark. Educ. 26(1):17-30.
Worthington AG, Wong PTP (1979). Effects of earned and assigned grades on student evaluations of an instructor. J. Educ. Psychol. 71(16):764-775.