Eurasian Journal of Educational Research, Issue 61, 2015, 118-136

Course and Instructor Characteristics Distinguishing Highest and Lowest Student Ratings of Instructors

Sevgi ÖZGÜNGÖR [*]
Erdinc DURU [**]

Suggested Citation:

Ozgungor, S. & Duru, E. (2015). Course and Instructor Characteristics Distinguishing Highest and Lowest Student Ratings of Instructors. *Eurasian Journal of Educational Research*, 61, 118-136. http://dx.doi.org/10.14689/ejer.2015.61.7

Abstract

Problem Statement: One way to delineate the main characteristics of effective teaching within the higher education system is to gather college students' opinions of an effective instructor. Research based on students' perceptions of efficient teaching revealed a series of teaching behaviors setting the distinction between good and poor teaching. However, studies also indicate differences across culture, and in Turkey, there has been little research on the topic.

Purpose of Study: The goal of this study was to determine instructor and course characteristics and teaching dimensions that discriminate between instructors who received the highest and the lowest student ratings within a Turkish college setting, by incorporating both quantitative and qualitative methods.

Method: A total of 23,814 students across different departments in the university rated 630 instructors on a scale developed to assess students' perceptions of instructors' performance. In addition, students were asked to respond to an open-ended question to provide their own impressions of each instructor. Then, students' ratings were analyzed by means of discriminative functional analysis, and written statements provided by students were analyzed via content-analysis techniques by using a combination of manual and computer-assisted methods (NVivo 9).

Findings and Results: According to the quantitative analyses, although course and instructor characteristics were weak in discriminating the

[*] Corresponding author: Assoc. Prof. Dr., Pamukkale University, Department of Educational Sciences, PDR.ABD e-mail: sozgungor@pau.edu.tr
[**] Prof. Dr., Pamukkale University, Department of Educational Sciences, PDR.ABD e-mail: eduru@pau.edu.tr

groups, all teaching dimensions (relationships with students, effective teaching, exams and evaluation, contribution to generic skills, class interaction, and organization and planning) were very useful in discriminating the instructors who received the best and the poorest ratings. Also, qualitative analysis revealed 4 themes consistently distinguishing the two groups: lecturing, relationship with the students, knowledge and expertise, and exams and evaluation.

Conclusions and Recommendations: This study replicates the existing literature on student perceptions of effective teaching, with a culturally different, large sample. It also adds support to the notion that there are teaching behaviors, such as lecturing skills, fair evaluations, respect and interest toward students, and demonstrating expertise, that help draw the distinction between good and poor teaching in the eyes of students and that could therefore assist the improvements efforts of teacher education.

Keywords: Instructor effectiveness, poor teaching, college teaching, student evaluations

Introduction

One of the important preconditions to ensure higher education quality is to employ effective instructors and maintain instructors' excellence through students' educational years. Additionally, teacher quality has been linked to long-term achievement going beyond school years (Chetty, Friedman, & Rockoff, 2014). Hence, attempts to define and delineate the main characteristics of the best instructors within a higher education system accumulated numerous studies. The most basic form of this research is conducted through asking college students to rate the most important features of an "effective or ideal" instructor on a researcher-provided scale. In such a study with a semantic differential task, Pozo-Muñoz, Rebolloso-Pacheco, and Fernández-Ramírez (2000) asked students to rate an ideal teacher based on a 7-point rating scale with two bi-polar adjectives listed on opposite ends (e.g., nice-unpleasant; expert-inexpert, etc.). They found four distinguishable factors defining ideal teachers: teaching competency, teachers' qualities, appearance, and directiveness. Zhang, Fike, and DeJesus (2015) reported that the highest rated qualities of ideal instructors were being knowledgeable and grading fairly.

More recent studies combining both survey and qualitative methods asked students both to rate the likelihood of each statement to define an ideal teacher on a scale and to list the most important features of their ideal teacher. In such a study, Okpala and Ellis (2005) reported five main components of quality teaching at higher education: caring for students and their learning, teaching skills, content knowledge, dedication to teaching and love of work, and good verbal skills. Khandelwal (2009) reported six dimensions: encouragement, course preparation and delivery, fairness, rapport with students, spending time with students outside of class, and control. Slate, LaPrairie, Schulte, and Onwuegbuzie (2011) identified four reoccurring themes:

knowledgeability, understanding, communication skills, and teaching well. However, they also detected some differences between undergraduate and graduate students' definitions in that graduate students placed more importance on being knowledgeable, connecting with the real world, passion for the job, and flexibility. In a comprehensive review, Feldman (2007) synthesized the current research on teacher efficiency and reported that the factors most related to teaching effectiveness varied depending on whether the students' own achievement or the results of the overall teaching evaluation is taken as the indication of teacher effectiveness. Within Turkish college settings, Tunca, Alkin-Sahin, Oguz, and Bahar-Guner (2015) obtained five themes by analyzing students' definitions of ideal instructors, namely the role and responsibilities of instructors, values, personal characteristics, social responsibilities, and ethical principles. These findings indicate that although some teacher behaviors are more salient and are better descriptors of effective teachers, still students' choices of best teachers do not necessarily correspond to best learning outcomes, indicating that students' characteristics and expectations might also affect the definition of excellent teachers.

Indeed, studies in which participants come from culturally different orientations other than Western countries seem to further support this possibility. In an earlier study, Bail and Mina (1981) compared American and Filipino students' ratings of 39 statements pertaining to instructor behavior according to their perceived importance for general teaching effectiveness. Although both groups perceived three dimensions as equally important (instructors' competence on the subject matter and their ability to relate it to class materials, quality of feedback and evaluation procedures, and instructor–student rapport), Filipino students placed more importance on qualities pertaining to authority and personal appearance for effective teaching than did American students. Watkins and Akande (1992) reported that although there were similar patterns of findings with Western studies, Nigerian students might hold a more "general view" of the instructors rather than distinguishing different aspects of teaching. Miller, Dzindolet, Weinstein, Xie, and Stones (2001) reported that U.S. students endorsed higher importance to teacher preparation, evaluation, and presentation items and opportunities for student inquiry in defining effective teaching than South African students, who in turn endorsed higher importance to these items than Chinese students. Taken together, these studies indicate that although there are identifiable themes consistent across studies, there are also differences of student opinion as to the best instructors, based on culture. Still, in spite of countless studies on the topic in the world's literature, only a few of them have been conducted in the Turkish culture. Therefore, it remains a need to define best-teacher qualities in the eyes of students. The topic is especially important within Turkish culture since it nests a variety of interesting contradictions compared to Western culture, where the majority of knowledge about teaching qualifications is obtained. Foremost of these contradictions comes from the recent years' radical change of educational policies.

The Turkish educational system has only recently discovered constructivist teaching (Akpinar & Aydin, 2007) and, despite great challenges—such as teacher

readiness needing to be reconciled with parental expectations for higher student scores on standardized tests such as the TEOG, YGS, and KPSS—efforts to establish an educational system based on constructivism remain strong. Constructivism requires students' active participation, good cooperative skills, as well as motivation for and enjoyment from learning activities. However, in a changing world of expectations and rules, students' surroundings that shape their motivation, skills, and even personalities have also been changed. Students are growing up in a technological world where there is less opportunity to practice social skills and fewer responsibilities at home, and exams push students to adopt a more results-orientation rather than focusing on learning and enjoyment. Within such an educational system hosting many contradictions and challenges, it is not easy for teachers to easily understand the students' expectations or define the teacher practices that best fit to these expectations. Hence, in spite of an already accumulated knowledge base on teaching quality, both the existence of differences across studies and the unique complexities of Turkish educational settings warrant further inquiry in this area.

Further, although much research has accumulated, a majority of it focuses on identifying the qualities of teachers who are defined as the best. Only a few compare the features of instructors delineated as either the best or the poorest within the same context. These studies revealed that organization of content, providing variety, knowledge, creating an enjoyable learning experience, and communication skills were the most commonly stated attributes of the most effective instructors, while poor course organization, poor communication, unfairness, and being boring were the most defining characteristics of the worst instructors across studies (e.g., Check, 2001; Epting, Zinn, Buskist, & Buskist, 2004; Fortson & Brown, 1998; Johnson, 1990; Young & Shaw, 1999). However, the majority of these studies—with the exception of only a few—asked students to define their best/worst instructors rather than to actually to rate their best or worst instructors (e.g., Marsh, 1977; Young & Shaw, 1999). Students might apply different criteria for rating a real instructor as opposed to defining an imagined one; when students are asked to rate their best instructors, they inadvertently activate their implicit schemas for an ideal instructor, causing some bias on ratings. For example, students might implicitly hold an ideal instructor schema that is more organized, highly knowledgeable, and highly dedicated to teaching, but they might award a higher rating to an easy grader, or a fun-loving or understanding instructor. Therefore, in this study students were asked to rate each of their instructors' performance and to provide an open-ended explanation of what they thought about the instructor without judging whether he or she is a good or poor instructor. Then, those instructors who received the highest evaluation points for three consecutive semesters were identified as the highest rated instructors. Likewise, those who received the lowest scores for three consecutive semesters were identified as the lowest rated instructors. The two instructor groups were then compared based on demographics such as age, years of experience, gender, academic rank, class size, instructors' total number of students, and total workload (as the number of total hours spent teaching courses in a week) to analyze if demographic variables would change the ratings. Although past studies have already indicated

that neither students' nor instructors' demographics have much effect on student evaluations (Marsh, 1984), in Turkey instructors have much higher workloads and student density than in Western universities where the literature mostly derives from. More importantly, in Turkey, unlike in Western countries, there is no determining utility of students' evaluations or even good teaching skills on hiring and promoting decisions. Hence, it is possible that those who devote more time to academic behaviors, such as publishing, rather than teaching behaviors, such as preparing for the course or caring for the students, might attain a higher academic rank compared to those who spend more time on teaching activities that leaves less room for academic duties. To restate, in spite of the existing studies to distinguish poor and effective teaching in college, the changing face of the educational needs of students as a result of technological advances, cultural needs, and the methodological differences across studies necessitates further research to define teaching-efficient criteria in the eyes of Turkish college students. Therefore, this study aims to extend current knowledge on teaching quality by comparing the most salient features of the instructors who received the highest/lowest evaluation points across the university within a Turkish higher educational setting, by using both qualitative and quantitative methods.

In summary, the main purposes of this study were to determine: (a) Which demographics (if any) discriminate those instructors who received the highest evaluations and those who received the lowest evaluations; (b) which instructional dimension(s) best discriminate those instructors who received the highest evaluations and those who received the lowest evaluations; and (c) whether there are identifiable patterns on students' written statements that distinguish the instructors who received the highest evaluations from the instructors who received the lowest evaluations, and do they support the quantitative analyses?

Method

Research Design

The students were asked to evaluate the performances of all instructors who lectured them in the Fall semester of the 2009–2010 school year, on a scale published on the school's website two weeks before final exams started. Students were also asked to respond to an open-ended question, which required students to provide their own impressions of the corresponding instructor in order to contribute to the improvement efforts of the university's instructional quality. The data were also collected at the end of the next two semesters. That is, feedback was collected in the 2009–Fall, 2010 Spring, and 2011 Fall semesters. An instructor was classified under the highest rated category if his or her evaluations were 1 standard deviation above the mean for three consecutive semesters. Likewise, an instructor was classified under the poorly rated category only if her or his evaluations were 1 standard deviation lower than the mean for three consecutive semesters. Using these criteria, 27 instructors were classified as the highest rated and 32 instructors were classified as the lowest rated. There were 26 male and 6 female instructors in the lowest rated

group, and there were 17 male and 10 female instructors in the highest rated group. For the lowest rated group, 6.2% were professors, 21.9% associate professors, 50% assistant professors, and 21.9% lecturers; for the highest rated group, 22% were associate professors, 51.9% assistant professors, and 25.9% lecturers.

Research Sample

This study used the data collected as a part of the "Pamukkale University Teaching Staff's Instructional Process Evaluation and Improvement Project." The project started in the 2009–2010 Fall semester, and data were obtained from all students attending to the university with the exception of medical school students. A total of 23,814 students (12,142 men, 51%, and 11,672 women, 49%) evaluated 630 instructors. There were 412 male and 218 female instructors. The ages of the instructors ranged from 23 to 64, with a mean of 40. 30 (ss=7. 97). Years of experience of the instructors ranged between less than 1 year to 33 with a mean of 10. 33 (ss=4. 76).

Research Instrument

Demographics. Demographics including students' GPA, instructor gender, course load, instructors' total number of students and class size, academic rank, age, years of experience, and discipline were obtained from the university's data processing unit.

Evaluation of instructors' teaching and educational processes. Students' evaluations of the instructors' performances were measured by a scale developed through modification of the most widely used questionnaires in the student evaluations literature, such as Course Experience Questionnaires (Ramsden, 1991) and SEEQ (Marsh, 1980, 1984; Marsh & Dunkin, 1997), based on the university's educational system's features and needs. Although the new scale was most closely related to SEEQ, its modifications involved reducing the number of items on all subscales and the addition of a subscale to measure students' perception of instructors' role on the acquisition of higher-order thinking skills such as critical thinking, new viewpoints, and flexibility in thinking. The resulting scale had 20 items compromising 6 subscales. Students were also asked to respond to two additional items: One structured item asking students to evaluate the instructor-made difficulty levels of the course after considering the innate nature of the course itself, and one open-ended question to write any useful information about the instructor's performance or how to improve it. The first subscale (Effective Teaching) taps the instructor's teaching skills such as being able to capture student interest and making the content meaningful and valuable; the second subcomponent (Course Organization and Planning) reflects the instructor's ability to organize and deliver the course content in a fluent and comprehensive manner; the third subscale (Exams and Evaluation) was made up from the items measuring the fairness and appropriateness of an instructor's evaluation practices; the fourth component (Relationship With Students) included items measuring the nature and closeness of instructors' communication skills and relationships with students; the fifth subscale (Class Interaction) taps the instructors' energy and behaviors to encourage students'

involvement in class activities; and, finally, the last subscale (Contribution of Generic Skills) measured instructors' ability to support development of students' thinking skills, such as critical thinking and gaining new viewpoints.

Validity and Reliability

Both test–retest and internal reliability values calculated to establish the scale's reliability levels were satisfactory, ranging between .70 and .98 for all subscales. Factor analysis initially revealed one dimension indicating a general view of instruction rather than capturing a multidimensional nature of the teaching behaviors. However, when principal component analysis was forced to create 6 dimensions, it revealed 6 subcomponents explaining 86% of total variance with all the items loading only on the expected dimensions.

Procedures and Statistical Analysis

Results were organized around the study's main purposes and presented accordingly. First, discriminate functional analysis was performed on the data to determine which (if any) demographics, as well as six teaching dimensions, differentiate between the instructors who were awarded the highest scores and those receiving the lowest evaluations. Second, a qualitative content analysis was run by using NVivo 9 to determine the possible repeating patterns in describing the highest and the lowest rated instructors. Finally, both qualitative and quantitative results were combined conceptually to synthesize the findings.

Findings and Results

Discriminant Analyses

At the first step of the statistical analysis, a discriminate functional analysis indicated that only three of the demographics—which were years of experience, student's course grade, and class size—were significant across all demographic variables; however, significant differences existed for all teaching dimensions between the two instructor groups.

The discriminant function produced a large canonical correlation of .97 and a coefficient of determinism of 100. The largest correlation with the discriminant function was Contribution of Generic Skills, followed by Class Interactions, Course Organization and Planning, Effective Teaching, Relationships with the Students, and Exams and Evaluations, in order. All three demographics had a negative contribution to discriminant function. The discriminant function for the stepwise analysis resulted in a correct classification of 100% of the respondents into their respective groups with no misclassification. The results are given in Table 1.

Table 1.

Discriminant Function Analysis of Demographics and Teaching Dimensions between Instructors Who Received the Best Evaluations and Those Who Received the Poorest Evaluations (N=59)

Part A: Correlation with Discriminant Function Variable

	Study Variables	Correlations
Instructional dimensions	Contribution to generic skills	.863
	Class interaction	.799
	Course organization and planning	.773
	Effective teaching	.766
	Relationships with students	.757
	Exams and evaluation	.702
Demographics	Student's grade	-.083
	Class size	-.073
	Instructor's gender	.051
	Academic discipline	.038
	Difficulty level of the course	-.033
	Instructor's total course hours	-.022
	Instructor's total students number	-.020
	Academic rank	.019
	Instructor's age	.005
	Years of experience	-.110

Part B: Tests of Equality of Group Means

Independent Variables	Wilks' Lambda	F	df1	df2	P value
Effective teaching	.093	553.827	1	57	.000
Class interaction	.086	603.545	1	57	.000
Relationships with students	.095	541.009	1	57	.000
Contribution to generic skills	.075	704.281	1	57	.000
Exams and evaluation	.109	464.993	1	57	.000
Course organization and planning	.092	565.220	1	57	.000
Student's grade	.897	6.524	1	57	.013
Class size	.919	5.042	1	57	.029
Difficulty level of the course	.982	1.023	1	57	.316
Instructor's gender	.958	2.499	1	57	.119
Instructor's age	1.000	.028	1	57	.868
Academic discipline	.977	1.353	1	57	.250
Instructor's total course hours	.992	.438	1	57	.511
Years of experience	.832	11.527	1	57	.001
Instructor's total students number	.993	.386	1	57	.537
Academic rank	.994	.341	1	57	.562

Part C: Canonical Discriminant Function

Eigenvalue	Canonical Correlation	Chi-square	df	P value
16,575	97.7	140.458	16	.000

Part D: Classification Result

Actual Group Membership	Predicted Group Membership	
	Best Evaluated	Poorest Evaluated
Best Evaluated	27	0
Poorest Evaluated	0	32

Percent of cases correctly classified = 100%

Qualitative Analysis of the Data

The written statements were analyzed via the content-analysis technique by using a combination of manual and computer-assisted methods (NVivo software for qualitative data management). Within this process, words and sentences that are conceptually similar were organized, coded, and interpreted to create meaningful themes and subthemes. After the organization of the codes, obtained themes and sub dimensions were expressed in percentages and given in a frequency table. The students' written statements were analyzed separately for the two instructor groups. There were a total of 945 responses for the open-ended question for the instructors who received the highest evaluations, and there were 1,393 responses for the instructors who received the lowest scores. However, only a total of 1,895 responses were codeable since most of them consisted of general statements such as "the lecturer was awesome," "this class sucks," or "worst teacher ever" without further explanations as to why. Below is a summary of the themes related to each group and the categories related to these themes (Table 2 and Table 3).

For the instructors who were rewarded the highest points, all the codeable statements were categorized into five themes: Lecturing, relationship with the students, knowledge and expertise, personality, and exams and evaluation. Lecturing characteristics further compromised six subthemes: teaching skills and techniques, value and utility of the content, enjoyment and interestingness of the lessons, instructor's positive attitude toward teaching and students, preparing students for the profession, and development of critical thinking and new viewpoints.

Table 2.

Number and Percent of the Frequencies of the Themes for the Highest Rated Instructors

Themes	Categories	Number of students mentioned the theme	% of the students mentioned the theme
Lecturing	Teaching skills and techniques	160	28.9
	Value and utility	96	17.35
	Enjoyment and Interestingness	71	12.83
	Attitude toward teaching and the students	35	6.33
	Preparing students for their profession	32	5.78
	Critical thinking and new viewpoints	15	2.71
	TOTAL	409	73.96
Relationships with the students		56	10.12
Knowledge and expertise		40	7.23
Personality		25	4.52
Exams and evaluation		23	4.15
	TOTAL	553	100

For the instructors who received the lowest evaluations, a total of five dimensions were obtained: lecturing, examinations and evaluations, content, relationships with the students, and expertise and knowledge. Lecturing had six subcomponents: inefficient or inadequate teaching skills, rote teaching, boring teaching style, providing insufficient examples, negative attitude toward the lessons, and inability to teach according to students' developmental levels, in the order of importance. Exams and evaluations had four subthemes: too difficult or above students' developmental levels, not fair, exams based on only factual information rather than deep understanding, and wanting more than what was given. Content compromised three subthemes: too hard, unnecessary or useless, and outdated. Students' responses about the relationships with the students theme revolved around two main subthemes: frequent use of insults and criticism, and negative attitudes towards students including the use of threats to make the exams or course content too difficult in order to maintain authority within the class.

Table 3.

Number and Percent of the Frequencies of the Themes for the Poorly Rated Instructors

Themes	Categories	Number of students mentioned the theme	% of the students mentioned the theme
Lecturing	Skills (lack of)	264	19.67
	Rote teaching	148	11.03
	Boring	136	10.13
	Providing insufficient example	62	4.62
	Negative attitude toward teaching	44	3.28
	Inability to teach according to students' developmental level	20	1.49
	TOTAL	674	50.22
Exams and evaluations	Difficulty levels/appropriateness to students' levels	97	7.23
	General comments /fairness	75	5.59
	Exam based on factual information	48	3.58
	Wants more than what is given	40	2.98
	TOTAL	220	19.37
Content	Inappropriate difficulty level	140	10.43
	Lack of utility	63	4.69
	Content is outdated	17	1.27
	TOTAL	260	16.39
Relationships with the students	Insults and criticism	79	5.89
	Negative attitudes toward students	90	6.71
	TOTAL	169	12.59
Expertise and knowledge		19	1.42
	GENERAL TOTAL	1342	100

When themes obtained from students' responses for instructors who received the highest scores and lowest scores were compared, it appears that obtained themes and subthemes were very closely matched for the groups. For both groups, the most frequently mentioned theme was related to the instructors' lecturing characteristics including teaching skills and techniques, which made the course effective or ineffective. More than 70% of the statements for the highest rated instructors and more than 50% of the statements for the lowest rated instructors were about the lecturing characteristics. While the statements for the instructors who received the highest evaluations emphasized the efficiency, joyfulness, usefulness (in a variety of ways, such as increasing critical thinking skills, knowledge, adaptation to real life), and gaining vocational skills, as well as a respectful and valuing attitude toward both the course and the students, statements for the lowest rated instructors were about a lack of teaching skills causing memorization rather than learning, the inability to consider students' developmental levels during instruction, boring lecturing, and a negative attitude toward both students and to the content being taught. Likewise, although the ordering based on mentioned frequency was different for each group, statements regarding both groups involved student and instructor relations, knowledge and expertise levels of the instructors, and exams and evaluations. With the exception of the exams and evaluation theme, for all themes, statements for each group were in essence bipolar statements. That is, the highest rated instructors were cherished for being caring, kind, and respectful to students and to their ideas, while the lowest rated instructors were described as being rude, inconsiderate to students and their needs, and disrespectful to their ideas. However, for the exam and evaluations, both groups' statements echoed the toughness of the content; while the highest rated instructors' students perceived this toughness as a necessary component of understanding and learning, students of the lowest rated instructors thought the toughness was created artificially by inefficient teaching behaviors, making it an obstacle to learning rather than a help.

Discussion and Conclusion

The purposes of this study were to determine which instructor and course characteristics and teaching dimensions best differentiate the instructors who received the highest evaluation points from those instructors who received the lowest evaluation points for three consecutive semesters, incorporating both quantitative and qualitative methods. For this aim, students' evaluations collected through a teaching evaluation scale were analyzed and compared to the obtained themes derived from students' open-ended statements. Quantitative analyses indicated that although the discriminative powers were very little, three demographics were useful in discriminating two groups, namely instructors' experience in years, the student's grade in the course, and the number of students in the course. Previous studies usually found little or no effect of either instructor or students demographics (e.g., Marsh, 1984; Beran & Violato, 2005), with some exceptions. This study's findings are compatible with previous findings since, although there were some effects, these effects were very little for all demographics

including the strongest of all three, which is experience. These results are in line with earlier studies which found a small but inverted U-shaped quadratic relation in which instructors receive higher evaluations initially, level up at some early point, and then decline slowly thereafter (Marsh & Roche, 2000). However, some cultural differences in academic settings might also be accountable for this result since the establishment of higher academic standards in the higher education system became important in Turkey only recently. Therefore, younger instructors might be more knowledgeable and inclined to use active learning techniques, and be more understanding toward students both because it is an important element of active learning and because they share a similar conceptualization of authority due to lesser cohort differences. In terms of class size, past research usually shows a significant but small effect of class size on student ratings (Hanushek, 2002). This study further extends the current findings by showing that, although small, this effect is also applicable for distinguishing between good and poor instructors in the eyes of students; the number of students in the class might cause an otherwise a regular instructor to be viewed as a very poor instructor. Likewise, the inefficiency of a poor instructor can be no longer compensated by the students' own efforts within crowded classes.

As to students' grades, findings on this issue are more complicated and less agreed upon. While some studies found significant correlations between students' evaluations and course grades (Cashin, 1995), some others found no or little relationship between these two (Marsh & Roche, 1999). Further, researchers' interpretation of the meaning of the existing link differ in that some consider this link as a sign of validity of the ratings since better instructors' students should learn better, resulting higher grades (Marsh & Roche, 1997). Others consider this link to be a sign of threat to the validity of student ratings, however, since high evaluations are attributed to students' appreciation of grading leniency (Greenwald & Gillmore, 1997). This study's findings add further support to the validity of students' ratings since, within the efficient teaching literature, the best instructors are usually defined as challenging, someone who sets higher standards and pushes students to do their best (e.g., Acker, 2003; Slate et al., 2011). Both qualitative and quantitative analyses of this study were in line with this notion. According to qualitative analysis, students in both groups frequently mentioned that the course was difficult and that the instructor had high expectations. However, the explanations for the difficulty levels were entirely different. Students thought that the instructors who garnered the highest ratings provided a rich learning experience and expected the best from the students in return. In contrast, students believed that the poorly rated instructors made the course harder by asking useless questions, relying on rote teaching, and under-teaching or avoiding certain concepts. According to the students, the highest rated instructors not only delivered the content very efficiently, but also they held high expectations for better performance. Students seem to perceive the high expectation for success as fair and useful as long as instructors themselves create the necessary conditions for it. On the other hand, poorly rated instructors' exams seem to be perceived as difficult only because they are unfair in the sense that these instructors ask more than what they provide and their exams contain rote learning

and memorization with useless content, making it hard to learn and get motivated to learn—which was described as unfair by the students.

While only a few demographics were useful in discriminating the two groups, all teaching dimensions were important in the classification of the groups. Teaching dimensions in order of the highest contributions to the discriminate variance were generic skills, class interaction, course organization and planning, effective teaching, relationships with the students, and exams and evaluations. These results are important in the sense that they lend further support to previous studies of teaching efficiency with data obtained from students with, culturally, a very different orientation. Good instructors might possess a variety of characteristics, however, some of these characteristics keep reappearing across studies: good teaching skills (using a variety of teaching methods, giving examples to illustrate concepts, increasing student interest and value), good organization and preparation, instruction in an enjoyable and enthusiastic way, being knowledgeable, caring deeply for students and teaching, having a passion and commitment to teach, being fair in exams and grading, creating a comfortable learning environment, and emphasizing and using active learning (e.g., Fortson & Brown, 1998; Khandelwal, 2009; Slate et al., 2011). Likewise, although much more limited, literature on poor instructors demonstrates that poor instructors also possess some characteristics which are comparable across studies. Poor instructors are boring (Fortson & Brown, 1998; Check, 2001), unable to deliver the content because of a lack of organization (Fortson & Brown, 1998; Check, 2001) or a lack of knowledge or ability (Check, 2001), and unable to relate to students due to poor communications skills (Johnson, 1990; Epting et al., 2004). In addition to being disorganized, they are unfair in evaluation practices (Johnson, 1990; Khandelwal, 2009). The current study's findings, combined with the existing literature, indicate that efficient and poor instructors are not very different from each other. Rather, they are the flip side of the same coin—meaning, that although there is no certain definition of good teaching, there are some elements that are very useful when possessed and very harmful when missing for efficient teaching, at least in the eyes of students.

Another contribution of this study was through converging quantitative and qualitative results to provide a better understanding for the distinction between the lowest and highest rated instructors. According to this analysis, although qualitative results were mainly supportive of the quantitative analyses, there were dimensions obtained from qualitative analyses not captured by the sole use of the scale, such as instructors' personality, the encouragement of the development of critical thinking, and knowledge and expertise levels. Patrick (2011) reported that the dimensions of openness and conscientiousness were the best personality predictors of both course and instructor ratings. The most frequently mentioned features of the instructors in written statements of the current study were also closely related to conscientiousness, openness, and agreeableness. Based on these findings, it is possible that there might be a relationship between an instructor's personal resources and the academic, developmental, and psychological needs of students. In other words, while the conscientiousness dimension of the highest rated instructors'

personalities might serve to meet the academic needs of students, the openness and agreeableness dimensions of these instructors' personalities might serve to meet the interpersonal needs of students, including their developmental and psychological requests. Taken together, these findings seem to lend further support to Chickering and Reisser's (1983) theory, which postulates that the postsecondary experience has the potential to foster feelings of intrinsic motivation by fulfilling students' autonomy, competence, and relatedness needs.

In short, this study replicates the existing literature on teaching effectiveness with a culturally different and large sample, and adds clear support that there are actual teaching behaviors that really draw a distinction between good and poor teaching, and therefore could assist efforts for quality improvements in a higher education system.

References

Acker, J. R. (2003). Class acts: Outstanding college teachers and the difference they make. *Criminal Justice Review, 28*, 215-231.

Akpinar, B. & Aydin, K. (2007). Egitimde degisim ve ogretmenlerin degisim algilari, [Change in education and teachers' perceptions of change]. *Egitim ve Bilim, 32* (144), 71-80.

Bail, F. T. & Mina, S. S. (1981).Filipino and American student perceptions of teacher effectiveness. *Research in Higher Education, 14* (2), 135-145.

Beran, T. & Violato, C. (2005). Rating of university teacher instruction: How much do student and course characteristics really matter? *Assessment and Evaluation in Higher Education, 30*, 593-601.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: teacher value added and student out comes in adulthood. *American Economic Review, 104* (9), 2633–2679.

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

Check, J.F. (2001). Positive traits of the effective teacher - negative traits of the ineffective one. *Education, 106* (3), 326-334.

Chickering, A. W. & Reisser, L. (1983). *Education and Identity*. (2nd ed.) San Francisco: Jossey-Bass.

Epting, L. K., Zinn, T. E., Buskist, C., & Buskist, W. (2004).Student perspectives on the distinction between ideal and typical teachers. *Teaching of Psychology, 31* (3), 181-183.

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. Perry & J. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 93-129). Dordrecht, The Netherlands: Springer.

Fortson, S. B. & Brown, W. E. (1998). Best and worst university instructors: The opinions of graduate students. *College Student Journal, 32*, 572-576.

Greenwald, A. G. & Gillmore, G. M. (1997) Grading lenience is a removable contaminant of student ratings. *American Psychologist, 52*(11): 1209-1217.

Hanushek, E.A. (2002). Evidence, politics, Oxford and the class size debate, in L. Mishel & R. Rothstein (Eds.) *The Class Size Debate*. Washington, DC: Economic Policy Institute.

Johnston III, G.P., (1990). Best liked/least liked teacher attributes: Herzberg's two-factor theory of job satisfaction. *Journal of Education for Business, 66* (2), 121-129.

Khandelwal, K. A. (2009). Effective teaching behaviors in the college classroom: a critical incidents technique from students' perspectives. *International Journal of Teaching and Learning in Higher Education, 21*(3), 299-309.

Marsh, H. W. (1977). The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. *American Educational Research Journal, 14*, 441-447.

Marsh, H. W. (1980). The influence of student, course and instructor characteristics on evaluations of university teaching. *American Educational Research Journal, 17*, 219-237.

Marsh, H.W. (1984). Students' evaluation of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76,* 707-754.

Marsh, H. W., & Dunkin, M. (1997). Student evaluation of university teaching: A multidimensional perspective. In Perry, P. R., & Smart, J. C. (Eds.), *Effective Teaching in Higher Education: Research and Practice*, (pp. 241-320).Agathon, New York.

Marsh, H. W. & Roche, L.A (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*, 1187-1197.

Marsh, H.W. & Roche, L.A. (1999). Rely upon SET research. *American Psychologist, 54*, 517–518.

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, and innocent bystanders. *Journal of Educational Psychology, 92*, 202-22.

Miller, J. L., Dzindolet, M. T., Weinstein, L., Xie, X., & Stones, C. R. (2001). Faculty and students views of teaching effectiveness in the United States, China and South Africa. *Teaching of Psychology, 28*(2), 138-142.

Okpala, C.O. & Ellis, R. R. (2005). The perceptions of college students on teacher quality: A focus on teacher qualifications. *Education. 126* (2), 374-383.

Patrick, C.L. (2011). Student evaluations of teaching: Effects of the big five personality traits, grades, and the validity hypothesis. *Assessment & Evaluation in Higher Education, 36* (2), 239-249.

Pozo-Muñoz, C., Rebolloso-Pacheco, E., & Fernández-Ramírez, B. (2000).The 'Ideal Teacher' implications for student evaluation of teacher effectiveness. *Assessment & Evaluation in Higher Education, 24*(1), 253-263.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education, 16*, 129-50.

Slate, J., LaPrairie, K. N., Schulte, D. P., & Onwuegbuzie, A. J. (2011).Views of effective college faculty: a mixed analysis. *Assessment & Evaluation in Higher Education, 36* (3), 331-346.

Tunca, N., Alkin-Sahin, S., Oguz, A. & Bahar-Guner, H. O. (2015). Qualities of Ideal Teacher Educators. *Turkish Online Journal of Qualitative Inquiry, 6*(2), 123-148.

Young, S. & Shaw, D. G. (1999). Profiles of effective college and university teachers. *The Journal of Higher Education, 70* (6), 670-686.

Watkins, D. & Akande, A. (1992). Student evaluations of teaching effectiveness: A Nigerian investigation. *Higher Education, 24*, 453-463.

Zhang, S., Fike, D., & DeJesus, G. (2015). Qualities university students seek in a Teacher. *Journal of Economics and Economic Education Research ,16* (1), 42-54.

En Yüksek ve En Düşük Öğrenci Değerlendirmelerini Alan Öğretim Elemenalarını Ayıran Demografik ve Öğretim Elemanı Özellikleri

Atıf:

Ozgungor, S. & Duru, E. (2015). Course and Instructor Characteristics Distinguishing Highest and Lowest Student Ratings of Instructors. *Eurasian Journal of Educational Research*, 61, 118-136. http://dx.doi.org/10.14689/ejer.2015.61.7

Özet

Problem Durumu: Yükseköğretimde kaliteyi sağlamanın en önemli yolu öğrencileri gelecekteki mesleklerine hazırlayan eğitim kadrosunun kaliteli ve yetkin olmasını sağlamaktır. Bu yüzden yükseköğretimde etkili öğretim elemanını tanımlamaya yönelik pek çok araştırma yapılmıştır. Bu araştırmalarda kullanılan en temel yöntem üniversite öğrencilerinden ideal, etkili ya da en iyi öğretim elemanlarını tanımlamalarını istemektir. Bu çalışmalarda yükseköğretimde etkili öğretim elemanını tanımlayan özelliklerin başında öğrenci ve öğrenme düzeylerini önemseme, öğretme becerileri, alan bilgisi, öğretmeyi sevme ve işini sevme, dersin iyi planlanması, adil değerlendirme gibi özellikler sıralamıştır. Ancak var olan çalışmalar, sonuçların lisans ve yüksek lisans öğrencileri için farklılaştığını ve öğrencilerin etkin öğretim elemanlarına ilişkin düşüncelerinin farklı kültürel yapılara göre değiştiğini göstermiştir. Alan yazında öğrencilerin tanımladığı etkili öğretim elemanı özelliklerinin kültüre göre değiştiğini öne süren çalışmalara rağmen yükseköğretimde etkili öğretim elemanı tanımlamasına ilişkin bulguların neredeyse tamamı yurtdışında yapılan çalışmalara dayalıdır. Oysa son yıllarda teknolojik, kültürel ve eğitim politikaları açısından çatışan özellikleri barındıran Türkiye için bu konunun araştırılması daha da önemlidir. Şöyle ki, bir taraftan eğitim politikalarındaki değişimler öğrencilerin aktif katılımını gerektiren yapılandırmacı eğitimi teşvik ederken, bir taraftan TEOG, LGS, KPSS gibi sınavlar öğrencileri daha hazırcı ve ezberci olmaya itmektedir. Bu tür çelişkileri aynı anda barındıran eğitim sisteminde görev yapan eğitimciler için öğrencilerin gözünde en yararlı ve bu

yüzden öğrenmeye güdüleyici öğretmen davranışlarının ne olduğunun belirlenmesi daha önemli hale gelmektedir. Aynı zamanda, ülkemizde öğretim elemanlarının üzerindeki öğrenci sayısı ve ders yükünün daha yoğun olması gibi unsurlar göz önünde bulundurulduğunda öğretim elemanlarının performanslarında kişisel özellikleri dışındaki özelliklerin de belirleyici olduğu düşünülebilir.

Araştırmanın Amacı: Bu çalışmanın temel amaçları (a) üniversite öğrencileri tarafından sunulan performans değerlendirmelerinde en yüksek ve en düşük puanları alan öğretim elemanlarını ayrıştıran demografik özelliklerin belirlenmesi (b) üniversite öğrencileri tarafından yapılan performans değerlendirmelerinde en yüksek ve en düşük puanları alan öğretim elemanlarını ayıran öğretim etkinliklerine ilişkin boyutların belirlenmesi (c) öğrencilerin en yüksek ve en düşük puanları verdikleri öğretim elemanlarına ilişkin açık uçlu soruya yönelik değerlendirmelerinin analiz edilerek iki grubu ayırmada işlevsel olan anlamlı örüntüler oluşturulup oluşturulamayacağının belirlenmesidir.

Araştırmanın Yöntemi: Çalışma kapsamında veriler Tıp Fakültesi hariç Pamukkale Üniversitesi'ne bağlı tüm akademik birimlere devam etmekte olan toplam 23814 (12142 erkek ve 11672 kız) öğrenciden derslerine giren öğretim elemanlarının performanslarını üniversitenin web sitesinde yayınlanan bir ölçek üzerinden değerlendirmeleri istenerek elde edilmiştir. Kullanılan ölçek Marsh tarafından geliştirilen SEEQ'nin Türkçe uyarlamasını kapsasa da, çalışmanın amacına uygun bazı değişimler de içermektedir. Bu değişimler madde sayısının kısaltılması ve öğrencilerin bilişsel becerilerinin gelişimini destekleyici davranışlara yönelik yeni bir boyutun eklenmesini içermektedir. Nihai ölçek altı alt boyut içermektedir: etkili öğretim, sınıf içi etkileşim, öğrencilerle ilişkiler, planlama, ölçme ve değerlendirme ve genel becerilere katkı boyutları. Veriler 2009 Güz, 2010 Bahar ve 2011 Güz dönemlerinin sonunda final haftasından önceki 2 hafta sürecinde toplanmıştır. Öğrencilere aynı zamanda değerlendirdikleri tüm öğretim elemanı için düşüncelerini belirtmeleri istenen açık uçlu bir soru sunulmuştur. Üç dönemin sonunda, her dönem için standart sapmanın 1 yukarısında ve 1 aşağısında değerlendirmeler alan öğretim elemanları belirlenmiştir. Bu şekilde üç dönemde standart sapmanın yukarısında ve aşağısında puan alan öğretim elemanları belirlenerek öğrenciler tarafından performansları en iyi ve en kötü olarak algılanan öğretim elemanları tespit edilmiştir. Bu sınıflama sonucu toplam 630 öğretim elemanının 32'si (26 erkek, 6 kadın) en düşük değerlendirmeleri alan, 27 (17 erkek, 10 kadın) öğretim elemanı da en yüksek puanları alanlar olarak sınıflandırılmıştır. Öğrencilerin açık uçlu yanıtları NVivo 9 programı kullanılarak içerik analizi yöntemiyle analiz edilmiştir.

Araştırmanın Bulguları: Araştırmanın birinci amacına yönelik olarak en yüksek puanları alan öğretim elemanları ile en düşük puanları alan öğretim elemanlarını, öğretim elemanının deneyimi, cinsiyeti, toplam ders yükü, öğrenci sayısı, unvanı, öğrencinin ders başarı notu, akademik birimi, sınıf mevcudu ve dersin zorluk derecesini içeren demografik özelliklerin ayırmada yararlı olup olmadığını belirlemek amacıyla discriminant analizi yapılmıştır. Analiz, sadece 3 demografik değişkenin (deneyim, öğrenci başarı notu ve sınıf mevcudu) grupları ayırmada etkili olduğunu, ancak bu etkinin zayıf olduğunu göstermiştir. Araştırmanın ikinci amacına yönelik olarak yapılan analizler öğretim etkililiğine ait tüm alt boyutların grupları ayırmada etkili olduğunu ortaya koymuştur. Son olarak, öğrencilerin açık

uçlu yanıtlarının içerik analizi yoluyla incelenmesi sonucu en yüksek puanları alan öğretim elemanlarını tanımlayan 5 alt boyut elde edilmiştir; dersin işlenişi, öğrencilerle ilişkiler, uzmanlık, kişilik ve ölçme değerlendirme. Benzer şekilde, en düşük puanları alan öğretim elemanlarına yönelik yanıtlar dersin işlenişi, ölçme ve değerlendirme, içerik, öğrencilerle ilişkiler ve uzmanlık olmak üzere yine 5 alt boyut oluşturmuştur.

Araştırmanın Sonuçları ve Önerileri: Bu çalışmanın temel amacı takip eden 3 dönem boyunca öğrenciler tarafından en yüksek ve en düşük değerlendirmeleri alan öğretim elemanlarını ayrıştıran demografik ve öğretim etkinliği boyutlarının belirlenmesidir. Yapılan analizler demografik değişkenlerden iki grubu ayrıştırmada etkili özelliklerin öğretim elemanının deneyimi, sınıf mevcudu ve öğrenci başarı puanı olduğunu ve öğretim etkinliği alt boyutlarının tamamının iki grubu ayrıştırmada etkili olduğunu göstermiştir. Öğrencilerin açık uçlu cevaplarına yönelik yapılan içerik analizleri nicel bulguları destekler niteliktedir. Analizler sonucu en yüksek puan alan öğretim elemanlarını tanımlayan 5 alt boyut oluşturmuştur. Öğrenci algılarına göre, en yüksek puanları alan öğretim elemanları, dersi işlerken farklı ve eğlenceli yöntemler kullanan, öğrencileri aktif kılan, derste öğrencilerle sürekli etkileşim içinde olan, alanında uzman, öğrencilere ve öğrenci gelişimine içten ilgi duyan, yaptığı işten haz alan ve bunu sınıf içi hareketleriyle sürekli sergileyen, öğrencilerin mesleki ve kişisel gelişimlerini destekleyen ve dünyaya ve insanlara karşı pozitif tutumlarıyla öğrencilerin yaşama bakışında radikal değişiklikler yaratan öğretim elemanlarıdır. Ayrıca bu öğretim elemanlarının sınavları zor ancak adil olarak tanımlanmaktadır. En düşük puan alan öğretim elemanları genelde aynı özelliklerin yokluğuyla tanımlanmıştır. Ders işleyişi sırasında sıkıcı bir tarz sergileyen, sunu tekniği kullanan, kitabı okuyarak ya da ezberden tekrarlayarak, öğrenci katılımını sağlamayan bu öğretim elemanları aynı zamanda içeriği güncelleştirememeleri ya da öğrencilerin günlük ve mesleki yaşamlarıyla ilişkilendirilmemeleri nedeniyle eleştiri almışlardır. Alanında yetersiz olarak algılanan bu bireyler, derse ve öğrencilere karşı olumsuz, aşağılayıcı davranışlarıyla tanımlanmıştır. Son olarak, bu öğretim elemanlarının anlatmadıkları konuları sınava dâhil ederek, yeterli örnek çözmeyerek, ezber sorarak sınavları zorlaştırdıkları ve adil olmak yerine kişisel ilgi ve ihtiyaçlarına göre puanlandırma yaptıkları belirtilmiştir. Özetle, bu çalışmanın bulguları, öğrencilerin en yüksek ve en düşük puan verdikleri öğretim elemanlarının alan yazında yer alan etkili öğretmen tanımıyla paralel olduğunu ve bulguların yükseköğretimde eğitim kalitesini artırmada yararlı olacağını göstermektedir.

Anahtar Sözcükler: Öğretmen etkinliği, yükseköğretimde eğitim kalitesi, öğrenci değerlendirmeleri, iyi öğretmen