International Journal on New Trends in Education and Their Implications
October 2012 Volume: 3 Issue: 4 Article: 08 ISSN 1309-6249

STUDENT EVALUATION OF TEACHERS: A CASE STUDY AT TERTIARY LEVEL

Assist. Prof. Dr. Evrim ÜSTÜNLÜOĞLU
School of Foreign Languages
Izmir University of Economics
Izmir-TURKEY

Assist. Prof. Dr. Seda CAN
Department of Psychology
Gediz University, Izmir-TURKEY

ABSTRACT

The increasing number of institutions using "Student Evaluation of Teaching" (SET) has caused a growing controversy in the literature. Studies on the use of students' ratings for evaluating teacher effectiveness have mostly questioned students as valid judges of teaching effectiveness, revealing positive and negative effects of SET. Based on the literature indicating that student ratings should be combined with data collected from different sources, the purpose of this study is to identify the relationship, if any, between the evaluation ratings of the students and coordinators. The study also aims to investigate the consistency of the ratings of the two groups over a two-year period. The participants consisted of 1028 Preparatory Program students, 99 teachers, 4 trainers and 4 Preparatory Program coordinators in the first year study, and 1211 Preparatory Program students, 99 teachers, nine coordinators in the second year. A Pearson's correlation addressed the relationship between the mean ratings for teachers by the students (M = 4.12, SD =.56) and by the coordinators (M = 4.7, SD =.46). The correlation between the ratings was found to be statistically significant, r (99) =.45, p <.001. This indicates that ratings for teachers given by the students and given by the coordinators were positively related. Another correlation was employed to determine the relationship in the second year. The correlation between the ratings of the students (M = 4.38, SD =.44) and coordinators (M = 4.80, SD =.30) was also significant, r (99) =.43, p =.00. The study suggests that data collected from students is a valid evaluation tool in evaluation of the teachers.

Key Words: Students' evaluation, feedback, teacher, teaching.

INTRODUCTION

One of the most important resources for student learning, teacher effectiveness and professional development is feedback and evaluation given by students. As the number of institutions using "Student Evaluation of Teaching" (SET) to evaluate teachers increases, questions have been increasingly raised about the validity of these surveys as an indicator of instructor effectiveness. This has caused a growing controversy in the literature, as student evaluations of teaching play a vital role in the promotion and tenure process.

Together with ongoing controversy on SET, a growing body of research has examined the question of whether students are valid judges of teaching effectiveness (Jones, 1989; Ory & Ryan, 2001). In their study, Chen and Hoshower (2003) investigated the importance of student evaluations, and found that students can offer meaningful feedback when they believe and see that their input is being valued. Renaud and Murray (2005)

suggest that student ratings are adequate in terms of reliability, which means that ratings are reasonably consistent across courses, years, and groups of raters; and they are valid and adequate, which means that they are relatively free of bias, and agree with evaluations made by other evaluators such as colleagues. Renaud and Murray (2005), emphasizing the importance of SET, claim that student evaluation of teaching is spreading rapidly across the world, and is having an impact in three areas: (i) faculty personnel decisions, (ii) improvement of quality of teaching, and (iii) academic standards. They add that student evaluation of teaching and, particularly, the written comments makes teacher evaluation data more convincing, meaningful, and contribute to improvement of teaching accordingly. Theall, Abrami and Mets (2001) agree that student evaluations of an instructor provide a reliable, valid assessment of that instructor's effectiveness, especially where assessments reflect the views of a significant number of students involved in different course settings. Believing that students' views should be taken into account in a teacher-supportive evaluation system, Murdoch (2000) considers that teacher-evaluation is often considered of secondary importance in comparison to issues such as revising curricula and teaching materials, and introducing new technology. He claims that this tendency leads to a poor evaluation in many institutions.

On the other hand, the drawbacks of SET have also been investigated. One has been found to be its negative effect on academic skills as it might cause grade inflation and lowering of academic standards. It is claimed that since faculty members know that student evaluations are used in personnel decisions, they are reluctant to give low grades. They assume that low grades will lead to students seeking revenge in the form of low teacher ratings, and teachers may respond by raising grades and thus, SET may not be a valid evaluation system (Murray, 1997; Renaud & Murray, 2005). Ryan, Anderson and Birchler (1980) note that at least one-third of teachers in their survey had substantially decreased their grading standards and level of course difficulty. It was also found in one of the studies that students believe instructors usually ignore the results of student evaluation of teaching and they get discouraged and stop reflecting their real thoughts (Koç and Çoşkuner, 2007). Summarizing all these above mentioned issues, McKeachie (2006) points out four major criticizisms of SET: (i) students are not able to make informed and consistent judgments about their instructors; (ii) students cannot make accurate judgments until a certain period of time has elapsed after the course; (iii) student ratings are negatively related to student learning; and (iv) student ratings are based upon expected grades in the course.

Despite drawbacks, the use of SET is becoming increasingly widespread. Therefore research has focused on raising awareness of the factors affecting SET and how to structure SET in terms of content and administration. This has lead to a variety of findings, with some studies suggesting that signed rating forms are more positive compared to anonymous ones; that ratings given during final exam week are generally lower than those given during a regular class period; and that discussion of the importance of the ratings with students slightly raises ratings (Feldman, 1977). Some studies indicate that while structuring and interpreting SET, factors such as the instructor's personality, level of the course, time of day, class size, different disciplines, gender, number of years experience, interest and academic ability should be considered carefully because all these affect student ratings. It has been noticed in these studies that higher level courses have better course ratings than lower level courses; ratings are slightly higher in classes where the majority of the students are the same gender as the instructor; new courses usually get lower than expected ratings the first time they are taught; and students with higher interest in a course give somewhat higher ratings to the instructors (Ory & Ryan, 2001; Theall & Franklin, 2001). Some studies which have investigated influence of teacher's gender on student evaluations of teachers suggest that students give high rankings to teachers, no matter what their gender is, as long as they know their subject matter, teach well, make fair assessments (Yurtkoru and Sipahi, 2003). Studies emphasizing instructor personality as a factor also yield controversial findings. Erdle, Murray and Rushton (1985) claim that student ratings may be influenced by the instructors' personality rather than their teaching practices. Highlighting that personality is a sensitive issue, these researchers suggest that instructional ratings should not be used in decision-making about faculty promotion and tenure, because charismatic and enthusiastic faculty can receive favourable ratings regardless of how well they know their subject matter. For this reason, rating

scales should avoid focusing on aspects of personality, such as charisma or similar attributes. Rather, the emphasis should be on instructor traits which are related to effective teaching, such as student-teacher interaction or concern for students' learning (Cooper & Simonds, 2007).

Summarizing all the issues mentioned above, Murray (1997); Renaud and Murray (2005) state that student evaluation forms can assess only those characteristics that are observable by students, such as covering learning objectives, keeping to teaching hours, fulfilling all teaching hours, speaking clearly, keeping the classroom environment positive for learning, knowing the names of the students, and choosing appropriate materials. However, Murray (1997) suggests that they cannot assess non-classroom factors such as course design or substantive factors such as instructor knowledge, academic standards and quality of assignments. He acknowledges that SET in itself is inadequate as a means of evaluation, and must, therefore, always be supplemented by other sources of data.

Based on the concerns shared by Renaud and Murray (2005), Ory and Ryan (2001) agree that student ratings are one, but not the only way to evaluate instruction, adding that student ratings should be combined with data collected from different sources such as peer review, teaching portfolios, classroom-observations, or self-evaluation. Murray (1997) and Marsh (1987) suggest that a well-structured evaluation process can be promoted by educating students on how to give precise and meaningful feedback, and clarifying the purposes for which the ratings will be used by the university. According to Murdoch (2000), the likelihood of a positive outcome for teachers can be increased by the inclusion of a follow up stage which could include professional development activities.

Studies conducted on SET emphasize the importance of collecting data from different sources and evaluating teaching or teacher effectiveness accordingly. Based on the literature indicating that student ratings should be combined with data collected from different sources and the process of SET should be well structured for a reliable and valid evaluation of teaching, the purpose of this study is to identify the relationship, if any, between the evaluation ratings of the students and coordinators. In addition, the study also aims to investigate the consistency of the ratings of the students and coordinators over a two-year period and to give suggestions about how the results of SET should be used. To this end, this study addresses the following questions:

1. How far does SET agree with the evaluations of coordinators?
2. Are the ratings consistent over a two-year period?
3. What use should SET be put to?

This study, the first of its kind conducted within the School of Foreign Languages, is expected to have implications not only for the evaluation of the teachers, but also for their professional development.

METHOD

Participants

The study was conducted at a Preparatory Program, English Medium University. Students failing to meet the proficiency level of English required to study at faculty level attend the Preparatory Program run by the School of Foreign Languages (SFL). This Program prepares the students for their Faculties through an intensive English Preparatory year, offering 25-30 teaching hours ranging from Elementary to Pre-Advanced levels, in classes of 19 or 20 students, with an age range of 18-22. The study was conducted in two academic years; and the participants consisted of 1028 Preparatory Program students studying at different levels, 99 teachers and nine coordinators at the Preparatory Program in the first year and 1211 Preparatory Program students, 99 teachers, and nine coordinators in the second year.

Questionnaires

To ensure a supportive evaluation system, and for the sake of fairness, data were collected through observations by questionnaires from students and coordinators. For an unbiased picture of an instructor's abilities, the emphasis was on teaching ability, rather than the instructors as individuals (Murdoch, 2000). The questionnaire was developed to collect data from the coordinators and the students. The questionnaire included 10 five-point Likert-scale items on teaching ability of the instructors. The Likert-scale included five points ranging from 5 (always) through to 1 (never). The participants rated how often the instructor does the items presented in the questionnaires. The questionnaire covered in-classroom teaching roles of a teacher and included the following ten items: ensuring that each lesson links with the previous lesson and the lesson that follows, making the aims of the lesson clear to students, implementing relevant methods and techniques related to the topic of the lesson, , adjusting content of the lesson according to student level, using aids and materials in a timely and appropriate manner, arousing interest and encouraging students to ask questions, checking achievement of lesson aims, identifying and correcting students' mistakes, concluding and summarizing the lesson clearly, and informing the students about the content and related sources for the following lesson.

The questionnaire was reviewed and pilot tested prior to the study, in order to refine and validate the instrument for students and coordinators separately. Factor analysis was performed on the data collected from 87 Preparatory Program students. Principle components analysis with varimax rotation was used because to identify the factors underlying the evaluation of teaching roles of a teacher. The initial eigenvalues for the students' questionnaire showed that the first factor explained 59.28% of the variance, the second factor 8.66% of the variance. One factor solution was also examined and it was preferred its previous theoretical support and the 'leveling off' of eigenvalues on the scree plot after one factor. No items were eliminated, because they all contributed to a simple factor structure with factor loadings higher than .40. Internal consistency for the questionnaire was examined using Cronbach's alpha. The reliability coefficient was 0.96. The item-total correlations for the each item were also examined. They were between .55 and 0.84. The results of the reliability analysis showed a high internal consistency for the students' questionnaire.

The factor analysis for the questionnaire of the coordinators was also performed to examine the factor structure of the data. Principle components analysis with varimax rotation showed a single factor with an explained variance of 43.97%. The Cronbach alpha coefficient was .85 and the item-total correlations were also high with values ranging from .38 to .75.

Procedure

In both years, all data were collected at the end of the academic year. In order to obtain the most accurate information, the data from students were collected through a questionnaire in May, in the last week of the Second Term, before the Final exams, based on their observations throughout the year. In order to prevent any bias and provide an atmosphere in which students could freely comment on aspects of the teachers' performance, the questionnaire was administered by assistants rather than teachers. The assistants explained the purpose of the questionnaire as being aimed at collecting data to support the improvement of teacher/teaching effectiveness and quality, and answered questions regarding the questionnaires. The data from coordinators were also collected through the same questionnaire based on their observations conducted throughout the year. In both years, the same coordinators were assigned to the same group of teachers (one coordinator in charge of 11 teachers) in order to conduct in-classroom observations throughout the academic year. Before the observations, coordinators participated in a series of workshops to ensure uniformity of focus and methodology. They drew up specifications regarding observations such as the criteria to be used, teachers to be observed, general description, timing, conditions, and behaviours to be observed. The observations, including pre and post observational sessions, were carried out over the course of the academic year twice or three times, as deemed necessary. Coordinators worked cooperatively and exchanged ideas on possible problems and specific teachers. In the second year, the same process was followed.

RESULTS

The purpose of this study was to identify how far the SET evaluations and the evaluations by coordinators correspond, whether there is a consistency between the evaluations of the two groups over the two years period and to draw some conclusions regarding how SET should be used. In order to discover whether there was a significant relationship between the evaluation ratings of instructors given by the students and the coordinators, the overall means were correlated.

A Pearson's correlation addressed the relationship between the mean ratings for teachers by the students (M = 4.12, SD =.56) and by the coordinators (M = 4.7, SD =.46). For an alpha level of .01, the correlation between the ratings was found to be statistically significant, r (99) =.45. This indicates that ratings for teachers given by the students and given by the coordinators were positively related. Another correlation was employed in order to determine the consistency of the relationship over the two years. For this reason, the association between the ratings was also investigated for the second year. The correlation between the ratings of the students (M = 4.38, SD =.44) and coordinators (M = 4.80, SD =.30) was also significant, r (99) =.43, p<.001.

DISCUSSION

That there is a significant relationship between the two groups over the two-year period can be considered a positive indication of the validity and reliability of SET, consistent with the results of similar studies conducted in the field. Theall et al. (2001) and Renaud and Murray (2005) conclude that students take evaluations seriously, that SET provides a reliable, valid assessment of instructor's teaching effectiveness and that ratings for any given instructor are reasonably stable and consistent across courses and years. One of the major criticisms of SET regarding that students can not make accurate judgements until a certain period of time has elapsed after the course, as pointed out by McKeachie (2006), does not seem as a negative effect in this study as the results indicate a significant relationship between the ratings given by both groups. The reason why a significant relationship was found between the ratings given by both the students and the coordinators over the two-year period could be the result of a careful methodology followed. This study was structured considering the drawbacks mentioned in the literature review (Cooper & Simonds, 2007; Murray, 1997; Jones, 1989). In this study, the purpose of the evaluation was well explained to the students; the questionnaires were completed in a regular class hour at the end of the term before the final exams; observable characteristics were covered in the questionnaire to ensure accurate evaluation; the importance of the ratings was discussed with students; the questionnaires were administered by assistants to ensure fair evaluation and to avoid judgment; and items regarding personality were excluded. Such precautions could be contributing factors to the positive relationship between the students and coordinators' evaluations found in this study.

The third purpose of the study was to inquire about what use SET should be put to. The increasing use of SET means that the ways in which such evaluations should be used has become a significant issue in the world of education. Although the results of this study indicate a significant relationship between the ratings given by both groups –students and coordinators-, SETs should be one of the evaluations used, but not the only one, if they are used for the purpose of faculty personnel decisions such as appraisal system. As pointed out by several researchers, student ratings should be combined with data collected from different sources such as classroom observations and self-evaluations since SET used alone would not provide sufficient reliability (Ory & Ryan, 2001; Murray, 1997). Such arguments point to one important conclusion that data about a teacher specifically for appraisal purposes should be collected from a variety of sources to ensure that teacher appraisal process is seen to be fair, and not based solely on the views of one individual's classroom observation of a teacher, or a student's judgment. However, where the purpose of evaluation is developmental rather than for assessment purposes, SET by itself may be appropriate. A pattern of low student ratings could be usefully brought to the teacher's attention. If a negative response pattern is identified, such as classroom management, giving

instruction, and using technology, this could become an area for future professional development, agreed between a teacher and a supervisor (Murdoch, 2000). It could be appropriate and useful for teachers to gather feedback themselves via questionnaires on their teaching and classroom management; this could lead teachers to self-reflection, self criticism and self-awareness. This process can provide the opportunity for making the necessary adjustments to their teaching and management.

The results of this study were considered within the SFL as part of the administration's restructuring of performance evaluation and teacher development policy. Data collected from students and coordinators were used for developmental purposes. After teachers were informed of the students evaluation results at the end of the academic year, all the teachers were asked to write an action plan for the following year and submit a copy to the trainers and coordinators, so that the points mentioned could be followed up in the next academic year. In addition, trainers and coordinators cooperated on a new observation and post-observation programme for the following academic year. Especially those teachers with lower ratings were asked to focus on areas in need of improvement mentioned by students, and by the trainers/coordinators who had worked closely with those teachers throughout the academic year. At the end of the following academic year, those teachers with areas for improvement were asked about the actions they had taken to develop themselves in respect to these areas. The results of the student evaluations were also used for revising the current program, updating goals and developing the curriculum. This gave an opportunity for the administration to review its policy, strategy and approaches since school culture has an impact on teacher effectiveness and thus teaching.

CONCLUSION AND RECOMMENDATIONS

The results of this study suggest a significant relationship between the ratings of students and coordinators and the results of two years' evaluation by the two groups are consistent and confirm that SETs are capable of providing instructors and administrators with useful feedback. This study is significant in terms of revealing the relationship between student and coordinator ratings, which indicate that students are in fact able to make judgments which accord with other judgments. However, the purpose of SET should be well identified. No matter what the purpose is, we should be wary of using such an evaluation by itself for administrative decision-making purpose such as promotion, dismissal, and tenure without reference to other evaluation methods. On the other hand, SET seems to be a reliable source for improvement in the quality of teaching, as long as the improvement process is well-structured, for example, by including action plans, observations and post-observations. This can raise the self-awareness, self-reflection and critical thinking skills of teachers. SET can also be used for material and program development by taking student perspective into consideration.

The scope of this study did not include the investigation of the relationship between student grade expectations and the students' rating of instructor effectiveness or between students' learning and instructor effectiveness in terms of low/high exam scores and mean ratings. As well as the investigation of the relationship between grading and rating, a follow-up could explore the students' written comments in a more qualitative study. These issues could be considered in future as an extension of the work done in this study

In conclusion, although SET remains controversial, data from students can be extremely illuminating. Student evaluation of teaching certainly has value and is worth the effort taken to collect it, but it would be a mistake to assume that student evaluation provides a complete assessment of all important aspects at the level of college or university teaching.

International Journal on New Trends in Education and Their Implications
October 2012 Volume: 3 Issue: 4 Article: 08 ISSN 1309-6249

BIODATA AND CONTACT ADDRESSES OF AUTHORS

Evrim ÜSTÜNLÜOĞLU is an assistant professor and the Director of the School of Foreign Languages, at the İzmir University of Economics, Turkey. She received her MA on TEFL-teaching English as foreign language-, and her Ph.D. on Educational Sciences. She has articles published in national and international journals. Her research interest is new approaches in teaching, program development, and teacher training.

Assist. Prof. Dr. Evrim ÜSTÜNLÜOĞLU
İzmir University of Economics
Balçova-Izmir- TURKEY
E. Mail: evrim.ustunluoglu@ieu.edu.tr

Seda CAN is an assistant professor at Gediz University, Department of Psychology, Turkey. She received her MA on Measurement and Evaluation and Ph.D. on Psychometrics. Her research interests are structural equation modeling and multilevel structural equation modeling. She has articles published in these areas.

Assist. Prof. Dr. Seda Can
Gediz University
Department of Psychology
Seyrek-Menemen/İzmir- TURKEY
E. Mail: seda.can@gediz.edu.tr

REFERENCES

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. Assessment and Evaluation in Higher Education , 28 (1), 71-88.

Cooper, P., & Simonds, C.. (2007). Communication for the classroom teacher. Boston: Pearson Education, Inc.

Erdle, S., Murray, H. G., & Rushton, J. P. (1985). Personality, classroom, behavior, and college teaching effectiveness: A path analysis. Journal of Educational Psychology, 77, 394-407.

Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers and courses. Research in Higher Education, 6, 223-274.

Jones, J. (1989). Students' ratings of teacher personality and teaching competence. Higher Education, 18, 551-558.

Koç, N., & Coşkuner, T. (2007). "Öğrencinin Öğretimi Değerlendirmesi". XVI. Ulusal Eğitim Bilimleri Kongresi 5 – 7 Eylül 2007: Bildiriler 1.Cilt: Tokat: Gaziosmanpaşa Üniversitesi Yayını, 226-231.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388.

McKeachie, W. J. (2006). Teaching tips. 12th Ed. Boston: Houghton Mifflin Company.

Murdoch, G. (2000). Introducing a teacher-supportive evaluation system. ELT Journal, 54 (1), 54-64.

Murray, H. G. (1997). Does evaluation of teaching lead to improvements in teaching? The International Journal of Academic Development, 2 (1), 8-23.

Ory, J. C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? In M. Theall, P. Abrami, and L. Mets (Eds.), The Student Ratings Debate: Are they Valid? How can we best Use Them? New Directions for Institutional Research, 109, San Francisco: Jossey-Bass.

Renaud, R. D., & Murray, H. G. (2005). Factorial validity of student ratings of instruction. Research in Higher Education, 46, 929-953.

Ryan, J. J., Anderson J. A., & Birchler, A. B. (1980). Student evaluation: the faculty responds. Research in Higher Education, 12, 317-333.

Theall, M., Abrami, D. A., & Mets, L. (2001). The student ratings debate: Are they valid? How can we best use them? New Directions for Institutional Research No. 109. San Francisco: Jossey-Bass.

Theall, M. &, Franklin, J. (2001). Using technology to facilitate evaluation. New Directions for Teaching and Learning, 88, 41-50.

Yurtkoru, E. S. &, Sipahi, B. (2003). "Öğretim üyesi performans değerlendirme kriterinin cinsiyete göre belirlenmesi üzerine analitik bir çalışma", İstanbul Ticaret Üniversitesi Dergisi, 13-37.