

Myths And Facts About Student Surveys Of Teaching The Links Between Students' Evaluations Of Faculty And Course Grades

Nitza Davidovitch, Ari'el University Center, Israel

Dan Soen, Ari'el University Center and the Graduate School, Kibbutzim School of Education, Tel-Aviv, Israel

ABSTRACT

The present study sought to examine the justification of faculty claims regarding bias in students' assessments of faculty performance that stem from external factors which do not include the quality of their teaching. Specifically, we sought to examine the hypothesis that there is a correlation between lecturer ranking and grades given by lecturers and between lecturer rankings, grades, and background variables. The framework of the research is the combination of three different stages: faculty, course, lecturer and the statistical manipulation, creating a complex image of reality and thereby offering an answer to the most classical question in the research literature. Findings of this study indicate that the alleged correlation between the students' grades and the lecturers is non-existent, and nothing but a myth amongst the academic body. However, the research still points out that there are some additional elements which are beyond the efficiency of teaching as we tap into different levels of interaction between student and lecturer.

Keywords: Course grades, students' evaluation of faculty, faculty performance, interaction between student and teacher, students' surveys.

PREFACE

Faculty evaluation questionnaires first appeared in the academic world at the beginning of the twentieth century, however their status and significance increased particularly from the 1970s (Hativa, 2008). In time the questionnaires were accepted as an efficient and inexpensive tool for evaluating teaching efficiency and a means for reaching operative conclusions concerning the improvement of teaching quality and evaluating faculty work. The use of questionnaires is particularly prevalent in the United States and Canada, where the emphasis is on turning teaching into a recognized profession and setting strict standards of knowledge and skills in the field of teaching (Hativa, 2003; Hativa, 2002). The evaluation questionnaires are perceived as a means of achieving these goals and they led European countries as well to adopt the method of evaluation despite the existing arguments concerning the questionnaires as a valid basis for applicable decisions (Reinderman & Schofield, 2001).

Initially, questionnaires served as a feedback tool with no far-reaching implications. However, in time, the assessment questionnaires acquired status and from feedback tools per se providing recommendations for improvement they became tools whose results were significant for academic careers – including tenure, promotion, and academic appointments (Ehie & Pararhanos, 1994; Harrison et al., 2004; Smith et al., 2004; Williams & Ceci, 1997).

These aspects inevitably influence salaries, perceived prestige, and career development possibilities. The results of the questionnaires receive much significance and provide a measure for the use of supervisory elements. The great dominance of evaluation questionnaires, known by some as the “assessment mandate” (Chandler, 1978) dictates and determines lecturers' career development to a great degree.

Due to these implications the surveys are a target of much criticism by the academic faculty and a subject of many studies which seek to examine the validity and reliability of the surveys, the degree to which students' perceptions should influence the career of academics. The perceptions can be divided in two groups – supporters and objectors. Some claim that this argument is the monumental argument in the history of teaching research.

On the one hand are those who see the mere existence of the survey as a judgmental means which is forced on the staff and damages their academic status (Avdor, 2006). Supporters of this approach feel that they are supposed to “deliver the goods” posed by their students, who credit them in return. This perspective, some say, stems from a lack of dialogue between the faculty and their supervision, between and faculty and students (ibid.). On the other hand there are those who are in favor of the surveys and recognize the importance of the feedback questionnaires as a means of promoting and improving the quality of teaching as well as the satisfaction of the student-consumer (Marsh, 1987; Marsh & Roche, 1994; Ramsden, 1991).

Despite the disagreement, researchers agree on the significance of promoting “efficient teaching”, improving the quality of teaching, and reducing the number of dropouts (National Survey, 2002). Both governments and universities attempt to develop various practices and policies aimed at encouraging, rewarding, and measuring “good teaching” (Ballantyne et al., 2000). Thus the disagreement does not focus on the need for control, but rather on its manner of performance and its degree of effectiveness. Therefore we must ask whether the academic world is excessively reliant on evaluation surveys? Are the surveys valid and reliable? Or are these reciprocal manipulations performed by faculty in order to receive high rankings and by students in order to receive high grades?

Many faculty members are opposed to the surveys (Hativa, 2008), claiming that the latter are biased for various reasons which include areas that are not under the control of the faculty, such as *Course characteristics*: level of difficulty, material mass (Wachtel, 1998; Feldman, 1978; Chen & Hoshower, 1998); *Characteristics of the lecturer*: appearance, sex, age, ethnicity (Anderson & Siegfried, 1997; Wachtel, 1998); *Characteristics of the student*: personal expectations, sex, motivation, class attendance (Davidovich & Notzer, 2004; Koerner & Petelle, 1991; Tatro, 1995; Anderson & Siegfried, 1997; Chen & Hoshower, 1998) and *Administrative characteristics*: class size, number of classes (Feldman, 1978; Chen & Hoshower, 1998; Wachtel, 1998).

Each of these factors is perceived as a potential cause of bias which might affect the ranking of the lecturer's teaching since teaching is not devoid of context and does not isolate the lecturer's skills and capabilities from other factors active in the student-lecturer-class interaction. This is the logical basis of the faculty claim that student feedback is not an efficient means of evaluating the quality of their work and all the more so as a means of dictating their future.

In the present study we sought to examine the justification of faculty claims regarding bias stemming from external factors which do not include the quality of their teaching. Specifically, we sought to examine the hypothesis that there is a correlation between lecturer ranking and grades given by lecturers and between lecturer rankings, grades, and background variables.

The topic of staff evaluation is the most widely studied subject in the field of evaluation methods. There are more than 2,000 studies in the ERIC reserve alone (Centra, 2003). However despite the many studies most of them are generic and do not deal with specific qualities of the teaching context. Research has examined teaching from an interdisciplinary point of view, focusing on the universal aspect of teaching (Hativa, 2003a). This matter is very significant as focusing on a certain domain or specific discipline enables intensification of the empirical meaning of the research (Worthington, 2002).

In order to give the present study some distinction, we chose to discern the various faculties and analyze them on three levels: student evaluations of staff, staff evaluations, and evaluations of courses given by staff. We hypothesized that this would achieve better results than past studies which presented simple descriptive statistics and regular regression as means of proving correlation between staff rankings and student grades, while putting aside the complexity of relationships between variables (ibid.).

STAFF EVALUATION AND STUDENT GRADES – THE LINK THEORY

One of the most prevalent claims against evaluation surveys states that staff rankings are no more than reflections of students' satisfaction with their anticipated grades (Remedios & Lieberman, 2008). This claim is well-known and sometimes independent of research findings. However many studies have attempted to provide an unequivocal response to these claims by presenting the validity and reliability of the surveys.

A review of the research from the 1970s to the present does not produce an unambiguous result regarding the existence or absence of a correlation between staff rankings and students' perceived grades (Centra, 2003). For example, Howard and Maxwell (1982) did not succeed in clearly determining that the cause of correlation between students' anticipated grades and staff rankings is a non-mediated correlation. This can be explained in terms of satisfaction with the level of teaching, leading to high grade expectations, or in contrast it is possible to claim that expectations of high grades led to high satisfaction and high rankings.

d'Appolonia and Abrami (2007) claimed that there is indeed a positive correlation between grades and rankings, however this does not indicate decreased grades or excessive generosity of the staff as a means of increasing rankings. They claim that the relationship between the (student's test) grade and the ranking (given as student feedback on staff) is problematic only when not compatible with students' work. The researchers claim that if students' evaluation of the staff reflects their work for the course then this correlation is natural and inevitable. However, if the opposite is true, we expect to see a correlation between low amount of work and higher expectations of grades.

Bortetz (2004) also claims that student evaluations do not cause "grade inflation" stemming from staff attempts to "bribe" students with high grades in order to receive high rankings (Kfir et al., 2003). She believes that the concept of inflation is used unnecessarily regarding the academic world and that students are not clients who require higher grades rather an integral part of the system. The correlation between grades and rankings led the researcher to conclude that students achieve higher grades due to improved communication between staff and students. This improvement is one of the results of evaluation surveys which facilitate improvement of staff performance and the amount of student's work.

Krautman and Sander (1999) also found a positive correlation between high grades and high staff evaluations. Their findings led them to conclude that this correlation inevitably indicates that the survey is an inappropriate means of staff evaluation.

Many other studies found a positive correlation between students' grade expectations and staff ranking (Marsh, 1987; Braskamp & Ory, 1994; Marsh & Dunkin, 1992; McPheson & Jewell, 2007). Most of the early research reports on a relatively moderate correlation with a mean of .20 (Centra, 2003). In later studies the correlation is in the range of .10 to .30 (Feldman, 1997). This correlation, although relatively moderate, is significant and demands explanation.

Thus, many studies present a positive correlation, however when assessing the importance and implications of this correlation it is necessary to relate to the problems inherent in forming a unidimensional model. Such a model must relate to the two variables of grade/rating without relating to mediating variables which inevitably have an independent direct interaction.

The positive correlation seen in studies mainly from the '70s to '90s was attributed primarily to the failure of the survey itself to predict what it was intended to predict – the quality of teaching (=its validity). Researchers saw the correlations as proof of a prevalent myth in universities (Hativa, 2008) – that the survey results were biased.

Despite the attempt to justify the results by attributing the causality to the survey, there is room to examine the correlations from a wider perspective, as presented by later studies. Studies from the last decade and a half, offer examination of a multi-variable model, regarding the grade-rating relationship.

For example, Wright and Palmer (2006) presented **three different models** aimed at explaining the relationship between evaluating ratings and student grade expectations. **The first model** is the simplest model, presenting a linear relationship between grade expectations and staff evaluations, **the second model** presents relationships between study efforts, grades, and staff ratings, while **the third model** is the most complex and explains the relationship by including motivation, student abilities, study efforts, and grades, as factors affecting staff evaluations. Each of the models was examined versus the statistical database.

The researchers found that the third model was most compatible with the data and explained the relationships between the variables most cogently. The compatibility was so significant that the hypothesis could not be rejected (ibid.). The second model was completely incompatible with the data and thus the researchers concluded that it does not explain the relationships between the variables. The linear model, which is still prevalent among academics (Hativa, 2008) was not compatible with the data either. The researchers concluded that students do not tend to rate lecturers higher only because they believe that they will receive higher grades. However they also do not tend to rate lecturers higher because they believe that the course widened their horizons. The researchers concluded that the relationship between grades and teaching evaluation presents a complex system of relationships which includes elements of motivation, personal ability, quantity of learning as perceived by the student.

Some suggest assessing survey results in a multidimensional statistical method (“Generalizability of multidimensional”) which presents the relationship between staff, students, and external factors (Rindermann & Neville, 2001). In this method researchers found that the evaluation of efficient teaching includes the staff, however their influence is part of a wideR equation which includes the student, conditions of interaction, and external conditions. Examination of all the influential factors turns the survey into a reliable and valid tool which enables evaluation of the quality of lecturers’ teaching.

Comprehension of the complexity of relationships and the various influences that may exist in students’ rating of staff, created hypotheses regarding the effect of other factors aside from grades. These factors, termed background characteristics, were also suggested as potential causes of survey results bias and invalid evaluation of staff quality.

BACKGROUND CHARACTERISTICS AND STAFF EVALUATION – MYTHS AND REALITY

Many studies have been published under the promising title of regulation and influences on student ratings, with suggestions and proposals for solving potential bias (Baldwin & Blattner, 2003). However while some attempt to present defensive means other suggest examining those factors which cannot be changed by the individual – characteristics such as age, ethnicity, sex. These demographic data might also theoretically take part in students’ evaluation of faculty.

GENDER

In a meta-analysis conducted by Jones and Dindia (2004), which included 32 studies sampled over a period of thirty years, *students’ gender* was found to have an effect on the student-lecturer interaction and on students’ perception of the lecturer. The researchers found that male students have more interactions with faculty in general and also have more negative interactions with faculty than female students. In the classroom, gender roles form a significant part of the interaction, as males dominate the classroom.

Another correlation was found between the *lecturers’ gender* and students’ participation in class. Male lecturers facilitate more participation of male students, however this effect does not exist in the case of female lecturers. These correlations were found in most of the studies indicating the existence of an interaction between the gender of the lecturer and the gender of the student (Hutchinson & Beadle, 1992; Dindia & Jones, 2004).

Despite the anticipated correlations, there is disagreement concerning the effect of student/lecturer gender on staff evaluation (Kelly, 1988; Dindia & Jones, 2004). Some claim that the interaction itself is very moderate and does not indicate a gender effect on staff evaluation (Aleamoni, 1999; Feldman, 1993; Fernandez & Mateo, 1997; Freeman, 1994; Wheelless & Potorti, 1989).

Some claim that the difficulty in reaching unambiguous conclusions is connected to the problem of isolating and neutralizing gender effects on staff evaluations (Spragu & Massoni, 2005). However the researchers disagree and some present findings that support the effect of staff gender on student evaluations. For example, Spragu and Massoni (2005) found congruence in students' remarks on male and female lecturers and that students' expectations change as a function of lecturers' gender. Lecturers who do not meet gender expectations will receive lots of hostility from students. The general burden of meeting gender expectations is born by both sexes, however women are required to prove themselves more than men and must work harder to meet student expectations (Spragu & Massoni, 2005).

FACULTY SENIORITY, RANK, AND AGE

Many longitudinal studies performed in various departments found that the quality of teaching decreases as a function of seniority. The research has found that lacking external interference there is a moderate but steady decrease in the quality of teaching (Ryans, 1960; Barnes, 1985). The researchers report a negative correlation between teaching experience and teaching efficiency. Seniority, age, and rank are usually examined together since they clearly overlap.

Feldman (1983) examined the manner in which the three variables affect faculty evaluations. The research results indicate a negative correlation between age and faculty ratings and between seniority and faculty ratings. These findings were found to be compatible with more recent studies which also found negative correlations and an inverse relationship between seniority and age on the one hand and the quality of teaching on the other (Blackburn & Lawrence, 1986; Feldman, 1997; Horner, Murray & Rushton, 1989; Marsh, 1987, 2007; Marsh & Dunkin, 1997; Marsh & Hocevar, 1991a; Murray, 1990, 1997; Renaud & Murray, 1996). Feldman (ibid.) also found a positive correlation between rank and staff evaluations, where higher rank indicated higher student evaluations.

Marsh (2007) claimed that the findings presented above are not a correct reflection of reality, as they are based on mainly cross-sectional data gathering. In order to support his claim he conducted a study presenting a new statistical method by which he performed a meta-analysis of the results of evaluation surveys of 195 lecturers and more than 6,000 classes and data gathered over 13 years. His study found that most of the staff evaluations remained constant and that staff remained at about the same efficiency and level of teaching.

ETHNICITY

The effect of ethnicity on student evaluations of staff is a field that has been neglected relative to other background characteristics. This variable is usually examined together with that of appearance or gender. Most of the studies examine this variable as secondary to other background elements perceived as more central (Anderson and Siegfried, 1997; Watkins, 1994). Lackritz & Ghorpade (1991) found that age and ethnicity caused bias in lecturer evaluations and present a degree of discrimination among students. The studies find a positive correlation, however it has not been examined sufficiently in order to determine the existence of effects of ethnicity. As of the present, there is no comprehensive study examining the effect of ethnicity on faculty evaluations (Worthington, 2002; Centra, 1993).

RESEARCH QUESTIONS

We chose to deal with this myth of the relationship between grades and ratings at the Ariel University Center (AUC). This study is based on student evaluations of faculty during the second semester of the 2007-2008 school year and their grades. We examined the existence of a direct relationship between students' grades in a specific course and their evaluation of the lecturer. Are there other variables which have an effect on the grade?

The relationship between these variables may be examined by using data gathered in teaching surveys. However since teaching surveys are anonymous, the coordination between the rating given by students to the lecturer and their grades was examined by an external firm ("Rensis") which performed the computerized teaching survey. The researchers used the data in an anonymous and collective manner and the students knew in advance that the data would be used ensuring their anonymity.

FINDINGS

The study included 16,484 student evaluations of 434 lecturers and 1,033 courses given by various lecturers. Since the very large number of evaluations, of both lecturers and courses, when examining the differences or correlations, we will refer only to significant results from the $p < .01$ level.

We shall present the data in reference to student evaluations of lecturers (each student evaluated different lecturers on different courses), in reference to the lecturers (each lecturer was evaluated by a group of students on the different courses given) and in reference to the course given by the lecturer (each course given by each lecturer was evaluated by a group of students).

In summary, in this study we performed regressions of instructor scores on distinction criteria in each field and student grades on each of the four instructor evaluations. In addition, we included score-grade interactions in these regressions and we found two interactions.

Aside from specific points, no correlations were found between student grades and instructor evaluations. The purpose of the study was to find these relationships.

Almost no differences were found in the correlations examined for various groups of instructors (by gender, origin, tenure, etc.). Entering these variables in regressions and the interactions between variables show that the correlation is varied. In addition, no varied correlations were found for the other faculties.

Nonetheless, we performed possible regressions in order to examine their possible contribution to the study. Five regressions were performed:

1. **Explanation of grades** according to 4 evaluations of instructors, instructor gender, age, tenure – scoring on distinction criteria, and seniority. **Only instructor gender was found to contribute to explaining grade variance.**
2. **Explanation of general evaluation of course instructors** according to grades, instructor gender, age, tenure, scoring on distinction criteria, and seniority. None of the variables contributes to explanation of the variance.
3. **Explanation of evaluation of course structure and design** according to grades, instructor gender, age, tenure, scoring on distinction criteria, and seniority. **Instructor age and tenure contribute to explanation of the variance.**
4. **Explanation of evaluation of lecture clarity** according to grades, instructor gender, age, tenure, scoring on distinction criteria, and seniority – Instructor gender and age contribute to explanation of the variance.
5. **Explanation of the evaluation of instructor attitude towards students** according to grades, instructor gender, age, tenure, scoring on distinction criteria, and seniority. None of the variables contributes to explanation of the variance.

CONCLUSION

The current research project, in many ways, only represents another drop in the sea of research that aims at mapping the relationship between private variables and value variables.

In addition, this research is rather successful at shedding some light on the manifestation from a different angle.

The framework of the research is the combination of three different stages: faculty, course, lecturer and the statistical manipulation, creating a complex image of reality and thereby offering an answer to the most classical question in the research literature.

Based on the significant high levels, we can see that the correlation between the students' grades and the lecturers are non-existent. This is nothing but a myth amongst the academic body and is not based on reality.

However, the research still points out, just as one breaks or dismantles a myth, that there are some additional elements which are beyond the efficiency of teaching as we touch different levels of interaction between student and lecturer.

The attitude of the lecturer, for example, becomes a key element which impacts the course. This is an important moral for all those who are in the business of teaching that efficient “pure” teaching does not only rely on transferring the material but also on the general attitude of the lecturer to the people in front of him. On this matter, it is possible that the statement of lecturers that their evaluation is not based on their teaching skills is correct – students are not only influenced by the academic possibilities of a lecturer but also from his ability to interact with them.

On the same level, there are other demographic elements such as sex and ethnic background which have a negative influence on the students’ valuation. In this case, the students are nothing but part of a cultural concept which is based on discrimination against women, racism against foreigners (mainly from the former USSR). Therefore we cannot claim that the students are not a mirror of society and represent what is happening in terms of prejudice and reality bites.

These results give our research a certain important weight where the central assumption was refuted and together with that the doubts of the faculty of the Ariel College – Students are human beings who are influenced by society stigmas but they do not project their personal successes or lack of on the skills of the lecturer. On the contrary, there seems to be a certain skill to separate between the interior and the exterior. However, this skill is not available to the student when he stands in front of a society larger than him – sexual and ethnic discrimination.

AUTHOR INFORMATION

Dr. Nitza Davidovitch holds a doctorate degree, awarded by Bar Ilan University, focused on the developmental trends of regional colleges and their impact on the higher education system in Israel. Dr. Davidovitch currently heads the Department of Academic Development at Israel's largest public college, Ari'el University Center. Dr. Davidovitch is active in community services in issues relating to the Holocaust and Jewish identity, cultivating the cultural history of Jewish sects, and moral education.

Prof. Dan Soen is currently Head of two Departments at the Ariel University Center: The Multi-Disciplinary Dept. for Social Sciences & Humanities, and the Department of Sociology & Anthropology. Dr. Soen is also teaching at the Graduate School, Kibbutzim School of Education, Tel-Aviv. He has been associated with most of the Israeli universities along the years, and also worked as private consultant in social planning as well as regional and urban planning. He worked on projects in Israel, Asia and Africa. He published extensively in Hebrew, English and French and edited and wrote about 30 books.

REFERENCES

1. Anderson, K.H & Siegfried, J.J. (1997) Gender differences in rating the teaching of economics , *Eastern Economic Journal*, 23(3), 347–357.
2. Avador, S. (2006). For whom and for what? Bias factors in student's evaluations at educational colleges and the question of benefit for the evaluated and the organization, *Dapim*, 41, 10-37 (Hebrew version).
3. Babad, E, Icekson, T., & Yelinek, Y.(2008). Antecedents and Correlates of Course Cancellation in a University "Drop and Add" Period. *Research in Higher Education*, 49(4), 293-316.
4. Baldwin, T. & Blattner, N. (2003). Guarding against potential bias in student evaluations. *College Teaching*, 51(1), 27-33.
5. Ballantyne, R., Borthwick, J. & Packer, J. (2000). Beyond student evaluation of teaching: identifying and addressing academic staff development. *Assessment & Evaluation in Higher Education*, 25(3), 221 – 236.
6. Barnes, B. (1985). *About Science*. Oxford: Basil Blackwell.
7. Blackburn, R. T. & Lawrence, J. H. (1986). Aging and the quality of faculty job-performance. *Review of Educational Research*, 56(3), 265–290.
8. Boretz, E. (2004). Grade Inflation and the Myth of Student Consumerism. *College Teaching*, (2)52 ,42-52.

9. Bowling, N. A. (2008). Does the Relationship between Student Ratings of Course Easiness and Course Quality Vary across Schools? The Role of School Academic Rankings. *Assessment & Evaluation in Higher Education*, 33(4), 455-464.
10. Boysen, G. A. (2008). Revenge and Student Evaluations of Teaching. *Teaching of Psychology*, 35 (3), 218-222.
11. Centra, J. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44(5), 495-518.
12. Chandler, T. (1978). The questionable status of student evaluations of teaching. *Teaching of Psychology*, 5(3), 150–152.
13. Chen, Y. & Hoshower, L. B. (1998) Assessing student motivations to participate in teaching evaluations : an application of expectancy theory, *Issues in Accounting Education*, 13(3), 531–549.
14. d'Apollonia, S. & Abrami, P. C. (1997). Navigating student ratings of instruction. *The American Psychologist*, 52 (1111), 1198-1208.
15. Davidovitch, N. (2004). Students' attendance and scores of teacher's evaluations, *Al-Agova*, 3, 38-41 (Hebrew version).
16. Ehie, I. & Karathanos, D. (1994). Business faculty performance evaluation based on the new AACSB accreditation standards. *Journal of Education for Business*, 69 (5), 257-262.
17. Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 368–395). New York: Agathon.
18. Feldman, K. A. (1983). The seniority and instructional experience of college teachers as related to the evaluations they receive from their students. *Research in Higher Education*, 18, 3–124.
19. Feldman, K. A. (1978) Course characteristic s and college students' ratings of their teachers : what we know and what we don't. *Research in Higher Education*, 9(2), 199–242.
20. Fernandez, J. & Mateo, M. (1997). Student and faculty gender in ratings of university teaching quality. *Sex Roles*, 37(11/12), 997-1003.
21. Freeman, H. (1994). Student evaluation of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology*, 86(4), 627-630.
22. Ghorpade, J. & Lackritz, J. R. (1991). Student evaluations: equal opportunity concerns. *Thought and Action*, 7(1), 61-72.
23. Harrison, P., Douglas, D., & Burdsal, C. (2004). The relative merits of different types of overall evaluations of teaching effectiveness. *Research in Higher Education*, 45(3), 311-323.
24. Hativa, N. (2008). Myths and facts about evaluation surveys by students, *Al-Agova*, 7, 13-14 (Hebrew version).
25. Hativa, N. (2003a). Do teachers at humanities and art are better than their colleagues at management and engineering? *Al-Agova*, 2, 38-40 (Hebrew version).
26. Hativa, N. (2003b). what teachers should know and do- student's reform at teaching, *Hed-Hachinoch*, 6, 14-19 (Hebrew version).
27. Hativa, N. (2002). New thought on university and college teaching: teaching as proficiency, *Al-Agova*, 1, 6-9 (Hebrew version).
28. Howard, G. S. & Maxwell, S. E. (1980) Correlation between student satisfaction and grades: a case of mistaken causation? *Journal of Educational Philosophy*, 72(4), 810–820.
29. Hutchinson, L. M. & Beadle, M. E. (1992). Professors' communication styles: how they influence male and female seminar participants. 8(4),405-18.
30. Jones, S. & Dindia, K. (2004,). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research*, 74(4), 443-471.
31. Kfir, D., Frasko, B., & Paul- Benjamin, A. (2006). Professional dues and grading inflation in higher education, *Megamot*, 42(2), 296-313 (Hebrew version).
32. Koermer, C. D. & Petelle, J. L. (1991). Expectancy violation and student rating of instruction. *Communication Quarterly*, 39(4), 341–350.
33. Krautmann, A. C. & Sander, W. (1999). Grades and Student Evaluation of Teachers. *Economics of Education Review*, 18(1), 59–63.
34. Kulik, J.A. (2001). Student Ratings: Validity, Utility, and Controversy. *New Directions for Institutional Research*, 109, 9–25.

35. Marsh, H. W. & Roche, L. A. (1994). The use of students' evaluations of university teaching to improve teaching effectiveness. *Final project report for the Evaluations and Investigations Program of the Department of Employment and Education*. Canberra: Australian Government Printing Service.
36. Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253–288.
37. Marsh, H. W. & Cooper, T. L. (1981). Prior subject interest, students' evaluations, and instructional effectiveness. *Multivariate Behavioral Research*, 16(1), 83.
38. McPherson, M., & Jewell, R. (2007). Leveling the Playing Field: Should Student Evaluation Scores be Adjusted? *Social Science Quarterly*, 88(3), 868-881.
39. National survey seeks to improve retention, graduation rates. (2002). *Black Issues in Higher Education*, 19(14), 18.
40. Petridou, E. & Katerina, S. (2004). Evaluation Research in Business Schools: Students' Rating Myth. *International Journal of Educational Management*, 18(3), 152-159.
41. Tatro, C. N. (1995) Gender effects on students evaluation s of faculty, *Journal of Research and Development in Education*, 28(3), 169–173.
42. Rindermann, H. & Schofield, N. (2001). Generalizability of multidimensional student ratings of university instruction across courses and teachers. *Research in Higher Education*, 42(4), 377-399.
43. Remedios, R. & Lieberman, D. (2008). I liked your course because you taught me well: the influence of grades, workload, expectations and goals on students' evaluations of teaching. *British Educational Research Journal*, 34(1), 91-115.
44. Renaud, R. D., & Murray, H. G. (1996). Aging, personality, and teaching effectiveness in academic psychologists. *Research in Higher Education*, 37(3), 323–340.
45. Ryans, D. G. (1960). Prediction of teacher effectiveness. In C. W. Harris (Ed.), *Encyclopedia of Educational Research* (pp. 1486–1491). New York: Macmillan.
46. Smith, K. & Pollak, M. W. (2008). What Can They Say about My Teaching? Teacher Educators' Attitudes to Standardised Student Evaluation of Teaching. *European Journal of Teacher Education*, 31(2), 203-214.
47. Sprague, J. & Massoni, K. (2005). Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us. *Sex Roles*, 53(11/12), 779-793.
48. Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student Evaluation of Teaching Quality in Higher Education: Development of an Instrument Based on 10 Likert-Scales. *Assessment & Evaluation in Higher Education*, 32(6), 667-679.
49. Lechuga, V. M. (2008). Assessment, Knowledge, and Customer Service: Contextualizing Faculty Work at For-Profit Colleges and Universities. *Review of Higher Education*, 31(3), 287-307.
50. Wachtel, H. K. (1998) Student evaluation of college teaching effectiveness: a brief review, *Assessment and Evaluation in Higher Education*, 23(2), 191–211.
51. Worthington, A. (2002). The Impact of Student Perceptions and Characteristics on Teaching Evaluations: a case study in finance education. *Assessment & Evaluation in Higher Education*, 27(1), 49-64.
52. Wright, R., & Palmer, J. (2006). A comparative analysis of different models explaining the relationship between instructor ratings and expected student grades. *Educational Research Quarterly*, 30(2), 3-18.

NOTES