Students' Grade Vs. Teachers' Evaluation: An Empirical Validation

Jesse S. Susada

Abstract

The importance of student evaluation for promotion leads teachers to be lenient in exchange of a good evaluation result. This study looks into the possible effects of midterm grades to the evaluation results of teachers at the end of the semester. The study utilized data mining approach to determine relationships between the two variables. Grades and evaluation results were taken from the Electronic Data Processing Center of the school. Both parametric and non-parametric analyses were employed in the study. Findings revealed a weak correlation between the midterm grades of students and their evaluation rating of teachers. Looking at grading leniency versus teacher status, it was found out that grades of full-time faculty was significantly higher compared to part time; and grades of tenured faculty was significantly higher compared to their probationary counterpart. Therefore, higher education institution administrators who use results of student evaluation as a key indicator of the effectiveness of teachers should be very careful as these results might not represent the true performance of a teacher inside the classroom.

Keywords: midterm grade, teacher's evaluation, education,

1.0 Introduction

Several studies have revealed the link between students' grades and results of student evaluation of teachers. Though results were varied, the connections between the two are one of the issues that beset almost every school, more particularly the private schools, where evaluation plays a significant role in the status of a teacher or faculty. Sander & Krautmann (1999) & Isely & Singh (2005) claimed that instructors can "buy" better evaluation through a more lenient grading system where students who expect higher grade gives more favorable evaluation of their teacher. While many literatures supported the same result, Centra (2003) and Pascale (1979) countered that after controlling all other variables, expected grades of students, generally, did not affect results of student evaluation. Therefore, there is no correlation between student's knowledge of grades and student's evaluation of teacher. Other authors have different claims over the two variables. McSpirit and Chapman (2000), in interpreting faculty comments, revealed that these opposing results in the research community reflect the contradictions on the issue among instructors. Thus, it is, therefore, the purpose of this study to validate the link between these two variables of by using midterm grades and actual results of student evaluation of teachers, and consequently, find whether the results differ according to the

employment status of a teacher.

A student grade is one indicator of his achievement inside the classroom. It measures the student's attainment and performance in all activities done inside or outside the classroom setting. However, several studies were conducted with regards to the connection of students' grades and instructor's rank. Researches on this topic have consistently revealed that lower ranking instructors gave higher grades compared to their high ranking counterpart (Johnson, A. E., Pitts, S. T., & Kamery, R. H. (2006)). In the studies of Bolge (1995), Clark (1990), Sonner (2000), Sonner and Sharland (1993), they all found that a student's grade is a function of an instructor's rank and revealed that part-time instructors assign higher grades on average than full-time instructors. The results suggest that grade inflation may be caused by teacher's status (whether fulltime or part time), and the grades given by part-time instructors put pressure on their fulltime counterpart because it sets unrealistic expectation about the grades a student must receive (Landrum, 1999). All of these bring to mind that a student grade is a function in the status of a teacher or faculty and may imply that the grade may not necessarily reflect student performance.

Another factor that possibly affects the grading is the student evaluation of teachers when the semester is over. A number of studies hypothesized that the higher grades a teacher gives to students, the more likely the teacher will receive high evaluation. While many authors were cited regarding the same results, others disagree about the extent of such problem (Wright, R. E., & Palmer, J. C. (2006)). According to d'Appolonia and Abrami (1997) and Boretz (2004), grading leniency is only a problem when it is unrelated to student's learning. As long as the grades of the student truly represent what is learned inside the classroom, there is nothing wrong with being lenient in giving grades. But whether student evaluation of teacher results is a true representative of teacher's performance inside the classroom, higher education institutions of learning are, therefore, cautioned on interpreting the result, especially, if such measure is subjective or problematic. Evaluation results are supposed to give teachers an understanding on the strengths and weaknesses of their teaching pedagogy. However, because of grade inflation by other teachers, especially part-time, this puts pressure on their full time counterpart to give high grades in order to meet student's expectations. Furthermore, the result of student's evaluation is not the only measure of teacher's effectiveness inside the classroom, as this may be useful only for specific group of students who have particular levels of motivation and ability that are in line with the instructor's expectation problem (Wright, R. E., & Palmer, J. C. (2006)).

All literatures reviewed revealed that there is a relationship that exists between the two variables, student's grade and evaluation results. However, their views were varied. While others found a correlation between grades and students' evaluation of teachers, others revealed otherwise. Most of these literatures used expected grades of students as the main independent variable because final grades are normally released at the end of the semester. The contrasting results of the two main variables in the study, led the researcher to further validate on whether or not grades of students, more particularly midterm grades (which are already known before they evaluate their teacher), will have an undue influence to students rating of teachers performance. Furthermore, this

study would also try to validate the literatures reviewed on whether part-time instructors do really give higher grades compared to their full time counterpart as well as probationary and tenured faculty. The results of this study would, then, be added to the body of knowledge to the existing studies and further validate existing theories on hand regarding grades and student's evaluation.

2.0 Conceptual Framework

This study is anchored on the principle that teachers can "buy" better evaluation results by giving high grades to students (Sander & Krautmann (1999) & Isely & Singh (2005). Although other studies may have revealed different results, this study is focused on validating, whether students midterm grade will have an undue influence on the rating of students of their teacher. Midterm grades are given in the middle of the semester. Therefore, students will already have an idea what could possibly be their final grade at the end of the semester. The study of Centra (2003) revealed that subjects that were rated by students as just right received high evaluation from students while those that were rated difficult or too elementary received a low rating. Given the importance of evaluation results to teacher's promotion and tenure, a teacher's temptation to manipulate grades just to get high evaluation from students is a big possibility. Figure A shows the conceptual framework diagram of the study.

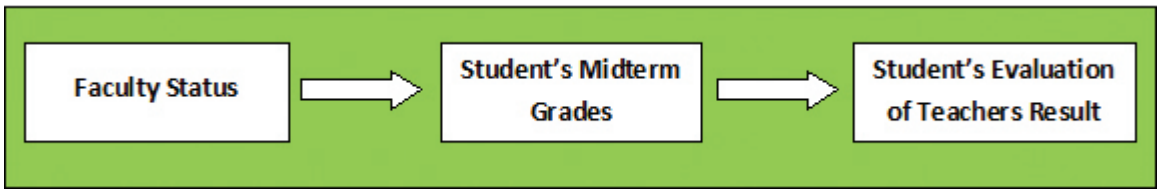


Figure A. *Conceptual Framework of the Study*

3.0 Design and Methods

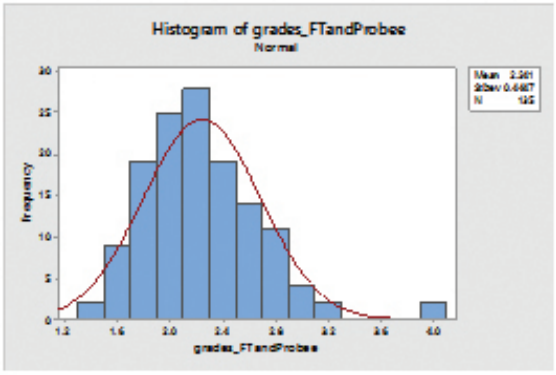
This study utilized exploratory data analysis or more commonly known as data mining. Student's midterm grades and evaluation of teachers' results were the main data used in the study. Grades and evaluation results were taken from the Electronic Data Processing center of the University. Midterm grades were used as predictors because these are the grades that students have prior knowledge before they will evaluate their teachers. Final grades are normally given after students finished their evaluation of teachers. Hence, most of the time, they are already done with their evaluation upon leaning their final grades.

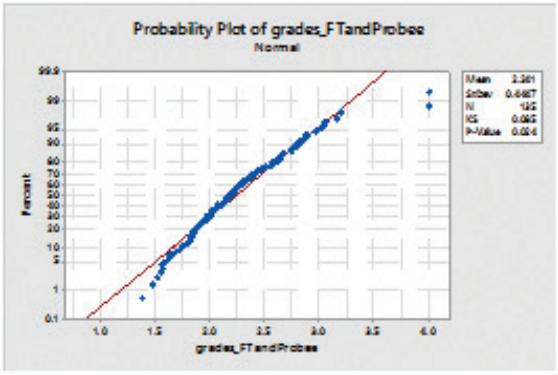
Both grades and evaluation were treated as continuous variables. Grades were measured from 1.0 to 5.0 (with 1 as the highest and 5.0 as the lowest) while evaluation results have 5.0 as the highest while 1.0 as the lowest. Since only real grades of students were available during the data gathering, the researcher took the average grades of the students in order to equate the number of samples for the two variables. After taking the average grades and evaluation results of students from each subject, a total of 262 grades and evaluation results were retrieved as subjects of the study. These data were subjected to normality tests to determine appropriate statistical analysis to be used.

4.0 Results and Discussion

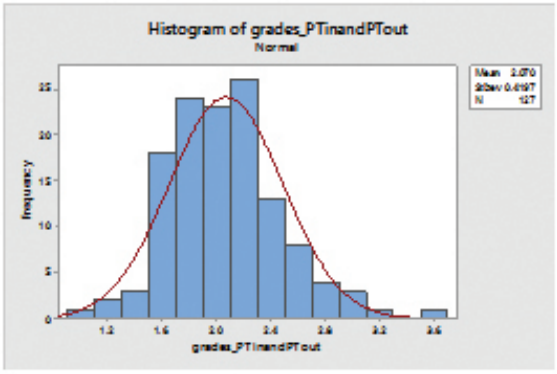
Figures 1 to 12 present the histogram and probability plot results of all variables in the study, namely: grades of full-time faculty (Figures 1 & 2) and grades of part-time faculty (Figures 3 & 4); grades of full-time tenured (Figures 5 & 6) and grades of full-time probationary (Figures 7 & 8); and grades of all faculty (Figures 9 & 10) and evaluation results (Figures 11 & 12). These presentations are necessary to identify whether data followed a normal distribution.

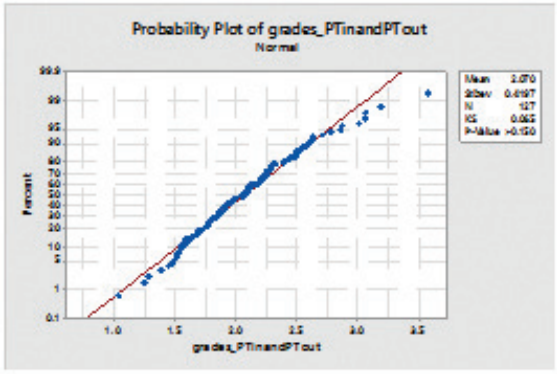
Normality Tests





Figures 1 & 2. Histogram and Normality plot of full time faculty midterm grades





Figures 3 & 4. *Histogram and Normality plot of part time faculty midterm grades*

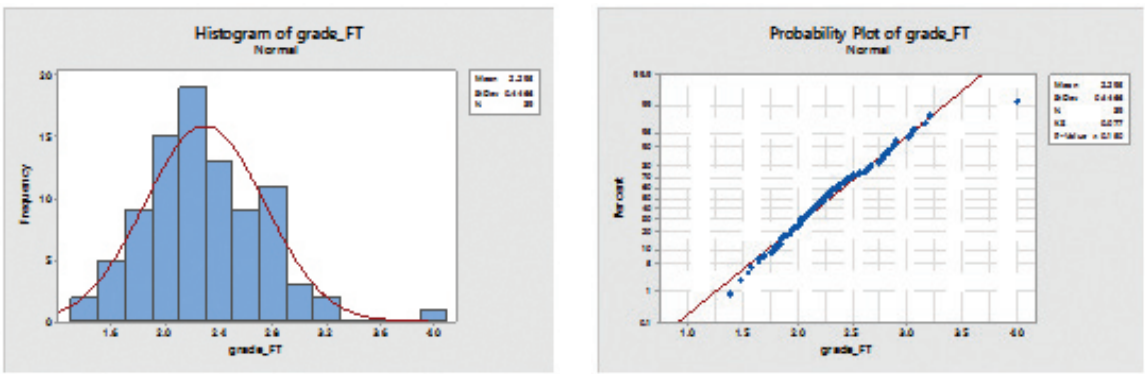
Figures 1 to 4 showed the normality test using Kolmogorov-Smirnov for both grades of full-time and part-time faculty members. The test revealed that grades for full-time (Fig. 2) does not follow a normal distribution with p-value of 0.024; while the grades of the part-time exhibits a normal distribution with p-value of greater than 0.150. With this result, a non-parametric test is employed since one variable of the study does not follow a normal distribution.

Table 1 presents the non-parametric test using Mann-Whitney to determine any significant difference of the medians of the two variables.

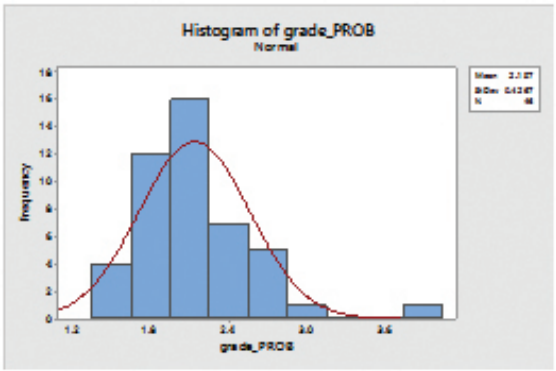
Table 1. *Mann-Whitney test for grades of full-time and part time faculty members*

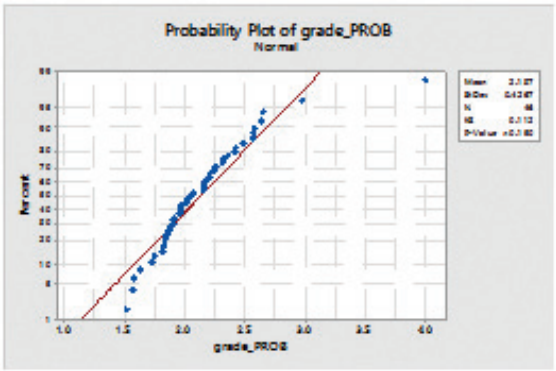
	N	Median
grades_FTandProbee	135	2.1827
grades_PTinandPTout	127	2.0560
Point estimate for η1 – η2 is 0.1593		
95.0 Percent CI for η1 – η2 is (0.0590,0.2601)		
W = 19655.5		
Test of η1 = η2 vs η1 ≠ η2 is significant at 0.0019		

A Mann-Whitney test for two independent samples revealed that median of the means of grades of full-time and part-time faculty members appears to be significantly different from each other with a test result of 0.0019 (Table 1). It further suggests that the same median of full-time and probationary faculty (2.1827) were significantly higher compared to the median of the part-time faculty member of 2.0560. The result appears to contradict the study of Bolge (1995), Clark (1990), Sonner (2000), Sonner and Sharland (1993) which they all found out student's grade is a function of an instructor's rank and that part-time instructors assign higher grades on average than full-time instructors. In this study, with the median of the means as the parameter being looked into, full-time revealed a higher result than its part-time counterpart. At the very least, the result neutralizes the claim of Landrum (1999) that higher grades given by part-time instructors gave the students unrealistic expectations as to the grades they should also receive, which, in turn, tended to put pressure on full-time instructors to follow. With this, students are expected to get fair grades whether they are with a full-time or part-time instructor.



Figures 5 &6. *Histogram and Normality plot of full-time tenured faculty midterm grades*





Figures 7 &8. Histogram and Normality plot of probationary faculty midterm grades

Separating the two subjects (probationary and tenured faculty) will give a glance on whether literatures are true that probationary faculty gives higher grade compared to tenured faculty because of the necessity of permanency requirement. But before any statistical analysis was conducted, normality tests were employed to determine the appropriate test to be done. Figures 5 to 8 showed the normality test for grades of both tenured and probationary faculty members. The test revealed that grades for both tenured (figure 6) and probationary faculty (figure 8) followed a normal distribution with p-values of greater than 0.150. With this result, a parametric test is deemed appropriate.

Table 2 presents the parametric test applied for grades of tenured and probationary faculty members. A T-Test for two independent samples was used for the two variables.

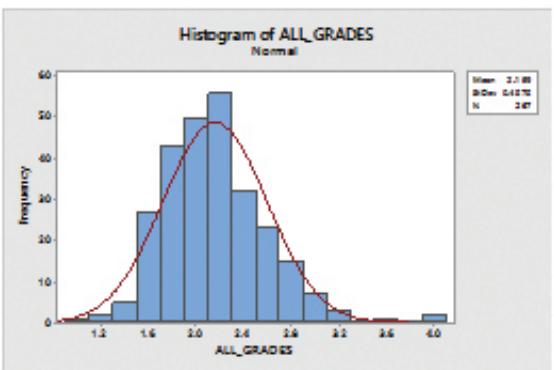
Table 2. *Two Sample T-Test for grades of tenured and probationary faculty members*

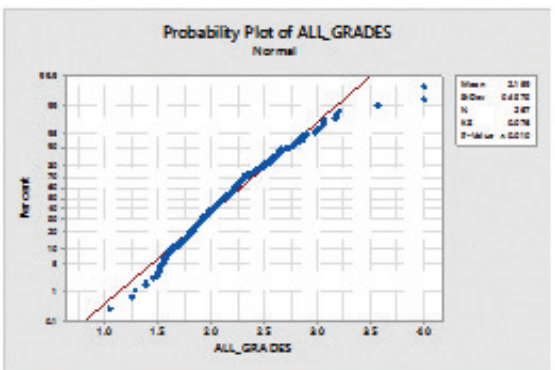
Two-sample T for grade_FT vs grade_PROB				
	N	Mean	StDev	SE Mean
grade_FT	89	2.295	0.447	0.047
grade_PROB	46	2.137	0.427	0.063
Difference = μ (grade_FT) − μ (grade_PROB)				
Estimate for difference: 0.1578				
95% CI for difference: (0.0014, 0.3141)				
T-Test of difference = 0 (vs ≠): T-Value = 2.00 **P-Value = 0.048** DF = 94				

A T-Test for two independent samples revealed that the means of midterm grades of tenured and probationary faculty are significantly different from each other with a p-value of 0.048. Looking at Table 2, tenured full-time faculty revealed a significantly higher mean compared to probationary teachers. These findings negate the results of researches that have consistently revealed that low-ranking instructors (in this case, probationary faculty) gave higher grades compared to their high ranking counterpart (Johnson, A. E., Pitts, S. T., & Kamery, R. H. (2006)) and that grading leniency is a factor

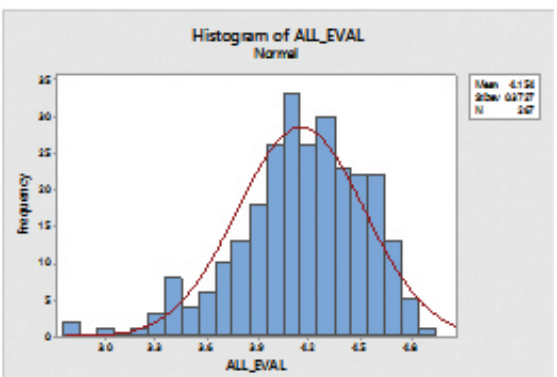
of getting a good evaluation result. This implies that grades may not necessarily be affected by the employment status of a faculty member.

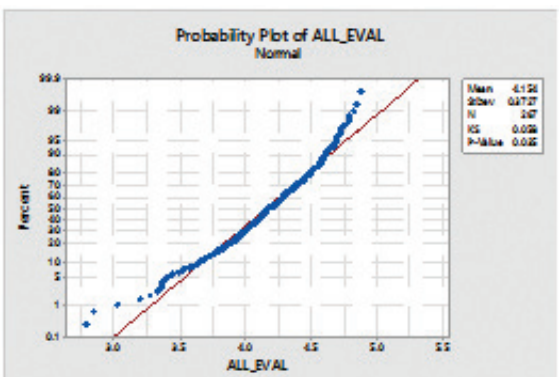
Several studies have also revealed that evaluation results are frequently used by individuals in making decisions for tenure and promotion (Ehie & Karathanos, 1994; Harrison, et al., 2004; Smith, 2004; Williams and Ceci, 1997). However, McKeachie (1997) posits that extra care should be given on how the administration utilized evaluation results for promotion and tenure. The necessity of probationary faculty to get a good evaluation result at the end of the semester would sometimes lead them to being lenient in giving grades to students, especially that midterm grades are given before students will answer the evaluation. But with the findings of this study, it seems that probationary faculty members do not care about the evaluation rating that they get from the students. The fact that probationary has lower mean grades compared to tenured faculty, it implies that they (probationary) are not afraid to get low evaluation from their students.





Figures 9 &10. Histogram and Normality plot of all midterm grades





Figures 11 &12. Histogram and Normality plot of all evaluation results

Figures 10 and 12 revealed that both grades and evaluation of all faculty members do not follow a normal distribution (<0.10 and 0.035 respectively) which warrants a non-parametric test to be used.

To determine relationships of grades and evaluation results of all faculty members, a Spearman Rho Correlation Test was employed in this study. Results are presented in Table 3.

Table 3. Spearman Rho correlation for Grades and Evaluation results

Spearman rho for ALL_GRADES and ALL_EVAL = **-0.163**
P-Value **= 0.008**

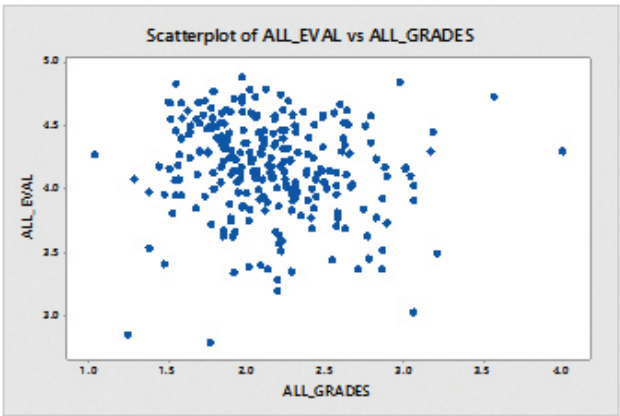


Figure 13. *Scatter plot of grades and evaluation*

A Spearman Rho Correlation Test for grades and evaluation of all teachers revealed that there is a weak negative correlation between the two variables as represented by a correlation coefficient of -0.163. However, the weak negative correlation is significant enough and does not happen by chance as represented by a p-value of 0.008. The result of the negative correlation is due to the fact that grades and evaluation ratings are inversely proportional with each other. Meaning, the highest possible grade that a student can get is 1 with a lowest of 5, while evaluation gave the highest of 5 and a lowest of 1. Nevertheless, the result revealed that indeed, few teachers who gave high grades to students would most likely get a high evaluation rating from them. Although, there is no trend or pattern as revealed in figure 13, the result of a weak correlation between the two implies that grades may have a little influence to the evaluation ratings. The results have supported the theory of Sander & Krautmann (1999) & Isely & Singh (2005) who claimed that instructors can "buy" better evaluation through a more lenient grading system and students who expects higher grade gives more favorable evaluation of their teacher. Some students may reward teachers' leniency while punishing rigorous instructors (Wright, R. E., & Palmer, J. C., 2006). This means that, indeed, there are few students who would reward their teachers with good evaluation because they (students) got a good grade. This is also where the subjectivity of the student comes in, because no matter what a teacher does inside the classroom, it will still boil down to what grade the student receives

from the teacher. However, other factors that lead to student's rating of teachers such as expected learning outcome and personal bias to the teacher may have contributed to the result to this relationship. But whether students' experiences were good or not, it will boil down to their personal feeling and opinion on each indicator once they answer the evaluation.

5.0 Conclusion

Based on the findings of the study, the principle on "Teachers can buy better evaluation results by giving high grades to students" is slightly manifested in the results. There is indeed a link between the two quality indicators of performance (grades and evaluation) when actual midterm grades are being looked into. However, the relationship between the two variables is more complex than simple. But whatever the outcome is, administrators in higher education institution, who use the results of student evaluation of their teachers as a key indicator of instructor's effectiveness, should be very careful as these results might not truly represent the performance of a teacher inside the classroom.

References

Barry, T. & Thomson, R. (1997). Some intriguing relationships in business teaching evaluations. Journal of Education for Business, 72(5), 303-311.

Bolge, R. D. (1995, February 28). Examination of student learning as a function of instructor status (Full-Time versus Part-Time) at Mercer County Community College. Retrieved from https://eric.ed.gov/?id=ED382241

Boretz, E. (2004). Grade inflation and the myth of student consumerism. College Teaching, 52(2), 42-47.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? Research in Higher Education,44(5), 495-518. doi:10.1023/a:1025492407752

Clark, B. A. (1990, March 31). Comparison of the achievement of students taught by full-time versus adjunct faculty in the chemistry of hazardous materials course: Governance and management. Retrieved from https://eric.ed.gov/?id=ED330261

D'Applonia, S. & Abrami, P. C. (1997, November). Navigating student's rating of instruction. American Psychologist, 52(11), 1198-1208.

Ehie, I. C., & Karathanos, D. (1994). Business faculty performance evaluation based on the new AACSB accreditation standards. Journal of Education for Business, 69(5), 257-262.

Filson, C. & Whittington, M. S. (2013, December). Engaging undergraduate students through academic advising. NACTA Journal, 57(4), 10-17.

Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. New York: Basic Books.

Harisson, P. D., Douglas, D. K., & Burdsal, C. A. (2004, May). The relative merits of different types of overall evaluations of teaching effectiveness. Research in Higher Education, 45(3), 311-323.

Isely, P. & Singh, H. (2005, February). Do higher grades lead to favorable student evaluations?. The Journal of Economic Education, 36(1), 29-42.

Johnson, A. E., Pitts, S. T., & Kamery, R. H. (2006, September). The effects of part-time instruction on final grades in the English composition course at a comprehensive Iia University. Academy of Educational Leadership Journal,10(3), 23-35.

Landrum, R. E. (1999, Winter). Student expectations of grade inflation. Journal of Research and Development in Education, 32(2), 124-128.

Marks, R. B. (2000). Determinants of students evaluations of global measures of instructor and course value. Journal of Marketing Education, 22(2), 108-119.

McSpirit, S., Chapman, A., Kopacz, P. & Jones, K. (2000, June). Faculty ironies on grade inflation. Journal of Instructional Psychology, 27(2), 104.

Pascale, P. J. (1979, Summer). Knowledge of final grade and effect on student evaluation of instruction. Educational Research Quarterly, 4(2), 52-57.

Krautmann, A. C. & Sander, W. (1999, February). Grades and student evaluations of teachers. Economics of Education Review, 18(1), 59-63.

Smith, G. S. (2004). Assessment strategies: What is being measured in student course evaluations?. Accounting Education, 13(1), 3-28.

Sonner, B. S. & Sharland, A. (1993). Grading differences between graduate teaching assistants and faculty members in the introductory marketing class. Journal of Marketing Education, 15(2), 44-49.

Sonner, B. S. (2000). A is for "Adjunct": Examining grade inflation in higher education. Journal of Education for Business, 76(1), 5-10.

Williams, W. M., & Ceci, S. J. (1997, September/October). "How' M I doing?". Change, 29(5), 12-23.

Wright, R. E., & Palmer, J. C. (2006). A comparative analysis of different models explaining the relationship between instructor ratings and expected student grades. Educational Research Quarterly, 30(2), 3-18.