

Non-modifiable Factors and Student Evaluation of Faculty Teaching Quality: An Examination of the Correlations

Neal Luo*, Daniel Stiffler

Department of School Leadership/ Middle & Secondary Teacher Education, Emporia State University, USA

Received September 4, 2019; Revised October 28, 2019; Accepted November 5, 2019

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Course evaluation by students has been widely used to produce essential information for university administrators and faculty in assessing instruction quality. This study focused on examining the non-modifiable factors possibly relating to student evaluation outcomes by analyzing the quantitative data of 259-course evaluations in a teachers college at a Midwest state university. Findings of multiple regression and univariate statistical analyses suggested that the mean score of students' ratings in a course was associated with the ranks of the faculty who taught the course and students' response rate of the course survey. Courses taught by higher-ranking faculty and rated by a lower percentage of students tend to have lower mean scores of course evaluation. The mean score of students' evaluation in a course was not correlated with the other variables of the gender of faculty, course level, course delivery method, and class size. Findings of this study lead to a broader impact on faculty teaching evaluation policy and implications for course evaluation practices.

Keywords Faculty teaching Assessment, Student Evaluation, Course-based Factors, Faculty Rank, Correlation Analysis

1. Introduction

Student evaluation remains an important and widely utilized tool providing essential information on the assessment of instructor teaching in spite of the controversy about its validity and reliability (Boring, Ottoboni, & Stark, 2016). In general, state legislators, higher education boards and administrators are very enthusiastic about assessment for accountability, quality control, and policy purposes. University administrators including provosts, deans, and department chairs need the

course evaluation information to make decisions and create policies regarding teaching effectiveness. They frequently use student evaluations as essential evidence for policy and decision making on faculty tenure promotion, reappointment, and pay increases. Student evaluations have become the primary, even dominant summative indicators on the overall performance on teaching competence (Galbraith, Merrill, & Kline, 2012). From the instructors' perspective, course evaluation by students is more effectively utilized for formative purposes to provide instructors with field-based information about student learning, allowing them to shape and improve their teaching in a practical manner (Hornstein & Law, 2017).

Researchers have largely focused on identifying the modifiable characteristics or factors impacting the interactions between effective instructors and learners. (e.g., Hornstein, & Law, 2017; Boring et al., 2016). These studies reveal that program and course innovativeness and appropriateness, teaching performance, and career-related issues are the essential sources of student satisfaction. In addition, faculty competencies, communication skills, attitudes, likability and appropriate use of humor are found to be positively correlated with student ratings (Duque, 2013; Hornstein, & Law). Student course grades, student characteristics and reasons for enrolling in a course are also predictors of student evaluation (Denson, Loveday, & Dalton, 2010; Hoefler, Yurkiewicz, & Byrne, 2012). However, whether student evaluation is related to the non-modifiable course features and instructor characteristics, has not been comprehensively evaluated (Morgan, et al., 2016). As the non-modifiable course-based and faculty factors constitute the important components used by university administrators and faculty for policy and practical decision making, it is essential to better understand what course-based and faculty factors may relate to student course evaluation outcomes. Further, evidence or research-based understanding and decision

making can reduce the possibility of making unfair or inequitable policies for evaluating faculty teaching that affects faculty motivation and even moralization.

The theoretical framework of factors regarding teaching effectiveness measured by student evaluation outcomes is complex. It not only involves modifiable modes of instructors' teaching and students' learning (Brown, 2004; Denson et al., 2010; Hoefler, Yurkiewicz, & Byrne, 2012), but also includes the various non-modifiable factors for instance, demographic and background characteristics of the instructor, and teaching context (Young & Shaw, 1999). This study intends to address a research question that is of significant interest to many higher education institutions by examining the non-modifiable faculty and course-based factors that possibly correlate to student course evaluation outcomes.

2. Literature Review

The literature examining the correlations between non-modifiable factors and student evaluation has primarily focused on the impact of demographics. Studies related to the effect of gender on student evaluation do not reach a final verdict and seem to be situational. The gender bias appears to affect how students rate the objective aspects of teaching effectiveness and can be significant enough to cause more effective instructors to get lower student evaluation ratings than less effective instructors (as cited in Hornstein & Law, 2017). In undergraduate classes, male instructors got better evaluation scores than females (McPherson, Jewell, & Kim, 2009). Less effective male instructors received higher student evaluation ratings than more effective female instructors (Boring et al., 2016; Kogan, Schoenfeld-Tacher, & Hellyer, 2010). Students were more likely to give higher ratings to faculties of the same gender as students (Grimes, Millea, & Woodruff, 2004; MacNell, Driscoll, & Hunt 2015). Female students rated female faculty higher than male faculty while male students' ratings on both male and female professors did not make a significant difference (Bachen, McLoughlin, & Garcia, 1999). Other studies have failed to find this type of evidence. Same-gender preferences have a minor effect on students' ratings and teaching style rather than gender may well explain these preferences (Centra & Gaubatz, 2000). Instructor gender and student evaluations are not significantly associated without controlling for teaching effectiveness, effort, or other variables (Bennett, 1982; Cole, Shelley, & Swartz, 2014).

Studies correlating student evaluations of teaching to faculty rank, course level, and class size were mostly found in the previous century. Highly significant differences were found in ratings assigned by students in freshman, sophomore, junior, senior, and graduate-level courses. In addition, there were also significant interaction effects of class size, course level, and faculty rank interaction effects

on student evaluations of teaching (Aleamoni & Graham, 1974; Feldman, 1978). Since student ratings are multidimensional, faculty rank and class size may bias student rating data by causing differences (Cashin, 1988). Non-tenure-track instructors receive better

SET scores (Hamermesh & Parker, 2005). Other studies reported that student evaluations of teaching were not significantly related to faculty rank (Hayes, 1971; Kogan, Schoenfeld-Tacher, & Hellyer, 2010), class size (Marsh & Roche, 1997) and students' choice of course section (Wilson, 1998).

Age is a factor related to the course level of undergraduate and graduate courses, and also possibly faculty rank. It is primarily included in studies of student evaluation of faculty teaching. The course evaluation rating seems to be also related to the age of the student. Older students tend to rate teacher evaluations higher (Worthington, 2002). Younger instructors are more popular than older ones in teaching undergraduate classes (McPherson et al., 2009). Older teachers tended to receive lower ratings than younger teachers. Unattractive middle-aged female teachers and unattractive old male teachers frequently received lower ratings (Goebel & Cashen, 1979).

With the increasing trend of colleges offering online programs, the course delivery method has become another focused non-modifiable factor that attracts researchers' attention to examine its impact on student evaluations of teaching. Univariate analysis revealed that students enrolled in the online course were significantly less satisfied with the course than the traditional classroom students (Summers, Waigandt, & Whittaker, 2005). Content analysis of anonymous student responses to open-ended questions showed no significant difference in the proportion of appraisal text segments by delivery method, but the existence of significant differences in the text segments for topical themes and topical categories by delivery method (Kelly, Ponton, & Rovai, 2007). Cole et al. (2014) found partially online courses were rated as somewhat more satisfactory than fully online courses. The reason for satisfaction with fully online courses is convenience while lack of interaction contributes to the most cited reason for dissatisfaction. Hybrid courses were preferred for meeting students' needs.

Within the existing research studies on investigating the relationship between non-modifiable factors such as gender, age, and course delivery, and SET, inconsistent findings and outdated literature exist. Since student satisfaction is a complex and multidimensional phenomenon influenced by various variables, we can assume that the potential difference in student evaluations of teaching due to these factors is also contextually or situationally bounded, which can be inconclusive. Another important issue can be the research methodology using small sample and univariate analysis without the capacity to control the extraneous variables. In addition, confusion

and discontent were also evident in ratings-based evaluations. However, student evaluations of teaching will continue to be used for personnel decisions. More comprehensively, methodology-sound, and context-based research in improving the process and addressing areas of concern are still needed (Algozzine et al., 2004).

The purpose of this study was to determine whether or not the assessment outcome of faculty instruction through student evaluations are correlated to factors of course features and instructor characteristics that university administrators should be considering when creating policy and making decisions on instructor evaluation. Factors of faculty gender, faculty rank, semester, course level, class size, course delivery method, and course survey response rate were included as independent variables in this study by analyzing the quantitative data of 259 course evaluations collected by the IDEA (Individual Development and Educational Assessment, 2016) survey in a teachers college at a teaching orientated Midwest state university.

3. Methods

3.1. Design, Variables, and Instrument

This study utilized quantitative correlational design to determine whether or not the assessment outcome of faculty instruction by student evaluation was related to factors of instructor characteristics and course features. Factors of faculty gender, faculty rank (assistant, associate and full professor), semester (spring vs fall) when courses were offered, course level (undergraduate vs graduate), class size, course delivery method (online vs face-to-face), and course survey response rate were included as independent variables in the analysis of this study (also see Table 1 and Table 2).

The assessment outcome of faculty instruction measured by the course-based mean score of student evaluation was the dependent variable in this study. Faculty teaching quality assessed was conceptualized by the course teaching objectives selected by the instructors. It was operationalized with the assessment outcome of instruction by using the course-based IDEA (Individual Development and Educational Assessment, 2016) cross-sectional student ratings of instruction. IDEA (2016) is a nonprofit organization dedicated to improving student learning in higher education through analytics, resources, and advice. The IDEA survey is a tool that allows students to rate their learning experience in online, hybrid, or face-to-face classrooms. It invites students to respond to items that ask them about the learning methods they experienced, how much they believe they've learned, how hard they worked and how much they wanted to take the class (IDEA, 2016). Specifically, the IDEA Survey measures 1) teaching effectiveness by asking the students the survey items in addressing the question of "Did you design or plan for this

coursework?" 2) student-reported progress on instructor-chosen course objectives on the question of "Did your students think they were learning what you intended to teach?" and 3) teaching methods on the question of "Did you use methods that supported your goals?"

This study utilized the IDEA (2016) data collected on the student-reported progress on instructor-chosen course objectives. There are a total of 13 learning objectives developed in the IDEA survey: 1) Gaining a basic understanding of the subject (e.g., factual knowledge, methods, principles, generalizations, theories), 2) Developing knowledge and understanding of diverse perspectives, global awareness, or other cultures, 3) Learning to apply course material (to improve thinking, problem-solving, and decisions), 4) Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course, 5) Acquiring skills in working with others as a member of a team, 6) Developing creative capacities (inventing; designing; writing; performing in art, music, drama, etc.), 7) Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.), 8) Developing skill in expressing oneself orally or in writing, 9) Learning how to find, evaluate, and use resources to explore a topic in-depth, 10) Developing ethical reasoning and/or ethical decision making, 11) Learning to analyze and critically evaluate ideas, arguments, and points of view, 12) Learning to apply knowledge and skills to benefit others or serve the public good, 13) Learning appropriate methods for collecting, analyzing, and interpreting numerical information.

Each course instructor is required to select at least three of the above 13 objectives that are deemed to be important or essential for students' learning during the course before students complete the survey. The IDEA response choice use Likert-scales ranging from 1 representing "no apparent progress", through 3 representing "moderate progress" to 5 representing "exceptional progress".

The IDEA Survey has been tested and provides some evidence on its reliability and validity. Average reliability coefficients for individual items ranged from .78 for class sizes of 10–14 students to .94 for enrollments of 50 or more, which provide evidence to support the degree of reliability in individual items. There are correlations between students' self-ratings of progress on the learning objectives and their ratings of how frequently the instructor emphasized 20 specific teaching methods (Hoyt & Lee, 2002). Students' ratings of their own learning correlate positively with the instructor's measure of how much they have actually learned (Benton, Duchon, & Pallett, 2013).

3.2. Data and Statistical Analysis

Data compiled for analysis of this study came from the 259 course evaluations collected in the spring and fall semester of 2016 using the IDEA survey in a teachers

college at a Midwest state university. All students who took the courses were notified to complete the online IDEA survey during the last week of the classes. The survey data used for this study represent all the student evaluations of the courses taught by tenured and tenure-track faculty in the spring and fall semester in the whole college. The course evaluation data also exclude courses with a student enrollment of less than five from the analysis. The objective summary mean score of the student-reported progress on all the instructor-chosen course objectives for each course was used for the unit of analysis (rather than individual student's ratings) measuring the dependent variable of faculty teaching quality in this study. The data of independent variables for analysis on the factors of faculty gender, faculty rank, semester, course level, class size, course delivery method, and course survey response rate were also collected by the IDEA survey.

The course-based data were analyzed using descriptive statistics, univariate tests, and a multiple regression to seek the results. Mean scores and standard deviations for the IDEA (2016) objective summary scores were calculated to present the levels of student ratings by the variables of course features and instructor characteristics. Independent t-tests were used to determine whether there were significant differences in student ratings of course

evaluations by faculty gender, course level, and course delivery method. A one-way ANOVA was applied to determine whether there were significant differences in student ratings by faculty rank. A multiple regression was finally used to examine what factors significantly relate to the course-based evaluation of student ratings.

4. Results

The overall mean of students' ratings of the 259 courses on their learning progress objectives in the calendar year is 4.39 ($SD=.58$). Table 1 lists the means and standard deviations for each of the levels by the factors of course features and instructor characteristics. Results of independent t-tests and ANOVA are also presented in the table. No significant differences in students' ratings were found between the groups by course level, course delivery method, and faculty gender. The ANOVA test revealed significant differences in students' ratings by faculty rank ($F(2, 256)=3.893, p=.022$). Pairwise comparison follow-up tests indicated that the mean score of student ratings of courses taught by full professors was significantly lower than those taught by assistant professors ($p=.017$).

Table 1. Descriptive and Univariate Test Statistics of the Student Ratings by Instructor Characteristics and Course Features

Factors	Levels	n	Mean	SD	Test Statistics	p
All Courses		259	4.39	0.58		
Faculty Gender	Female	136	4.39	0.63	t=0.228	0.82
	Male	122	4.4	0.52		
Faculty Rank	Assistant Professor	86	4.49	0.32	F=3.893	.022*
	Associate Professor	120	4.41	0.59		
	Full Professor	53	4.21	0.81		
Course Level	Undergraduate	107	4.31	0.7	t=1.806	0.072
	Graduate	152	4.45	0.47		
Course Delivery Method	Online	147	4.35	0.64	t=1.328	0.185
	Face-to-face	112	4.45	0.49		

Note. The IDEA response choice scales were 1=No apparent progress, 2=Slight progress, 3=Moderate progress, 4=Substantial progress, 5=Exceptional progress. * p is significant at .05.

Table 2. Multiple Regression Coefficients on the Correlations between Non-Modifiable Factors and Student Ratings

Variables	Beta (unstd.)	Std. Error	Beta (std.)	t	p	VIF
(Constant)	4.218	0.208		20.325	.000	
Faculty Gender	0.068	0.078	0.059	0.877	0.381	1.182
Faculty Rank Dummy 1(Full Prof. vs Associate Prof.)	0.076	0.082	0.062	0.928	0.354	1.118
Faculty Rank Dummy 2(Full Prof. vs Assistant Prof.)	-0.252	0.103	-0.176	-2.439	.015*	1.373
Semester (Fall vs Spring)	-0.075	0.072	-0.065	-1.049	0.295	1.01
Course Level (UG vs GR)	0.008	0.024	0.033	0.353	0.725	2.286
Course Delivery Method(Online vs Face-to Face)	0.046	0.101	0.039	0.456	0.649	1.989
Class size	-0.004	0.005	-0.052	-0.755	0.451	1.259
Response Rate of a Course Survey	0.003	0.002	0.116	1.819	.050*	1.07

Note. The IDEA response choice scales were 1=No apparent progress, 2=Slight progress, 3=Moderate progress, 4=Substantial progress, 5=Exceptional progress. * p is significant at .05.

Multiple regression indicates that the mean score of students' ratings in a course was associated with the ranks of the faculty who taught the course and students' response rate of the course survey. Courses taught by higher-ranking faculty and rated by a lower percentage of students tend to have lower mean scores of course evaluation. The mean score of students' evaluation in a course was not correlated with the other variables of faculty gender, semester, course level, class size, and course delivery method. The model significantly predicts the mean scores of a course rated by students ($F=2.001$, $p=.046$), explaining 10.23% of the variances of the mean scores. The VIF (variance inflation factor) values are all well below 10 indicating that the data analyzed in this study meet the assumption of no collinearity among the predictor variables in the regression. Table 2 presents the coefficients and other statistics of the regression model.

5. Discussion and Conclusions

This multivariate analysis of 259 courses including seven variables as factors in a teachers college at a teaching orientated Midwest state university adds to the literature body of student evaluation of faculty teaching. We found that the non-modifiable variables including faculty gender, course level, course delivery method, and class size did not make significant differences in students' ratings in their course evaluation. These findings differ from the previous studies of student evaluations of faculty, which demonstrated significant differences based on course delivery method (Summers et al., 2005; Cole et al., 2014), class size (McPherson et al., 2009), and faculty gender (Boring et al., 2016; Hornstein, & Law, 2017). The inherent differences in the research context and time between our study and the previous ones may have contributed to the dissimilar findings, which further support our assumption that the differences in student evaluation rating due to these factors are bound by the situation, context, and even time. Context-based and timely analysis of studies seems to be important and needed in future areas of inquiry. From a practical perspective, the previous findings of our study provide evidence to guide college administrators' decision and policy makings in evaluating faculty teaching. Specifically, in using student evaluations of teaching for setting criteria for faculty tenure promotion, reappointment, and pay increases, administrators seemingly do not need to use multiple standards due to the differences in faculty rank, course level, course delivery method, and class size.

Our study found that faculty with higher rank tended to have lower course evaluation scores, which is not consistent with the previous research finding of no differences based on rank (Kogan et al., 2010), but seems to support the results of some other studies focusing on the factor of age that older instructors tended to receive lower ratings than younger ones (McPherson et al., 2009). One

contextual explanation for this finding can reasonably be drawn in utilizing student evaluation ratings as the only criterion measuring teaching success for tenured promotion while such criterion seems to have little impact on tenured faculty. It is unclear whether the finding of our study is due to the above notion or to the notion that the lower the evaluations, the better that student performance tends to be because good instructors require their students to exert effort that students dislike. Student evaluations sometimes just reflect the utility they enjoy from the course (Braga et al., 2014). Therefore, it is worthwhile to conduct further or expanded research with larger samples and in-depth qualitative explorations on this assumption.

On the other hand, a positive correlation between non-tenure-track faculty status and student evaluations of teaching can be valid as these instructors inspire higher future student achievement and provide higher quality instruction than their peers (Figlio, Schapiro, & Soter, 2015). With similar results confirmed by further research, conversations and discussions leading to additional tenured faculty development opportunities and even policy development regarding maintaining and enhancing tenured or senior faculty instruction quality may be needed.

This study has several limitations. It examined students' quantitative self-reported perceptions of their course evaluation rather than analyzing and using the students' qualitative feedback or peer-review data in observing the classes. This possibly can impact the objectivity and accuracy of the measurements. It is also important to pay careful attention to student comments, understand their scope and limitations, spend more time observing the teaching and looking at teaching materials to improve teaching and evaluate teaching fairly and honestly (Stark & Freishtat, 2014). The course evaluation of this study was delimited on student ratings on courses of the disciplines of education and psychology in a teachers college at a Midwestern teaching-oriented state university. This context can limit the generalizability of the research results to the course evaluations of other disciplines or other institutional settings. Lastly, since course evaluation results are affected by various factors including, for instance, course content and student characteristics, this study focusing on analyzing the non-modifiable course-based and faculty characteristics, does not have the capacity to examine the integrative features of the comprehensive factors on students' course evaluation. Future studies of similar topics should be further examined by including more comprehensive factors, combining modifiable and non-modifiable factors.

REFERENCES

- [1] Aleamoni, L. M., & Graham, M. H. (1974). The relationship between CEQ ratings and instructor's rank, class size and

- course level. *Journal of Educational Measurement*, 11(3), 89-202. <https://doi.org/10.1111/j.1745-3984.1974.tb00990.x>
- [2] Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F., Mohanty, G., & Bray M. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching*, 52(4), 134-141. <https://doi.org/10.3200/CTCH.52.4.134-141>
- [3] Bachen, C. M., McLoughlin, M. M., & Garcia, S. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education*, 48(3), 193-210.
- [4] Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, 74, 170-179. <http://doi.org/10.1037/0022-0663.74.2.170>
- [5] Benton, S. L., Duchon, D., & Pallett, W. H. (2013). Validity of self-reported student ratings of instruction. *Assessment & Evaluation in Higher Education*, 38(4), 377-389.
- [6] Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71-88. <http://doi.org/10.1016/j.econedurev.2014.04.002>
- [7] Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. Retrieved from Science Open Research. <http://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZ.C.v1>
- [8] Cashin, W. E. (1988). Student ratings of teaching: A summary of the research. IDEA Paper No. 20. (ERIC, Document Reproduction Service No. 302567) Retrieved from <https://eric.ed.gov/?id=ED302567>
- [9] Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 71, 17-33.
- [10] Cole, M. T., Shelley, D. J., & Swartz, L. B. (2014). Online instruction, e-learning, and student satisfaction: A three year study. *The International Review of Research in Open and Distributed Learning*, 15(6).
- [11] Denson, N., Loveday, T., & Dalton, H. (2010). Student evaluation of courses: What predicts satisfaction? *Higher Education Research & Development*, 29(4), 339-356. <http://doi.org/10.1080/07294360903394466>
- [12] Duque, L. C. (2013). A framework for analyzing higher education performance: Students' satisfaction, perceived learning outcomes, and dropout intention. *Total Quality Management and Business Excellence*, 25(1-2), 1-21.
- [13] Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, 9(3), 199-242.
- [14] Figlio, D. N., Schapiro, M. O., & Soter, K. B. (2015). Are tenure track professors' better teachers? *Review of Economics and Statistics*, 97(4), 715-724.
- [15] Galbraith, C., Merrill, G., & Kline, D. (2012). Are student evaluations of teaching effectiveness valid for measuring student outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education*, 53, 353-374.
- [16] Grimes, P., Millea, M., & Woodruff, T. (2004). Grades: Who's to blame? Student evaluation of teaching and locus of control. *The Journal of Economic Education*, 35(2), 129-147. Retrieved from <http://www.jstor.org/stable/30042585>
- [17] Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24(4), 369-376.
- [18] Hayes, J. R. (1971). Research, Teaching, and Faculty Fate. *Science*, 172(3980), 227-230. <https://doi.org/10.1126/science.172.3980.227>
- [19] Hoefler, P., Yurkiewicz, J., & Byrne, J. C. (2012). The association between students' evaluation of teaching and grades. *Decision Sciences Journal of Innovative Education*, 10, 447-459. <http://doi.org/10.1111/j.1540-4609.2012.00345.x>
- [20] Hornstein, H. A., & Law, H. F. E. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1). Retrieved from <https://www.tandfonline.com/doi/full/10.1080/2331186X.2017.1304016> <http://doi.org/10.1080/2331186X.2017.1304016>
- [21] Hoyt, D. P., & Lee, E. (2002). Technical Report No. 12: Basic data for the revised IDEA system. Manhattan, KS: The IDEA Center.
- [22] Kelly, H. F., Ponton, M. K., & Rovai, A. P. (2007). A comparison of student evaluations of teaching between online and face-to-face courses. *The Internet and Higher Education*, 10(2), 89-101. <https://doi.org/10.1016/j.iheduc.2007.02.001>
- [23] Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P. W. (2010). Student evaluations of teaching: Perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, 15(6), 623-636. <https://doi.org/10.1080/13562517.2010.491911>
- [24] MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291-303.
- [25] Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187.
- [26] McPherson, M., Jewell, R., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economics Journal*, 35(1), 37-51. <https://doi.org/10.1057/palgrave.eej.9050042>
- [27] Morgan, H. K., Purkiss, J. A., Porter, A.C., Lypson, M. L., Santen S. A., Christner J. G., Grum C. M., & Hammoud, M. M. (2016). Student evaluation of faculty physicians: Gender differences in teaching evaluations. *Journal of Women's Health*, 25(5), 453-456. <http://doi.org/10.1089/jwh.2015.5475>
- [28] Nowell, C., Gale, L. R., & Handley, B. (2010). Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment & Evaluation in*

Higher Education, 35(4), 463-475. <http://doi.org/10.1080/02602930902862875>

- [29] Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. <http://doi.org/10.14293/S2199-1006.1.SQR-EDU.AOFRQA.v1Stark>. Retrieved from Science Open: <https://www.scienceopen.com/document/vid/42e6aae5-246b-4900-8015-dc99b467b6e4?0>
- [30] Wilson, R. (1998, January 16). New research casts doubt on value of student evaluations of professors. *Chronicle of Higher Education*, 44, 12–14.
- [31] Worthington, A. C. (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. <https://doi.org/10.1080/02602930120105054>