Post-Secondary Student Evaluations of Teachers: The Debate of Usefulness Continues

Heather Dana[1], Scott Morrissette[1] & Sheree Nelson[1]

[1] School of Business & IT, Purdue University Global, West Lafayette, Indiana, USA

Correspondence: Heather Dana, School of Business & IT, Purdue University Global, West Lafayette, Indiana, USA.

Abstract

Student evaluations of teachers (SETs) are collected by colleges and universities across the country. Having only been introduced in the early part of the twentieth century, these evaluations are a fairly new tool for higher education administrators to receive feedback and assess the effectiveness of curriculum and instructors. Although implemented as a tool to provide students a medium to share their perspectives, with the goal to improve academic processes, there are concerns regarding their effectiveness, reliability, purpose, and necessity. Further, the literature reflects that students are not well versed by college administrations or faculty members regarding the desired impacts and purpose of SETs, so they are often not completed in a manner that includes cognitive engagement, accurate recall, or the genuine desire to provide constructive feedback and assessment. Even with these limitations, college and university administrators have grown to rely upon SETs to provide constructive insights for instructors to help them improve their teaching effectiveness and summative feedback for committees to use when making promotion, tenure, and compensation decisions. The disconnect between SET objectives and the actual outcomes, however, is problematic. Students often don't view SETs as impactful, so their level of cognitive engagement is lacking, which can result in skewed, or even false assessments. In fact, since most SETs are completed with the promise of anonymity, they have been used as a weapon by disgruntled students against instructors, regardless of whether the negative feedback is deserved. Finally, SETs have been directly correlated to grade inflation, which has numerous negative implications. The following literature review illustrates the myriad shortcomings of SETs, with the hope that further research will help to discover how they can be re-structured to foster academic excellence in a productive and reliable manner.

Keywords: student evaluations, evaluation of teachers, evaluations, education level, geographic area

1. Introduction

Online reviews have become extremely popular. In fact, many people rely upon them to make everyday decisions. The online customer satisfaction website Yelp has become, for many consumers around the world, a relied upon source to determine where they will shop, eat, and seek entertainment. World-wide internet access via smartphones has also enabled consumers to locate customer satisfaction reviews in a matter of minutes, resulting in a highly interconnected and informed public. In turn, consumers have become accustomed to business owners being responsive to online reviews, as well as fearful of them, reinforcing the common saying that "The customer is always right." In fact, McKenna (2018) noted that, due to the fear of bad Yelp reviews by patients, prescriptions for antibiotics may be written for demanding patients who may not actually need them. This level of empowerment, according to McKenna (2018), has enabled patients to take an unqualified and potentially dangerous role in their own healthcare. Luca (2016) identified the impact that a single increase in a 1-to-5-star rating scale on Yelp had on restaurants, which was correlated with up to a 5.4% increase in revenue. This tangible measurement demonstrates the relevance of Yelp reviews and consumer empowerment.

Current literature, which will be addressed in this manuscript, shows that similar consumer expectations have become evident in the college setting, both in person and online. The line between customer and student has become blurred, with students viewing college professors as people from whom they are buying a service. Coupled with the increased online platform, which removes the face-to-face element between students and professors, the dynamics of the student/professor relationship have changed. In fact, the literature supports the perspective that students have developed an expectation of more accountability from professors, while professors

have developed a decreased expectation of accountability from students. Further, the literature suggests that colleges and universities have created an environment that supports this shift in student/professor dynamics by placing particular emphasis on the importance of end-of-term student evaluations that are used to assess faculty performance, determine faculty course assignments, and contribute to decisions regarding career advancements. In extreme cases, end-of-term student evaluations can even compromise job security and foster an environment of fear and resentment.

1.1 Problem with Student Evaluations of Teachers and its Importance

The literature, reviewed in this manuscript, reflects that concerns about the effectiveness and reliability of SETs have been consistently expressed since their inception. The literature also demonstrates that college and university administrations continue to rely upon SETs to assist with their decision making, despite the numerous studies that have revealed that they do not provide reliable feedback. Although there is a consensus that student feedback is necessary and having a reliable way to assess instructor effectiveness is important, the current methods and instruments used to meet these objectives are inefficient and, in extreme cases, potentially harmful. To that end, research has also shown that the reliance upon SETs for the decision-making regarding faculty promotions, tenure, pay increases, and course selections has led to a shifting in the dynamics in the college classroom whereby students have become more like customers whose negative reviews could adversely affect the careers of faculty members. Further, there is substantial evidence that the fear of bad evaluations from students has fostered a steady increase in grade inflation. In turn, grade inflation has contributed to decreased student accountability, a reduction in student critical thinking requirements, frustrated higher performing students, and less prepared college graduates.

The questions the authors hoped to answer with this literature review include:

1) Has the increased importance of end-of-term student evaluations given students too much authority over faculty success, and is there evidence that more lenient professors are being promoted more often than stricter professors?

2) Has the increased importance of end-of-term student evaluations contributed to grade inflation and/or resulted in quality professors being penalized for holding students accountable for higher quality performance?

3) Has the increased access to the internet and online education platforms resulted in a less personal environment whereby students are more likely to provide harsh feedback than in traditional face-to-face higher education settings?

The following literature review addresses these three questions. Further, gaps in the literature will be identified and suggestions for further research will be offered.

2. Review of Relevant Literature

2.1 History of Student Evaluations

Prior to the 1920s, the teaching profession did not have a formal evaluation process. The growth of student organizations across campuses and the increased discontent with college instruction in general resulted in a call to action for professor accountability and more student-driven course design and implementation (Gelber, 2020). At the same time, college campuses around the country were experiencing increased population and diversity among students, which perpetuated the changing needs and demands. The Free Speech Movement on the University of California campus, as well as other campuses, exposed the desire for an overhaul of the faculty evaluation process which still relied heavily on informal conversations, rumors, and personal relationships (Gelber, 2020).

Student evaluations were created by Herman H. Remmers of Purdue University and Edwin Guthrie of the University of Washington in the 1920s (Stroebe, 2020). Not all faculty members welcomed being formally evaluated by students, though, as many people considered professorship as a field whose members had "earned" the privilege of being taken at face value, respected, and mostly above reproach (Gelber, 2020). Regardless, by 1950, one-third of colleges and universities in the United States had implemented some form of structured evaluation process (Gelber, 2020). Further, the Civil Rights Movement and anti-war protests contributed to the questioning of authority and the status quo, and college campuses were no exception. In 1961, the American Council on Education reported that only 12 percent of colleges and universities employed a systematic process for collecting student feedback, and only 14 percent conducted observations of faculty for performance evaluation purposes (Gelber, 2020). According to Stroebe (2020), as of 1973, 29 percent of colleges conducted some form of student surveys. This increased to 68 percent by 1983, 8 percent by 1993, and incrementally increasing to 94 percent by 2010 (Stroebe, 2020). Further, a study conducted in 2010 that surveyed college deans

about the importance of student evaluations resulted in nearly all deans acknowledging that end of course surveys are the main source of information to measure the quality of classroom teaching (Stroebe, 2020).

In addition to the students' demands for accountability, personal factors were called into question when evaluating how faculty members were hired, assessed, and promoted. In 1972, Title VII of the Civil Rights Act of 1964 was extended to prevent the discrimination of faculty on the basis of race, religion, sex, or national origin (Gelber, 2020). This was no easy task, however, as judges and scholars were hesitant to place constraints upon academic decision makers. Pacholski (1992) noted that The Supreme Court has addressed this concern by clarifying that the extension of Title VII of the Civil Rights Act of 1964 only applied to decisions that could be substantiated as being discriminatory, rather than on legitimate academic criteria. This called for faculty evaluations to include empirical evidence, which led to many colleges implementing and relying upon end-user assessment. This end user refers to the ultimate customer for evaluation, which in education, naturally, falls upon students (Gelber, 2020).

A result of providing a means of measurable assessment to help academic decision makers comply with anti-discrimination laws, anonymous student evaluations, whether collected in person at the end of a term or online, provided a means for academic institutions to evaluate their strengths and weaknesses and help to guide the colleges toward improved higher education goals (Gelber, 2020). Further, Uijtdehaage and O'Neal (2015) stated that student evaluations of teachers are widely used for feedback that with the support of department chairpersons and curriculum coaches can lead to course improvements and better teaching practices, as well as a collection of insights by students that can be used by tenure and promotion committees.

2.2 Benefits and Shortcomings of End of Term Student Evaluations

2.2.1 Questions of Validity

Concerns about the validity of students' evaluations of instructors is not new. Reynolds (1977) found that students rated a film on sexuality and communication higher than a lecture on the history of psychology, both of which the students had neither seen nor heard, indicating the lack of reliability of student feedback. These findings are similar to what Uijtdehaage and O'Neal (2015) revealed with their research that suggested that student's evaluations are not reliable because they are often completed mindlessly, without careful consideration of the faculty members' performance. In fact, their study revealed that two thirds of the students had submitted evaluations for professors whom they had never had as an instructor. In response to Reynolds' 1977 article, Gibson (2015) noted that grade inflation is still a current problem in colleges and universities. Echoing Reynolds' concerns, Gibson reported that grade inflation continues to cause instructors to be overly cautious, which results in lectures and conversations that are lacking substance and higher order analysis. The result, Gibson noted, is a conflict between students' short-term desires for easier material and higher grades and academic institutions' responsibility to maintain academic integrity, promote higher order thinking, and award diplomas to students who are prepared to successfully enter their respective fields of study.

Contributing to the concern about quality assurance and confidence in student rating, Utll et al. (2017) concluded that student learning is not related to student ratings of faculty owing to numerous biases including faculty gender, ethnicity, attractiveness, body language, and level of leniency. Additionally, student fatigue, type of course, timing, weather, difference in academic disciplines, and whether the students view evaluations as being useful also influence students' ratings (Utll et al., 2017). Research has also shown that many students either do not complete SETs in a mindful manner or they put forth sufficient effort only when they are extremely unsatisfied with an instructor or their grades, or both (Dunegan & Hrivnak, 2003). Dunegan and Hrivnak (2003) noted similar concerns when they identified three risk factors that affect the usefulness and reliability of students' evaluations of teachers. Noting that the frequency of students' evaluations and the lack of understanding of both how and why student evaluations are used, Dunegan and Hrivnak (2003) stated that student evaluations were: 1) cognitively taxing, 2) not perceived by students as having any impact on curriculum, and 3) just another routine chore. These factors, according to Dunegan and Hrivnak (2003), contributed to their study's results that revealed that student evaluations of their teachers were not reliable. Students were more likely to complete their anonymous, end-of-term evaluations in a mindful manner when they were not satisfied with their professors, while students who were satisfied with their instructors completed the evaluations in a mindless manner (Dunegan & Hrivnak, 2003).

A recent study by Roxa et al. (2021) addressed some of the benefits and shortcomings of end-of-term student evaluations and claimed that although student ratings of their instructors can be used as a tool for improving the learning process this can only occur when they are followed up with reflection and discussion on how to do so. In fact, Roxa et al. (2021) suggested that the research ascertained that the shortcomings can be detrimental for

both students and teachers if there is no follow up conducted. The dissatisfaction with the lack of quality assurance and biases in the instruments used for measuring quality has resulted in alienated instructors, weakened student voices, and the expression by some researchers that student evaluations should be completely abandoned (Roxa et al., 2021).

2.2.2 Faculty Perceptions

Additionally, studies show that online surveys may further skew the data with consistently lower response rates (Morrison, 2013). Some faculty have a difficult time interpreting results (Wong & Moni, 2014) and while many review and reflect on the data they understand, very few aim to interpret results that are confusing to them.

In order for faculty ratings to be beneficial the faculty must have a positive perception of these student ratings. Literature shows that even though faculty know there are major flaws in using student ratings to measure teaching quality they still feel positively about the student ratings (Stein et al., 2013). Other studies show that faculty are skeptical and mistrust them (Hammer et al., 2018). Due to the differing perceptions faculty have on student ratings Roxa et al. (2021) argue that these should not be the driver of teaching quality and innovation and can sometimes make matters worse for teaching effectiveness and capturing the voices of students.

Ultimately, the question remains whether student ratings should continue to be used and play a role in higher education. The authors of one study (Roxa et al., 2019) feel that when too much emphasis is placed on student ratings to measure quality it is likely to impede innovation in the classroom. For new teachers, this can negatively impact working conditions and actually alienate the student's voice. Roxa et al. (2021) are not suggesting that student ratings should go away but instead they should be the starting point to discussions regarding which teaching practices would be the most effective. It is imperative that we engage students in discussions about their learning experiences. Roxa et al. (2021) concluded that the problem with student ratings is not the quality in the actual student rating but a continuation in multi-sided efforts to find solutions or generate improvements in teaching strategies. The question remains if this is actually possible to start the discussions of learning quality with student survey results instead of it being the ending point.

2.2.3 End of Term Students' Evaluations, Grade Inflation and Accountability

The ability to instantly communicate with people throughout the world, utilize advanced technology, and deliver goods and services to any and every country has created a highly competitive global economy. Whether it's supply-chain management, talent recruitment, or resource management, being a successful employee or business owner requires a high level of analytical skills and attention to detail. Similarly, as information and knowledge increase, other professions are also experiencing rapid advancements in their fields. The National Research Council (2012) referred to a deeper-learning blend of knowledge and skills as "21st century competencies" that are necessary for individuals to be able to apply critical thinking skills and problem-solving techniques in increasingly sophisticated and complex work environments. Mainert et al. (2017) supported this perspective when they noted that higher-order thinking skills that are necessary for solving complex problems are critical for occupational success.

At the same time, however, academia is demonstrating that students are being challenged less to acquire these necessary skills. Blum (2017) noted that there is evidence that grade inflation has weakened the standards of excellence within educational institutions which has made it difficult to accurately assess student knowledge and competency levels. Blum (2017) also noted that grade inflation results in students being less able to assess their strengths and weaknesses, so they are unable to objectively evaluate what they have learned or where they truly rank among their peers. Further, it prevents students from experiencing the tough situations that occur from receiving honest and accurate feedback which is crucial for higher learning and the development of necessary cognitive and decision-making skills (Blum, 2017). The result is graduating students who are not fully prepared for the workforce.

There is substantial evidence that grade inflation is a reality. Chowdhury (2018) noted that grade inflation has become the norm in colleges and universities around the world and is a problem that, although fully aware of it, administrators are either not willing or not able to fix. The implications are very dangerous, though, for students, instructors, educational institutions, and entire societies, as students are receiving similar grades as students in the past but acquiring less knowledge, or they are receiving higher grades than students in the past, but acquiring similar amounts of knowledge (Chowdhury, 2018).

Chowdhury (2018) examined the factors that contribute to grade inflation and how grade inflation creates problems for students, faculty and society. He noted that the factors that contribute to grade inflation for faculty can be summarized into four different areas. First, many academic institutions use student feedback to evaluate

an instructor's performance in the classroom. This student feedback has a direct correlation with how faculty is perceived and the opportunities they have within their institutions. Ultimately the feedback can affect job security and have financial consequences (Chowdhury, 2018). Second, faculty can save time by grading and justifying low grades to students if the faculty takes the easier road to inflate grades. Third, students tend to gravitate toward faculty who award higher grades which causes those classes to fill. Part-time faculty have a higher risk because their job security varies by semester ratings and those ratings can determine if they are offered courses or not. If their classes are not as popular as other faculty members who tend to award higher grades for mediocre work, they risk not being offered a course every term. Finally, some faculty may use grade inflation out of concern for student motivation and psychology and life prospects.

Chowdhury (2018) also discussed the factors that contribute to grade inflation for institutions. First, there is competition from student enrollment between institutions. Second, there is the idea that faculty are here to serve and please students. Third, institutional staff subscribe to the myth that high grades and success are linked. Chowdhury argued that these factors contribute to grade inflation for institutions even if the institutions claim that they are against grade inflation.

Higher grades that are not reflective of higher learning can mislead students regarding their capabilities and preparedness (Chowdhury, 2018). Further, grade inflation can hurt the students whose grades don't get inflated. For example, a higher performer that has the same grade as a lower performing student can encourage mediocrity and the "why should I bother" attitude among star students (Chowdhury, 2018). This can deteriorate the work ethic for all students. Additionally, grade inflation could ultimately influence the course selection process for students where they may take a course that is not in their best interest but, due to believing that the teacher is an easy grader, they choose the course anyway (Chowdhury, 2018).

Grade inflation can also create problems for institutions. Grade point averages (GPA) have typically been a good indicator of proof of credentials and mastery of concepts (Chowdhury, 2018). Grade inflation can cause the GPA to lose value over time and lead to concerns about the credentials of an institution. Additionally, students with inflated GPAs may enroll in respected institutions that they are not truly prepared for, and, ultimately, this could negatively impact the academic standings of these schools and cause their degrees to have less value in the market (Chowdhury, 2018). Graduating students who received inflated grades could also enter the workforce unprepared, which will also eventually reflect negatively on the institutions that have awarded them their degrees. Blum (2017) shared these beliefs and stated that the effects of grade inflation are far reaching, as it contributes to a less-prepared workforce, damaged college reputations, the decreased perceived value of earning a college degree, and various other negative societal and business-related effects.

Grade inflation can create problems for society, as it can increase social disparities and inequalities (Chowdhury, 2018). Students who are able to afford to go to schools that inflate grades may have increased opportunities for future advantages in society. Additionally, if some schools inflate grades where others don't, the students from the institutions that use grade inflation have greater educational opportunities when they may not have the mastery of concepts needed to be successful in the next level of education (Chowdhury, 2018).

There are numerous consequences of grade inflation and its implications are especially important for online universities, as the online delivery platform for education is a relatively new concept. Recognizing the gap in the literature regarding grade inflation at online universities, Blum (2017) conducted a study of three regionally accredited online universities in the United States.

Blum (2017) conducted a qualitative case study of online universities, to better understand the issue of grade inflation at online universities and explore potential solutions to reduce it at U.S. online universities that are regionally accredited. According to Blum (2017), data saturation was reached with 27 participants who were interviewed via Skype with the condition of confidentiality. While two participants expressed that they either did not believe that grade inflation existed or that it was an issue, the remaining professors stated that they believed that it does exist and that it is a problem in higher education (Blum, 2017). The participants identified various areas that they believed may contribute to grade inflation. Based upon the data, Blum (2017) listed nine potential solutions including: (a) use grading rubrics, (b) revise student evaluations, (c) reevaluate academic policies, (d) use objective exams, (e) provide instructor training programs, (f) take instructors out of grading, (g) use a pass/fail grading system, (h) replace GPA with ranking, and (i) implement best practices that prevent good faculty members from being penalized for unsubstantiated negative student evaluations. Blum (2017) stated that a particular concern is that a faculty members' careers may be negatively impacted by disinterested students who are asked to assess an instructor's teaching effectiveness. Equally troubling is the admission of some faculty members to purposefully inflating grades out of fear of negative student reviews. In addition to the disconnect,

Blum (2017) suggests that there is evidence that the faculty who inflate grades often experience increased job security and financial benefits, more than instructors who grade conscientiously. Finally, Blum (2017) suggested that the pressure to participate in grade inflation actually begins at the universities' administration levels, as student retention and financial aid requirements may be more concerning to administrators than whether or not faculty are being unfairly evaluated by their students.

2.2.4 Absence of Face-to-Face Interaction and the Keyboard Warrior Effect

As online education and working remotely have become more commonplace, face-to-face interaction has decreased. Further, the increased usage of online social media platforms has fostered a negative change in traditional communication etiquette, respectfulness, and acknowledged roles of authority. The ability to lash out at people behind the safety of a computer screen, and with the ability to do so anonymously, has resulted in an onslaught of vitriol that can cause unprecedented and undeserved harm to people. According to the Cambridge Dictionary (2022), a keyboard warrior is "a person who makes abusive or aggressive posts on the internet, typically one who conceals their true identity." Further, although anonymity can foster the sharing of true feelings without the fear of judgment or repercussion, it can also provide a breeding ground for destructive behaviors and the lack of accountability. The nature of globalization has increased interfaces among more persons and more often across many places on earth. The Markkula Center for Applied Ethics at Santa Cruz University (2017) noted that the internet has provided a platform for people to take on a persona that would not normally be revealed in person. The internet has proven to be an invaluable tool for multiple sectors of employment, education, and information, but the darker side of it has also revealed itself in the form of unsubstantiated information, hurtful dialog, and detrimental cyberbullying. The safety of internet anonymity has contributed to a relatively uncontrolled source of hate speech and vengeance. In fact, although it enables people in extremely remote areas to connect to the rest of the world and provides billions of people with the ability to access seemingly endless information in a matter of minutes, it has also provided an extremely powerful weapon to those who, either knowingly or unknowingly, can have devastating and irreparable effects on people's careers, mental health, and lives.

The foregoing discussion has outlined how and why student evaluations have proven troublesome for university instructors. Although well-intentioned as a source of information to contribute to improved learning environments and teaching practices, they have also become a tool for students to retaliate against instructors with whom they are not pleased, regardless of whether their disgruntled perspectives are warranted or deserved. Further, they have also proven to be an unreliable, yet frequently relied upon, source for administrators when making decisions regarding faculty members' promotions, course assignments, and pay.

3. Discussion

There is a growing body of literature that reflects the ever-increasing negative effects of student evaluations of teachers. Although there are various potential positive outcomes of SETs, there is a strong concern about their negative implications. The current literature revealed that faculty members have experienced detrimental personal, professional, and financial impacts as a result of negative SETs. The literature has revealed that, due to these negative impacts, faculty members have become fearful of holding their students accountable. Further, coursework has become less challenging and grade inflation has been demonstrated throughout the educational institutional system. It has also been demonstrated that faculty have become fearful of end-of-term evaluations as a performance review by unqualified and untrained evaluators who do not usually have an understanding of how or why the courses are structured and taught the way they are, as students don't usually have access to the course development or lesson plan processes. Having a vested interest in receiving a high score for class performance, the students' responses in an end-of-term survey could be swayed by not receiving the desired score, even when it was undeserved. Therefore, feedback that administration may view as being helpful in improving faculty- and institution-performance can be used as a form of retaliation from a disgruntled student. Further, rather than being used to benefit all stakeholders, the SETs can be used to hurt, if not destroy, careers of innocent and good instructors. Finally, the quality of education has been affected, as instructors have been pressured to reduce student workloads and decrease opportunities for students to apply critical thinking, build analytical skills, and engage in problem-solving activities. Educational quality affects the ultimate goal of students to become employed and be successful in adequately fulfilling work tasks and responsibilities. Instructors, students, and education institutions could, indeed, benefit from providing a platform for students' voices to be heard. However, when that resource is used in a vengeful manner, and when the feedback of disgruntled students is taken at face value and not validated, the negative effects have proven to be harmful to the entire educational system and learning process.

4. Conclusion

The above literature review reveals that the answer to the first two questions posed by the authors is yes. The literature emphasizes how these negative impacts often lead to faculty members becoming fearful of holding their students accountable. It is evident that the increased importance of end-of-term student evaluations has given students too much authority over faculty success, and that there is evidence that more lenient professors are being promoted more often than stricter professors. Further, the increased importance of end-of-term student evaluations has contributed to grade inflation and/or resulted in quality professors being penalized for holding students accountable for higher quality performance.

5. Limitations and Recommendations

The question the authors were not effectively able to answer via a current literature review is the final question: Has the increased access to the internet and online education platforms resulted in a less personal environment whereby students are more likely to provide harsh feedback than in traditional face-to-face higher education settings? Even though there is a growing body of literature that is addressing this concern, the Yelp! Effect and the Keyboard Warrior Effect are both emerging phenomena. Further, the current literature is mostly related to customer service and business scenarios. Therefore, further research is needed to answer this final question that is directly related to academia. The Corona Virus pandemic forced education institutions around the world to adopt online teaching platforms. In turn, many of the institutions discovered that remote learning is a viable and effective educational platform. Additionally, students have demonstrated that the online platform, either solely or in combination with the traditional classroom format, is a choice that they want to have beyond the pandemic. Therefore, the opportunities for studying online student evaluations and how they compare to face-to-face student evaluations are growing.

References

Blum, D. (2017). Nine potential solutions to abate grade inflation at regionally accredited online U.S. universities: An intrinsic case study. *The Qualitative Report*, *22*(9), 2288−2311. https://doi.org/10.46743/2160-3715/2017.2914

Centra, J. A. (2009). *Differences in responses to the student instructional report: Is it bias*? Educational Testing Service.

Chowdhury, F. (2018). Grade inflation: Causes, consequences, and cure. *Journal of Education and Learning*, *7*(6), 86−92. https://doi.org/10.5539/jel.v7n6p86

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? *Journal of Marketing Education*, *31*(1), 16−30. https://doi.org/10.1177/0273475308324086

Dunegan, K. J., & Hrivnak, M. W. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *Journal of Management Education*, *27*(3), 280−303. https://doi.org/10.1177/1052562903027003002

Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*, *43*(4), 483−501. https://doi.org/10.1023/A:1015579817194

Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2008). Attractiveness, easiness and other issues: Student evaluations of professors on Ratemyprofessors.com. *Assessment & Evaluation in Higher Education*, *33*(1), 45−61. https://doi.org/10.1080/02602930601122803

Gelber, S. (2020). *Grading the college: A history of evaluating teaching and learning*. Johns Hopkins University Press.

Greenwald, A., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*(11), 1182−1186. https://doi.org/10.1037/0003-066X.52.11.1182

Juola, A. E. (April 1976). *Grade inflation in higher education: What can or should we do*? National Council on Measurement in Education Annual Meeting, San Francisco, CA.

Lanning, W., & Perkins, P. (1995). Grade inflation: A consideration of additional causes. *Journal of Instructional Psychology*, *22*(2), 163−168.

Luca, M. (2016). *Reviews, reputation, and revenue: The case of Yelp.com*. Harvard Business School Working Paper.

Mainert, J., Niepel, C., Murphy, K. R., & Greiff, S. (2019). The incremental contribution of complex problem-solving skills to the prediction of job level, job complexity, and salary. *Journal of Business and*

Psychology, *34*(6), 825−845. https://doi.org/10.1007/s10869-018-9561-x

McKenna, M. (2018). *The Yelping of the American doctor.* Wired.

National Research Council, Division of Behavioral and Social Sciences and Education, Board on Science Education, Board on Testing and Assessment, Committee on Defining Deeper Learning and 21st Century Skills, Hilton, M. L., & Pellegrino, J. W. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* National Academies Press.

Pacholski, S. L. (1992). Title VII in the university: The difference academic freedom makes. *The University of Chicago Law Review*, *59*(3), 1317−1336. https://doi.org/10.2307/1599866

Roxa, T., Ahmad, A., Barrington, J., Maaren, J., & Cassidy, R. (2021). Reconceptualizing student ratings of teaching to support quality discourse on student learning: A systems perspective. *Higher Education*, *83*, 35−55. https://doi.org/10.1007/s10734-020-00615-1

Stone, J. E. (1995). Inflated grades, inflated enrollment, and inflated budgets: An analysis and call for review at the state level. *Education Policy Analysis Archives*, *3*(11). https://doi.org/10.14507/epaa.v3n11.1995

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, *42*(4), 276−294. https://doi.org/10.1080/01973533.2020.1756817

Uijtdehaage, S., & O'Neal, C. (2015). A curious case of the phantom professor: Mindless teaching evaluations by medical students. *Medical Education*, *49*, 928−932. https://doi.org/10.1111/medu.12647

Uttl, B., & Smibert, D. (2017). Student evaluations of teaching: Teaching quantitative courses can be hazardous to one's career. *Peer J*, *5*, e3299. https://doi.org/10.7717/peerj.3299

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22−42. https://doi.org/10.1016/j.stueduc.2016.08.007

Zirkel, P. A. (1999). Grade inflation: A leadership opportunity for schools of education? *Teachers College Record*, *101*(2), 247−260. https://doi.org/10.1177/016146819910100203

Copyrights