The Influence of Teaching Effectiveness and Grading of Students' Work on how Students Evaluate Their Lecturers

Tadios Chisango, Wilberforce Kurevakwesu, Itai Mafa, Simango Tapiwa, and Samson Mhizha

Abstract

This article presents an experimental test of the effects of teaching effectiveness and grading on evaluation of lecturers by students. Although lecturers' grading of students' work presents a key confounding variable in studies that investigate the influence of teaching effectiveness on lecturer evaluations by students, most existing studies use correlational studies. This makes it difficult to separate the effects. In the present study, teaching competence and lecturer's grading of students' work were manipulated orthogonally, in a between-participants design, with a sample of Zimbabwean students, to test their effects on the students' endorsement of the lecturer , and also on potential lecturer evaluation. Hence, there were four experimental conditions: low-teaching competence-low grading, low teaching competence-low grading, high teaching competence-high grading and high teaching competence-high grading. The study tested the following specific hypotheses: (1) The low teaching competence-low grading condition would receive the lowest ratings on dependent measures; (2) The high teaching competence-high grading condition would receive the highest ratings,

ABOUT THE AUTHORS: TADIOS CHISANGO (Email: tadios.chisango@gmail.com, tchisango@staff.msu.ac.zw), WILBERFORCE KUREVAKWESU, and SIMANGO TAPIWA, Midlands State University, Zimbabwe; ITAI MAFA, Women's University in Africa; SAMSON MHIZHA, University of Zimbabwe, Zimbabwe

Chisango, T., Kurevakwesu, W., Mafa, I., Simango, T., & Mhizha, S. (2025). An Experimental Investigation of the Effects of Lecturers' Ostensible Teaching Effectiveness and Grading of Students' Work on how Students Evaluate them . *International Journal of African Higher Education, 11*(2), 42-57. https://doi.org/10.6017/ijahe.v11i2.18075

Results were largely in line with the hypotheses. These results show the benefits that accrue to instructors through giving students high grades. In particular, lecturers with low teaching competence can 'buy' better student ratings by assigning higher grades to students' work, while those with high teaching competence can enhance their ratings even more by giving high grades. Importantly, competent lecturers who grade their students lowly seem to be at the greatest disadvantage, in that they receive rather low ratings. The results indicate the flaws inherent in student evaluations of lecturers when their (lecturers') levels of competence are also taken into consideration. The results are further discussed in regards with the necessity to refining related research, and more rigorous evaluation methods of lecturers' performance in the classroom.

Key Words: lecturers, teaching effectiveness, student grading, quasi experiment

Résumé:
Nous présentons un test expérimental des effets de l'efficacité de l'enseignement et de la notation sur l'évaluation des professeurs par les étudiants. Bien que la notation des travaux des étudiants par les enseignants soit une variable confusionnelle clé dans les études qui examinent l'influence de l'efficacité de l'enseignement sur les évaluations des enseignants par les étudiants, la plupart des études existantes utilisent des études corrélationnelles. Il est donc difficile de distinguer les effets. Dans la présente étude, la compétence pédagogique et la notation ont été manipulées de manière orthogonale, dans le cadre d'une étude entre participants, auprès d'un échantillon d'étudiants zimbabwéens, afin de tester leurs effets sur l'anticipation de l'enseignement dispensé par le professeur, ainsi que sur l'évaluation potentielle de ce dernier. L'étude a testé les hypothèses spécifiques suivantes : (1) la condition « faible compétence pédagogique - faible notation » recevrait les notes les plus basses sur les mesures dépendantes ; (2) la condition « forte compétence pédagogique - forte notation » recevrait les notes les plus élevées. Les résultats ont été largement conformes aux hypothèses. Les résultats sont discutés en fonction de la nécessité d'affiner les recherches connexes et de mettre en place des méthodes d'évaluation plus rigoureuses des performances des enseignants dans les salles de classe.

Mots clés:
Conférenciers, efficacité de l'enseignement, notation par les étudiants, quasi-expérience.

The university classroom is a contentious place, 'administered' as it is both formally and informally, directly and indirectly, by the lecturer him/herself, the students and the university administration per se (Golish & Olson, 2000; Jacob & Lawan, 2020; Shor, 1996). On the face of it, the lecturer has almost complete control over the proceedings that take place within the confines of the classroom, with his/her influence spilling over even into the examination room. This includes selection of specific content to teach, the content and timing of coursework, leading all the way to the setting and grading of examinations (O'Brien et al., 2022).

Almost invariably, a lecturer's job is a deeply satisfying one, one characterised by a lot of discretion, flexibility, and a usually amiable and liberal scholarly environment (Baporikar, 2015). Teaching and inspiring a young adult population, researching on topics one pleases, managing own research grants, vacations, attending conferences and sharing research outputs with colleagues in the field, among other positive features of a lecturer's job, constellate into a special career. In particular, the discretion with which a lecturer conducts his or her duties is premised on the belief that having successfully scaled most or all of the existing steps of the academic ladder, he/ she is unquestionably a competent somebody (Chisango et al., 2022).

However, the lecturer still has to be answerable both to the body of students and the university administration. That forms the nexus from which most murkiness in a lecturer's job derives. In as much as lecturers are assumed to be a competent lot, their 'actual' levels of competence must still be assessed. For this, universities depend to some extent on peer evaluations, self-assessment, peer review, but mostly on student evaluations of lecturers (Rafsanjani, et el., 2020). Yet another measure of lecturers' levels of competence lies in the quantity, significance and impact of research output. In general, this is a much more objective measure, but this has problems of its own, in the sense that there can usually be an inverse relationship between time spent on, and therefore outcomes derived from research activities and teaching. A lot has been written, for instance, on the conflicting roles of lecturers as both researchers and teachers, and on which yardstick should take precedence in appraising lecturers' levels of competence (e.g., Karagiannis, 2009; Rafsanjani et al., 2020). A lecturer with an impressive research profile is certainly a great asset for a university, but he or she is simultaneously (or even primarily) a classroom teacher, who should also be valued as an asset by and to the students through excelling in the classroom, for that primarily is what they are at university for and what they pay fees for. Therefore, student ratings of lecturers are regarded seriously by universities worldwide. With its roots in American universities, the practice of student evaluation of teaching (SET) is now a

norm virtually across the world (e.g., Dev & Qayyum, 2017). As would be expected, SET has its own set of pros and cons. On the positive side, it can be a useful tool in fixing any glaring deficiencies within teaching practices and methods (Mohammed & Pandhiani, 2017)."Good teaching and good learning are linked through students' experiences of what we (lecturers) do. It follows that we cannot teach better unless we are able to see what we are doing from their point of view" (Ramsden, 2003, p. 2).

In Middle Eastern countries like Saudi Arabia and the UAE, which have in recent years registered leaps and bounds in extension of tertiary education, and where there has been a deliberate prioritization of quality of pedagogy over mass production of graduates, SET has become an issue of major focus. The widespread belief in these countries is that, holding other factors constant, competent lecturers will produce competent graduates who will perform well not only in the classroom and examination assessment, but also in the job market. As such, external national bodies such as the Commission for Academic Accreditation and Assessment (NACAAA) in Saudi Arabia have been bestowed with authority and autonomy to closely monitor the quality of education and ensure adherence to quality standards. In countries like Australia, a push towards performance-related funding has now made SET mandatory, with the ultimate goal of ensuring and safeguarding quality of higher education (Shah & Sid Nair, 2012).

However, greater focus has been given by faculty on the negative attributes of SET, given that it is plagued by problematic issues which have cast a shadow of doubt over the reliability and validity of the commonly used measures (Stroebe, 2020). Todd et al. (n.d.) notes three broad categories of such issues: (1) students are not qualified to evaluate faculty for teaching effectiveness; (2) the evidence for the validity of SETs for measuring individual teaching effectiveness has always been weak and is recently waning; and (3) many sources of bias in SET scores exist.

As such, the majority of academics view SET scores with skepticism; the general view is that that they are unreliable as an indicator of teaching effectiveness (e.g., Reckers, 1995). Critics of SET have mounted a multiplicity of arguments against its use over the years. On an ominous level, SET scores can be used by university administrations for punitive purposes, giving them control of faculty over a wide range of issues such as tenure, promotion, professional advancement and even dismissal (Ngware & Ndirangu, 2005). Furthermore, there is evidence converging on the conclusion that there are a lot of extraneous factors such as difficulty of course material, class size and demographic factors such as faculty gender, nationality, age and language of instruction in high school, which co-vary

with SET scores (e.g., Liaw & Goh, 2003; Boring, 2015; Pounder, 2007). The above argument serves as a pointer to another problem that has been noted of SET measures – their multiplicity and discretionary use across studies. For example, Mohammed and Pandhiani (2017) point out that despite there being a substantial body of literature around SET, there is little if any consensus among scholars in regards to their nature and number of items to be used to capture and measure valid and reliable scores. Indeed, an inventory they conducted on items from different studies revealed "diverse and contradictory results" (p. 3). In a more recent research, Esarey and Valdes (2020) argue that even unbiased, reliable and valid student evaluations can still be unfair and that even under "ideal" circumstances, SETs still yield an "unacceptably high error rate." They argue that unless the correlation between student ratings and teaching quality is far, far stronger than even the most optimistic empirical research can support, common administrative uses of SETs very frequently lead to incorrect decisions.

Perhaps the most striking problematic issue is how SET scores are related to grades, whether expected or actual. A robust finding in existing literature is that both actual and expected grades have a direct effect on SET ratings. In essence, expectations of a high grade have been found to lead to high SET ratings, whereas expectations of poor grades have led to low SET ratings (Pounder, 2007). Similarly, actual grades have also been shown to be directly related to SET scores. In one study that investigated this effect, Crumbley and Reichelt (2009) found that over fifty percent of faculty surveyed knew of colleagues who had deflated grading standards and lightened course content in order to boost their SET scores. Many other authors (e.g., Sacks, 1996; Schneider, 2013; Simpson & Siguaw, 2000) have reported similar tendencies, which include spoon-feeding examination content. The need to manipulate SET scores can reach very desperate proportions, as reported by Simpson and Siguaw (2000), who found that some university teachers go out of their way to dish free snacks on the day of evaluation, verbally reinforcing the students on their performance, or having an entertaining "fun" activity prior to the evaluation. By so doing, they would be trying to classically condition the students to give them high ratings, as demonstrated by some classic experiments in Psychology (e.g., Galizio & Hendrick, 1972; Janis et al., 1965).

To Crumbley and Reichelt (2009), all this represents a sad state of affairs in tertiary education characterized by an insidious "power shift and a move from professors running universities and colleges to students/administrators controlling higher education; where students simply punish ...professors on their SET scores" (p. 382). According to Dev and Qayyum (2017), poorly performing students may also give low rating to their lecturers in order

to protect their self-esteem, an argument that is in tandem with cognitive dissonance theory (Festinger, 1957). Hence, according to Tanner (2016), the current trends in SET have had the inadvertent impact of undermining teaching in universities and colleges

Study Objective and Hypotheses

One issue that is clear from the literature examined above is that it has been hard to disentangle the issue of the "actual" levels of teaching competence from other surrounding issues such as the difficulty of course content and student characteristics such as expectations of high versus low grades. Hence, rather than illuminating the issue of lecturer effectiveness, the existing trends in SET practices have further obfuscated its comprehension, leaving incessant cycles of controversy in their wake. Still, student evaluations have over time become a key indicator of teaching effectiveness at most universities and colleges worldwide (Liaw & Goh, 2003). Therefore, if there are misgivings about their validity, the way forward is to continue to prune them of their flaws, rather than dismissing them altogether.

One very important factor that has emerged in past research as a key deflector of teaching effectiveness from accurately predicting SET scores is student grade, whether prospective or actual (Crumbley & Reichelt, 2009; Dev & Qayyum, 2017). However, because most existing research on SET and student grading has been correlational (e.g., Culver, 2010; Waller, 2015), it has not been possible to empirically tear apart the causal effects of teaching competence vis-á-vis grade (or indeed any other such confounding factor) on SET scores. Although some existing research (e.g., Powell, 1977; Smith et al., 2011) tested the impact of student grading on instructor evaluation, this has never been done orthogonally against perceived lecturer competence. This leaves open the question whether these two variables share some interactive and tangible effects.

In other words, the fundamental issue is whether grades (actual vs. expected) would carry more or less weight than actual teaching competence scores in predicting SET scores. This is the main objective of this study. In essence, we orthogonally manipulated scores of teaching competence (low vs. high) against expected grade (low vs. high) in order to determine their interactional effects on SET-like scores. This means that we had the following four conditions in our design: (1) low teaching competence-low grading, (2) low teaching competence-high grading, (3) high teaching competence-low grading, and (4) high teaching competence-high grading. Although all the above permutations of competence and grading were tested,

we were particularly interested in the following two hypotheses:

1. The low teaching competence -low grading condition would receive the lowest ratings on dependent measures.
2. The high teaching competence-high grading condition would receive the highest ratings.

However, of more direct interest to this research was whether grading would offset the effect of teaching competence. To enable the testing of such effects would therefore behoove planned contrasts of the means of the conditions that had the same competence levels but different grading levels. This means that the following two contrasts were planned:

1. Low teaching competence-low grading versus low teaching competence-high grading.
2. High teaching competence-low grading versus high teaching competence-high grading.

If grading had no impact on the rating of teaching effectiveness, the conditions with the same teaching effectiveness scores would thus be expected to attract similar ratings.

Methods

Participants

Two hundred-and-one undergraduate students (102 men, 99 women) from the Midlands State University, Zimbabwe, took part in this study. Power analysis gave a coefficient of .812,, which indicated that the sample size was adequate. They reported ages ranged from 18 to 23 (mean = 20.72 years, SD = 1.14). Seventy (34.8%) of the participants were Social Work students, 88 (43.8%) were from the Department of Human Resources Management (HRM) and the rest (43 = 21.4%) were Psychology students. Thirty-five (17.4%) of them were first-year students (Mean age = 19.09 years, SD = 1.07); 92 (45.8%) were second-year students (Mean age = 20.77 years, SD = 0.87), whereas 73 (36.3 %) were fourth-year students (Mean age = 22.27 years, SD = 0.82). They were all black Africans.

Design, Procedure and Measures

This study employed a between-subjects experimental design with teaching competence (low vs. high) and student grading (low vs. high) as orthogonally manipulated independent variables. The dependent

variables were: (1) lecturer endorsement, measured by four items and (2) potential rating the student would give the lecturer, measured by a single item (see Appendix for the items). These two newly constructed measures showed a high criterion-related validity. To assess criterion-related validity, items were adapted from the teaching and learning dimension in the Professional Student Satisfaction Pre-Test Scale developed by Wang & Wu (2016). Exemplar items include 'I would be satisfied with the lecturer's teaching methods', 'I would be satisfied with the way the lecturer evaluates students' and 'I would be satisfied with the learning atmosphere provided by the lecturer in the class'. Correlations between lecturer endorsement and this scale was .63. In turn, potential lecturer rating also showed a high correlation of .58 with this scale. Hence, criterion-related validity of the new measures was high.

The experimenters, who are Lecturers in the School of Social Work and the Departments of HRM and Psychology, introduced the study as a student evaluation of a prospective lecturer who would soon take over from one of the lecturers that would be leaving his job. They then presented information which portrayed the in-coming lecturer as (1) low in teaching competence and low in grading, (2) low in teaching competence but high grading, (3) high in teaching competence and high in grading, and (4) high in teaching competence but low in grading. For instance, the high competence-low grading condition was introduced in the study as follows:

> As you might or might not know, one of our lecturers in the department/ school will leave the MSU soon to take a post at a regional university. A new lecturer will soon take over. The lecturer has provisionally passed his interviews at MSU. However, HR would like students to be involved in determining whether they are making the right choice, in line with new regulations. From the lecturer's previous employment, HR got the following details:
>
> 1. His/her average competence ratings as a lecturer as per peer evaluations (i.e., by two lecturers unknown to each other and unknown to the lecturer) on a scale of 1(bad) to 5(good) is 4.10 (this represented high teaching competence).
> 2. His/her average grading of students on essays marked out of 20 is 11/20 (this represented low student grading). This was followed by four self-report items which assessed the students' endorsement of the lecturer, all anchored on a 1(not at all) to 5(very much) scale (Cronbach's α= .88). Last was presented a single item that assessed the students' potential evaluation of the lecturer on a 7-point scale of 3(very poor) to 3 (very good).

Results

Preliminary Analysis

We computed descriptive statistics pertaining to lecturer endorsement and potential rating, as a function of condition (lecturer competence vs. student grading), as indicated in Table 1. As we had predicted, the high competence-high grading competence condition received the highest ratings, in terms of both lecturer endorsement and potential evaluations. The next highest ratings went to the low competence-high grading condition ahead of the high-competence-low grading one, giving initial support to our suspicion that high grading holds sway on students' ratings of lecturers. Predictably, the low-competence-low grading condition received the lowest ratings.

Table 1: Means and Standard Deviations for Lecturer Endorsement and Potential Evaluation

No	Condition	M	SD	M	SD
1	Low Competence-Low Grading	2.38	0.94	-0.93	1.73
2	Low Competence-High Grading	2.86	1.18	1.76	1.68
3	High Compentence-Low Grading	2.57	1.02	1.00	1.54
4	High Competence-High Grading	3.07	1.03	3.05	0.77

Main Analysis

The main analysis was in the form a one-way MANOVA, with Condition ([1] low teaching competence-low grading, [2] low teaching competence-high grading, [3] high teaching competence-high grading, and [4] high teaching competence-low grading) as the independent variable; Lecturer endorsement and potential lecturer evaluation were the two dependent variables. The multivariate main effect of condition was statistically significant, Wilks 'A = .61, $F(6, 478) = 18.65$, $p = .001$, partial $\eta^2 = .26$. In turn, the univariate main effects of condition on lecturer endorsement and potential evaluation scores were both statistically significant, MS =4.68, $F(3, 190) = 4.38$, $p = .006$, partial $\eta\ 2 = .09$; MS = 92.15, $F(3, 190) = 40.27$, $p = .001$, partial $\eta 2 = .44$, respectively.

To follow up on the univariate effects of condition, planned contrasts were conducted. The results confirmed that the high teaching competence-high grading condition was rated higher on lecturer endorsement scores than the average of all the other conditions, $t(197) = 18.42$, $p = .000$. Also as predicted, the low teaching competence-low grading condition was rated lower than the average of the other conditions, $t(197) = 16.04$, $p = .000$. Similar contrasts on the potential evaluation scores yielded the same results: $t(197) = 24.46$, $p = .000$; $t(197) = 3.38$, $p = .002$, respectively.

Further contrasts were run to test the differences in mean scores of the following conditions, as alluded to before, with regards to both lecturer endorsement and potential evaluation scores, in order to effectively test the differential effects of teaching effectiveness versus grading:

- Low teaching competence-low grading versus low competence-high grading, and
- High teaching competence-low grading versus high competence-high grading.

The results presented in Table 2 confirmed that the low teaching competence-high grading condition was rated higher on both dependent measures than the low teaching competence-low grading condition. Again, as was expected, the high-competence-high grading condition had higher ratings on both measures than the high-teaching competence-low grading condition.

Table 2: Contrasts Testing Effects of Teaching Competnce against Student Work Grading

Contrast		*df*	*t*	*p*
Low teaching effectiveness- low grading vs.low teaching effectiveness higher grading	Lecture Endorsement	197	5.15	.000
	Evaluation	197	22.27	.000
Low teaching effectiveness-low grading vs low teaching effectiveness higher grading	Lecture Endorsement	197	24.36	.000
	Evaluation	197	24.30	.000

Discussion

This study involved experimentally manipulating teaching competence against student grading to test their differential effects on lecturer endorsement and potential evaluation by students. As we pointed out earlier, most of existing research around lecturer evaluations by students has been correlational, hence rendering it difficult to empirically tear apart the causal effects of teaching competence versus grading (or indeed any other such confounding factor) on the evaluation scores. In the present study, we employed a quasi-experimental between-subjects design with teaching competence (low vs. high) and student grading (low vs. high) as orthogonally manipulated independent variables to test their interactive effects on students' possible evaluations of lecturers, in the form of lecturer endorsement scores and potential evaluation scores.

Initial, descriptive analysis provided preliminary support for the expected pattern of findings and the specific hypotheses. Indeed, as had been predicted, the high teaching competence-high grading condition received the highest ratings, in terms of both lecturer endorsement and potential evaluations. Also in line with the hypotheses, the low teaching competence-low grading condition received the lowest ratings. Interestingly, the low competence-high grading condition had higher ratings than the high-competence-low grading one, in this way underscoring that the effects of high grading can sometimes trump the effects of teaching competence.

The findings from the main analysis corroborated those from the preliminary analysis. For instance, the high teaching competence-high grading condition was rated higher whereas the low teaching competence-low grading conditions was rated lower than the average of all the other conditions. Similarly, as expected, the low teaching competence-high grading condition was rated higher than the low teaching competence-low grading condition on both dependent measures. Again, as was expected, the high teaching competence-high grading condition also had higher ratings on both measures than the high teaching competence-low grading one.

These findings clearly demonstrate the advantages granted by high grading to lecturers. Specifically, lecturers who score low on teaching competence can gain significantly higher ratings from students through higher grading of students' work, whereas those who score highly on teaching competence can further boost their ratings through high grading. In this whole set up, the biggest losers are the otherwise competent lecturers who "spoil the party" through giving students low ratings, and thus stand to be 'penalized'

by students through low ratings. In sufficient numbers and consistency, such low ratings can have a serious effect of stunting and/or derailing a lecturer's career (see Ngware & Ndirangu, 2005). Plagued with such glaring loopholes, it is thus no wonder that students' evaluations of lecturers are viewed with skepticism by the majority of academics, with the general view being that they are unreliable as an indicator of teaching effectiveness (e.g., Reckers, 1995). This renders the college and university background a place where uncanny strategy reigns, with lecturers bargaining for high ratings through high grading, and students rewarding them with high ratings or penalizing them with low ratings as they deem fit (Crumbley and Reichelt (2009). Lecturers may be driven to engage in other non-standard and/or unethical practices such as deliberately lightening course content, spoon-feeding examination content (e.g., Sacks, 1996; Schneider, 2013; Simpson & Siguaw, 2000), or even 'bribing' students with 'fun activities' and/or snacks, on the day of evaluation (Simpson and Siguaw (2000). All these effects constellate to form a fundamental deflector of teaching effectiveness from accurately predicting students' evaluations of lecturers (Crumbley & Reichelt, 2009; Dev & Qayyum, 2017). Viewed together with other problems typical of the evaluation process and outcomes, the common trend is that there is a shift away from, rather than towards, accurate evaluation of lecturers, creating in its wake problems for lecturers themselves, for students, between students and lecturers, between lecturers and university administration, and for the long-term future of academia in general.

Therefore, at least with regard to grading, there is an urgent need to formulate strategies that can help to tear apart its effects on student evaluations of lecturers from those of teaching effectiveness per se. Any such approaches need to see grading as an inherent confounding variable in the evaluation processes, such that without controlling for its effects, the evaluations themselves should be considered of no major significance. Considering that the evaluations carry a huge potential as a useful tool in fixing any glaring deficiencies within teaching practices and methods (Mohammed & Pandhiani, 2017), and that "good teaching and good learning are linked through students' experiences of what we (lecturers) do..." (Ramsden, 2003, p. 2), there is a need to make the evaluations much more objective, especially in relation to grading. One way of doing it would be to have lecturers rated at different universities from the ones where they teach, by similar sets of students, with more rigorous procedures such as reliability analysis being put into place. Furthermore, research into related issues may need to become more experimental, and less correlational, so that the findings become more robust and less prone to controversy and doubt. This can be better done through combining approaches, methods and ideas from related disciplines, such as Education, Psychology and Sociology.

Reference

Baporikar, N. (2015). Collegiality as a strategy for excellence in academia. *International Journal of Strategic Change Management*, 6(1), 59-72. DOI: 10.1504/IJSCM.2015.069522.

Boring, A. (2015, August 14). Female lecturers 'suffer from gender biases' in student ratings. https://www.timeshighereducation.com/news/female-lecturers-suffer-gender-biasesstudent-ratings.

Chisango, T., Muridzo, N. G., Mafa, I., & Khombo, S. (2022). An extrapolation of the stereotype content model to university lecturers. *Journal of Applied Social Science, 16*(1), 346-355. DOI: http://dx.doi.org/10.1177/19367244211026915.

Culver, S. (2010). Course grades, quality of student engagement, and students' evaluation of instructor. *International Journal of Teaching and Learning in Higher Education, 22*(3), 331-336.

Crumbley, D., & Reichelt, K. J. (2009). Teaching effectiveness, impression management, and dysfunctional behavior: Student evaluation of teaching control data. *Quality Assurance in Education, 17*(4), 377-392. Doi: 10.1108/09684880910992340

Dev, S. S., & Qayyum, N. (2017). Major factors affecting students' perception towards faculty evaluation of teaching (SET). *Journal of Social Studies Education Research, 8*(3), 149 – 167.

Esarey, J., & Valdes, N. (2020). Unbiased, reliable, and valid student evaluations can still be unfair. *Assessment & Evaluation in Higher Education, 45*(8), 1106-1120, Doi: 10.1080/02602938.2020.1724875.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Galizio, M., & Hendrick, C. (1972). Effect of musical accompaniment on attitude: The guitar as a prop for persuasion. *Journal of Applied Social Psychology, 2*(4), 350–359 Doi/10.1111/j.1559-1816.1972.tb01286.x

Golish, T. D., & Olson, L. N. (2000). Students' use of power in the classroom: An investigation of student power, teacher power, and teacher immediacy. *Communication Quarterly, 48*(3), 293-310. DOI: 10.1080/01463370009385598.

Jacob, O. N., & Lawan, A. (2020). Public universities administration in Nigeria: Challenges and the ways forward. *International Journal on Integrated Education, 3*(11), 163-169. DOI: 10.17605ijiev3i11.877

Janis, I. L., Kaye, D., & Kirschner, P. (1965). Facilitating effects of "eating-while-reading" on responsiveness to persuasive communications. *Journal of Personality and Social Psychology, 1*(2), 181–186. Doi: 10.1037/h0021644

Karagiannis, S. N. (2009). The conflicts between science research and teaching in higher education: An academic's perspective. *International Journal of Teaching and Learning in Higher Education, 21*(1), 75-83.

Liaw, S., & Goh, K. (2003). Evidence and control of biases in student evaluations of teaching. *International Journal of Educational Management, 17*(1), 37-43.

Mohammed, T. A. & Pandhiani, S. M. (2017). Analysis of factors affecting student evaluation of teaching effectiveness in Saudi Higher Education: The case of Jubail University College. *American Journal of Educational Research, 5*(5), 464-475. DOI: 10.12691/education-5-5-2

Ngware, M. W., & Ndirangu, M. (2005). An improvement in instructional quality: Can evaluation of teaching effectiveness make a difference? *Quality Assurance in Education, 13*(3), 183-201.

O'Brien, R., McGarr, O., & Lynch, R. (2022). Students' perceptions of lecturer power and authority in a higher education PBL business programme. *Teaching in Higher Education, 27*(5), 615-631.

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education, 15*(2), 178-191. Doi: 10.1108/09684880710748938

Powell, R. W. (1977). Grades, learning, and student evaluation of instruction. *Res High Educ 7*, 193–205. https://doi.org/10.1007/BF00991986

Rafsanjani, M. A., Pamungkas, H. P., Prakoso, A. F., & Sholikhah, N. (2020). Does teacher-researcher role conflict influence psychological well-being among the lecturers? *Tadris: Journal of Education and Teacher Training,* 5 (2): 287-296.

Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). Routledge Falmer.

Reckers, P. M. (1996). Know thy customer. *Journal of Accounting Education, 14*(2), 179-185.

Sacks, P. (1996). Generation X goes to college. *ERIC Digest*. ED399906 Open Court.

Schneider, G. (2013). Student evaluations, grade inflation and pluralistic teaching: Moving from customer satisfaction to student

learning and critical thinking. *Forum for Social Economics Journal, 42*(1), 122-135.

Shah, M. and Sid Nair, C. (2012). The changing nature of teaching and unit evaluations in Australian universities. *Quality Assurance in Education, 20*(3), 274-288. https://doi.org/10.1108/09684881211240321

Shor, I. (1996). When students have power: N*egotiating authority in a critical pedagogy.* University of Chicago Press.

Simpson, P. M. and Siguaw, J. A. (2000). Student Evaluations of Teaching: An Exploratory Study of the Faculty Response. *Journal of Marketing Education, 22*(3), 199- 21.

Smith, D. L., Cook, P, & Buskist, W. (2011). An experimental analysis of the relation between assigned grades and instructor evaluations. *Society and the Teaching of Psychology, 38*(4), 225-228. Doi: 10.1177/0098628311421317.

Steury, T. D., Dhanasekaran, M., Blair, M. (n.d.). Review on the use of student evaluations of teaching (SETs), with *recommendations for assessment of teaching faculty for promotion, tenure, and annual review.* The Teaching Effectiveness Committee, Auburn University.

Stroebe, W. (2020). Student evaluations of teaching encourage poor teaching and contribute to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology, 42*(4), 276-294. Doi: 10.1080/01973533.2020.1756817

Tanner, S. J. (2016). Storying the classroom: Storytelling and teacher evaluation. *English Teaching: Practice & Critique, 15*(2), 208-220.

Todd, K. (n.d.) We've misunderstood how to evaluate teachers for 100 Years: Plus bonus strategies for how to get students to adore you. https://toddkashdan.substack.com/p/we-misunderstand-teacher-evaluations.

Waller, L. (2015). An analysis of grades, class level and faculty evaluation scores in the United Arab Emirates, *International Education Studies, 9*(2), 66-72.

Wang, Y. S., & Wu, T. J. (2016). Student satisfaction scale development and application for sport management in China. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(5), 1429-1444.https://doi.org/10.12973/eurasia.2017.00678a.

APPENDIX
Endorsement of Lecturers Scale

No	Emdorsement	Not to all				Very much
		1	2	3	4	5
1	If this lecturer's module were not a compulsory one, I would be happy about taking it.					
2	If this lecturer's module were compulsory, I would be fascinated about being taught by him/ her.					
3	I believe this lecturer would provide a positive working and learning environment.					
4	Altogether, I suppose that I would be satisfied with attending this lecturer's module.					

<u>Potential Evaluation of Lecturers Item:</u>

Overall, what rating do you suppose you would give this lecturer in terms of teaching effectiveness, on a scale of **-3 (very poor) to three (very good)?**

Very poor						Very good
-3	-2	-1	0	1	2	3