

A Review of The Quality - Reliability and Validity - of Student/Teacher Evaluation Report Instruments

Ricardo Rodriguez*

University of the Southern Caribbean, Trinidad West Indies

***Corresponding Author:** Ricardo Rodriguez, University of the Southern Caribbean, Trinidad West Indies.

Received: May 30, 2024; **Published:** June 08, 2024

Abstract

Teacher evaluation forms in their current format are not useful appraisal instruments for performance testing and improving academic courses. These tools fall short in accuracy, validity and reliability in their design. A more useful approach should visit the potential for interactions between factor levels within the instrument together with hypothesis testing for variations in the mean responses solicited from the subjects which can harbor latent errors, misconceptions and imperfections in the findings.

A two-way policy agreement between teacher and student is therefore proposed in this paper for sign off by both teacher and student. The agreement is then tested for violations from both parties and whether there is any bias in the responses.

A statistical analysis is adopted to test the significance in the population differences in the means of dependent variables of choice in the instrument and for variance in the means of the factor levels used in the instrument. Logistic regression is finally instituted for self- test evaluation to determine the odds of a good performance rating (from a teacher), if you are a teacher over if you are not (student) for the same policy instrument. The scores for the teacher from the student on the teacher's instrument and the scores the teacher gives himself from the same teacher's instrument are compared.

The final product offers schools and universities an improved method for justifying students' evaluation of their lecturers while allowing the lecturers to evaluate themselves and can be extended for use in business organizations for conducting appraisals of workers. The instrument's reliability and validity are also tested as a quality control measure.

Keywords: Teacher evaluation; statistical analysis; odds ratio; appraisal instruments; quality control

Introduction

In a fast pace world students are demanding more from their lecturers not solely for knowledge but for success, high scores and grades in their course given the growing competition for jobs. There is therefore a movement away from knowledge based learning to time based learning given that access to the job market must be quick. Unfortunately, in students' haste to complete their electives, many knowledge gaps are left open. This does not auger well for job placement at the workplace as graduates may not have the required level of competence, expertise and work ethic. This complex model of students' wants, lecturer needs, institution policy and employer demand has to be addressed from the source which is the education system design.

In this light, the framers of student evaluation forms must think deeper as to what constitutes a useful assessment tool that exudes useful information about the lecturer's performance and course value. Other factors play an important role in how students respond and some of these include: course content and student's perception of its importance, environment and ambience for learning, ease of accessing information from staff students deem critical for their success, before trying to solicit feedback from students on any teacher appraisal (evaluation). Other facets of satisfaction in the learning experience that coexist include correct fees, refunds, speed of correspondence and strict policy guidelines for the student. (Davidovitch, Nitza. 2011).

Students blame their lecturers for receiving failing grades and poor performance but the genesis of the blame is usually not properly defined or explained with certainty for adjustments to be made to the teaching strategy thereby making the exercise frustrating and confusing. Remarks are based on the student's performance, personality, capability and obtaining high grades even when not deserved. The time is right for constructing suitable investigative student feedback tools in the teaching environment to determine the source of grievance and justification of responses on evaluation forms. Such good investigative tools cannot be a cut and paste exercise from non-reputable quarters. (Davidovich1 and Eckhaus, 2019).

The increased competition among local educational institutions to attract student enrolment exacerbates the problem as students are not likely to pursue courses at a university they perceive will take too long to complete. (Davidovitch & Eckhaus, 2019). Schools are therefore over extending themselves by committing to promises that are not in keeping with best academic practice which now allows the student to enter the classroom with a preconditioned assurance of success in a course even with sub-standard assignments, high absenteeism from classes and exams, and, little or no class participation. The student then expects the lecturer to give a good passing grade for poor work. This is a dangerous precedent especially in areas of science, engineering and medicine where accuracy and precision are critical. This trend however should not be assigned only to these disciplines but all others for improving quality assurance. The question is then whether there is an association between what student advisory communicates to the students and the expectations of the student and if so, what other variables are of importance for testing the accuracy or justification of the students' feedback of their lecturers?

While education can be considered a right, it is also a responsibility which lies with the student to monitor progress without criticizing educators when there is failure in the course. The problem is further amplified when universities allow too much freedom to the students to criticize lecturers through student evaluation forms which can be considered a weapon for revenge and hate rather than a tool for building relationships, mutual understanding, development and trust.

The recommendation is for a student-teacher academic performance policy instrument that presents and measures a set of agreed-to teacher deliverables, student responsibility and assessments with deadline dates. Proper record keeping and documentation will be expected for accurate accounts of behaviour. The student can no longer request make-up exams and extensions on assignments and exam dates if not in the policy. Each student will be assessed by the lecturer and each lecturer assessed by the student for having kept his part of the policy agreement. The instrument must be constructed with small tolerance and clearly defined variables of interest. A comparison can then be made between teacher evaluation of the student and student evaluation of the teacher using statistical analysis of the findings, based on the respective policy agreement variables from both parties.

Literature Review

Students assessments of lecturers are sometimes personal and do not reflect the quality of teaching experienced. This may breach student teacher thresholds leading to tense relationships. Lecturers are inclined to give high grades to students in exchange for good evaluation scores which may contribute immeasurably to tenure at the institution. The environment for learning includes a classroom that is conducive to learning with suitable air circulation and lighting, benches, room size and board. (Chen & Hoshower, 1998), (Wachtel, 1998). These all add to the students' experience and tendency to introduce errors of bias in teacher evaluation forms. The time for lectures is also a factor, especially if coming after a difficult one or sometime later in the afternoon when the student is fatigued. The

comments made are in no way to condone student feedback or indicate as irrelevant but rather whether effective and efficient teaching will be obtained. With strict guidelines on reporting the feedback there is better chance for a more practical report that can assist in teacher delivery and development. (Ballantyne, Borthwick, & Packer, 2000).

Problem factors that give evaluation forms a high level of bias against the lecturer include: poorly performing students usually find it difficult to keep up and are not willing to seek the lecturer's help, they are also the students who tend to be the first to fill out the teacher evaluation form. In some circumstances misinterpretation of how a lecturer may have communicated with them may see the student take unwarranted offence. Students also do not like their work being critiqued leading to misunderstanding of the importance of this exercise in academia and in their preparation for the world of work. These all promote the unreliability and non-validity of the evaluation instrument. In all, students' interpretations would vary widely if all attempted the evaluation and too narrowly if only a few who were dissatisfied or satisfied participated.

Methodology

The lecturer's persona, competence in the subject, care, delivery style, are on test from observers who are neither experienced nor competent. These are however quality attributes and are in the opinion of the beholder for each person has a different perception of quality whether it is fit for purpose, use or what the student wants out of the course. In this way, the evaluation has to be thoughtfully designed with bias error avoidance in mind. A better evaluation will be one that facilitates a policy with agreed-to variables for both teacher and student and statistical tests for determining: the validity and reliability of the instrument (policy agreement variables), the variance in performance criteria for the chosen variables (factor levels) of the instrument, statistical significance in the difference in the mean scores for policy variables of interest and a regression model to determine the relationship between the odds of a high performance score for various predictors in the agreement.

Teacher-Student Agreement

This agreement is based on a policy statement and policy performance measurement instrument with variables that I have agreed to and willing to uphold all statements made by me.

Student agreement example:

1. I will be on time for classes.
2. I will be present for all classes.
3. I will be present for all exams.
4. I will be present for all exam feedback sessions.
5. I will complete all assignments on time.
6. I will visit all office hour sessions assigned to me.
7. I will query any grade and sign-off before the last day of teaching.

Teacher agreement example:

1. I will be on time for class.
2. I will be present for all classes.
3. I will give all exams on time and day as published
4. I will give feedback solutions on all exam work.
5. I will give assignments two weeks in advance.
6. I will facilitate students during my office hours
7. I will mark all assignments and assessments within one week of its completion and facilitate grade queries by end of teaching date.

I participated in this agreement, read and understood the policy index and this and willing to abide by the policy guidelines.

Teacher Name

Signature and date

.....

Student Name

Signature and date

.....

Procedure

As outlined, currently used evaluation instruments are too subjective and require a more calculated approach with count data and/or proportions. There are therefore a few steps to be taken before depending on the results from an appraisal.

Step1: An academic policy (instrument) must be agreed upon between student and teacher. The policy is built from student buy-in to determine whether there is agreement by the students based on a policy they helped generate and signed off to at the start of class. In this regard, both teacher and student will fill this evaluation form and we can perform hypothesis tests for statistical significance of the findings.

Step 2: Validity test for instrument measurement. Use of a statistical test incorporating ANOVA one factor analysis to determine which policy criteria influence teachers' evaluation scores from the students the most (the effects of policy criteria on performance rating or score by the student). This is repeated over four measures namely; start of term, mid-term, after in-house assessments and after final exams. This will indicate whether the mean evaluation scores are all equal or which are different if any during the semester. For example, there are seven different levels for the factor policy agreement. Is there any difference in performance rating score given by the student for the seven different policy levels over the four period testing? See table 1.

Factor - policy agreement	Evaluation score Responses over 4 period term			
1.I will be on time for class.	X_{ij}	X_{ij}	X_{ij}	X_{ij}
2.I will be present for all classes.	X_{ij}	X_{ij}	X_{ij}	X_{ij}
3.I will give all exams on time and day as published	X_{ij}	X_{ij}	X_{ij}	X_{ij}
4.I will give feedback solutions on all exam work.	X_{ij}	X_{ij}	X_{ij}	X_{ij}
5.I will give assignments two weeks in advance.	X_{ij}	X_{ij}	X_{ij}	X_{ij}
6.I will facilitate students during my office hours	X_{ij}	X_{ij}	X_{ij}	X_{ij}
7.I will mark all assignments and assessments within one week of its completion and facilitate grade queries by end of teaching date.	X_{ij}	X_{ij}	X_{ij}	X_{ij}

Table 1

Step 2: Reliability test. To be administered using student grades over the four periods and evaluation scores as response to see if variations in grades obtained in assignments produces a difference in the mean evaluation score for each level and if so which ones? This test will indicate if there is any statistical significance in the claim that teacher evaluation scores from their students are influenced by grades. See table 2.

Factor - period	Factor - grades			
	A	B	C	F
Start of term	X_{ij}	X_{ij}	X_{ij}	X_{ij}
Mid term	X_{ij}	X_{ij}	X_{ij}	X_{ij}
In house	X_{ij}	X_{ij}	X_{ij}	X_{ij}
Final exam	X_{ij}	X_{ij}	X_{ij}	X_{ij}

Note that instead of using scores proportions of students responding A,B,C or F can be used as the scores.

Table 2

Step 3: Significance test. It would be interesting to determine if there is any statistical significance in the difference between violation scores obtained by both teacher and student. This will aid in us determining who is the 'bigger' violator of the policy agreement terms. Should it be the teacher, then low evaluation scores from the student are more justified. If however, the student is the greater violator, how can the teacher's low evaluation score from the student be justified? A t-test can achieve this objective by obtaining the mean of the scores of policy 1 to 7 for teachers and same for students. See table 3 for a sample of 20 students and 17 teachers.

Variables	Students. Total Score for NO to agree that teachers keep policy items 1 through 7	Teachers. Total Score for NO to agree that teachers keep policy items 1 through 7
1	13	11
2	14	12
3	11	8
4	5	6
5	8	10
6	10	12
7	16	14
Mean score	11	10.43

Yes =1, No = 1 are the scores for each answer.

Table 3

The mean scores for all the policy variables from both teacher and student are calculated together with the standard deviations. The two means are seen to be 11 for students not agreeing that teachers keep all the policy agreements and 10.42 respectively, for teachers not agreeing that students kept all the policy agreements. Is this difference statistically significant for a difference in population means for us to justify that teachers violate more than students? Sample sizes being 20 and 17 respectively for students and teachers. A hypothesis test for this can be conducted for support.

Step 4: Self-analysis test. We finally conduct a self-analysis test with the teacher instrument, using logistic regression to determine the odds of a good performance rating (from a teacher), if you are a teacher over if you are not (student). The scores for the teacher from the student on the teacher's instrument and the scores the teacher gives himself from the same teacher's instrument are compared. Example, for variable 4 on the teacher's instrument, what are the odds that a student will give his teacher a high score (agreement) as compared with what the teacher gives himself? The odds ratio is: odds that a teacher gives himself a high score / odds that a student gives him a high score. In words, the odds of a teacher giving himself a good performance rating over a student giving him one can be determined. This can also be done for the student.

Using each of the variables in the policy agreement as a predictor we can determine: odds for good performance score, p (good performance = y / student). The probability of a good performance score of the teacher given you are a student is $p(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_7 x_7}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_7 x_7}}$. The odds of occurrence = $p(x)/1-p(x)$ for both subject sizes. The log odds changes for every one unit change in the predictor 4.

Conclusion

Step 1 sets up the policy instrument agreed to by both parties.

Step 2 allows for us to test the instrument to determine if it actually measures its design purpose.

Step 3 gives the reliability based on bias of student marks and hence performance in the course over the semester.

Step 4 gives the statistical significance in the difference of the mean scores obtained on performance from both teacher and student when there is cross examination.

Step 5 allows for self-assessment of teachers and can also be done for students using the instruments separately. It gives the chance of getting high performance scores from a student over that of a teacher and, vice versa as the instrument changes. The test can also be used by including exam scores over the four term period.

The logistic regression sets the stage for predicting the change in odds for a one-unit increase in the independent variable (predictor). Say, for predictor variable 6, 'I will facilitate students during my office hours' what are the odds of a good performance rating from a student who has a low overall grade during the semester over one who has a high grade? This will give a prediction of the odds that a student will think highly of his lecturer for each of the instrument variables. Student grade of fail/pass in the course can also be added and used for analysis on odds of giving a high performance to the teacher using student input only. This will give some indication of bias on the student's part based on the grades they received.

The work here is completed but only just started for educators who would appreciate an appraisal instrument that protects them from unwarranted and unjustified criticism. It is hoped that in the unveiling of time whatever instrument so construed can be readily calibrated using some of the recommendations advanced in this paper. This appraisal design can be applied to other work environments for justification in worker evaluations.

References

1. Ballantyne R., et al. "Beyond Student Evaluation of Teaching: Identifying and addressing academic staff development needs". *Assessment & Evaluation in Higher Education* 25.3 (2000): 221-236.
2. Chen Y and LB Hoshower. "Student evaluation of teaching effectiveness: An assessment of student perception and motivation". *Assessment & Evaluation in Higher Education* 28.1 (2003).
3. Davidovich N and Eckhaus E. "Student Evaluation of Lecturers - "What do Faculty Members Think About the Damage Caused by Teaching Surveys?". *Higher Education Studies; Canadian Center of Science and Education* 9.3 (2019).
4. Davidovitch N and Eckhaus E. "What Do Academic Faculty Members Think of Performance Measures of Academic Teaching? A Case Study From a 10-Year Perspective". *SAGE Open* 13.2 (2019).
5. Davidovitch N. "Myths and Facts About Student Surveys of Teaching the Links Between Students" Evaluations of Faculty and Course Grades". *Journal of College Teaching & Learning (TLC)* 6.7 (2011).
6. Wachtel H. "Student Evaluation of College Teaching Effectiveness: a brief review". *Assessment & Evaluation in Higher Education* 23.2 (1998).

Volume 6 Issue 6 June 2024

© All rights are reserved by Ricardo Rodriguez.