Evaluating Teacher's Classroom Performance

Raul Ramirez-Velarde1[0000-0001-7186-1914], Laura Hervert-Escobar[0000-0003-2465-7106] and Neil Hernandez-Gress[0000-0003-0966-5685]

Tecnologico de Monterrey. Eugenio Garza Sada 2501 Sur, Col. Tecnológico, Monterrey, N. L., Mexico, 64849.

Abstract. Teacher evaluation in the classroom is a multifaceted challenge, with no one-size-fits-all solution. While student evaluations provide valuable feedback, they are limited by students' inability to accurately assess their own learning. This reliance on student evaluations alone can lead to biased assessments, grade inflation, and misconceptions about learning. To address these issues, a multidimensional approach to teacher evaluation is proposed. This approach incorporates various assessment methods, including on-site evaluations, standardized tests, portfolio reviews, and student surveys, to provide a comprehensive view of a teacher's performance.

Keywords: Teacher evaluation, Classroom assessment, Student feedback.

1 Introduction

As in any professional environment, the evaluation of professors at any university is an exercise that not only aids the institute in improving its educational processes but also assists professors in discovering their strengths and areas for improvement. This paper will exclusively discuss the evaluation of professors in the classroom, without considering their roles in administrative positions, scientific production, or participation in projects.

The evaluation of professors in universities plays a crucial role in the educational process as it aims to ensure the quality and effectiveness of teaching. This mechanism allows academic institutions to gather valuable feedback from students regarding the performance of teachers, thereby opening the door to continuous improvement. Evaluation has become an essential tool for measuring aspects such as clarity in communication, the ability to motivate students, and currency in content knowledge, among others.

However, its implementation poses challenges, such as the inherent subjectivity in student opinions, mapping the personal and professional characteristics of the professor to a score in the evaluation, intellectual production, and the need to balance quantitative and qualitative evaluation. In this context, the evaluation of professors emerges as a dynamic process aimed at raising educational standards and promoting academic excellence.

This paper aims to conduct a scientific evaluation of performance evaluation of professors in the classroom and the role that the Students Evaluation of Teaching (SET) plays in it by reviewing recent available literature. Based on the research findings, a reference model for the classroom performance evaluation of professors is proposed.

The assertion that the outcome of the SET significantly influences professor evaluation, salary increments, and future career trajectory is widely acknowledged. SETs serve as the primary means to measure a professor's performance in their interaction with students. There is a prevailing assumption that a higher evaluation score on the SET correlates with better student learning outcomes. However, is this true?

In this paper, we will address the following points:

1. Students' perception of their own learning.
2. Evaluation of professors and actual learning outcomes.
3. The illusion of learning: a definition of teaching fluency.
4. Factors biasing the evaluation of professors.
5. The persistence of professor evaluation: correlations over time and space.
6. The relationship between professor evaluation and their level of rigor.
7. The relationship between learning, professor evaluation, and grades.
8. Does professor evaluation incentivize poor teaching practices?
9. The role of social networks.
10. Alternatives for evaluating professors' performance in the classroom: A Tec Model.

To address these topics, a literature review has been conducted, summarized as follows.

2   Students' Perception of Their Own Learning

How proficient are students in assessing their own learning? It is reasonable to assume that students are adept at estimating how much they have effectively learned, especially considering that they have spent decades learning and dedicating most of their time to education. Initially, in the mid-20th century, experiments seeking to measure students' perception of their own learning found substantial evidence that students were adept at gauging their level of learning [1]. However, more careful measurements conducted later revealed that these results were **influenced by three biases**: a) publication bias, b) small sample bias, and c) disciplinary bias. **Publication bias** occurs because **results with a positive correlation are more likely to be published** in a journal than those with a null or negative correlation. **Small sample bias** implies that **correlations found in small or non-carefully selected samples must be very large to be statistically significant**. That is, if the sample size is small and the correlation is not close to 1, the result is not valid because the confidence interval in small samples is very wide. The most significant bias is **disciplinary bias**. The field **of Pedagogy consistently**, suffering from the two biases mentioned above, **finds a positive correlation between perception of learning and actual learning, while disciplinary areas such as business, science, and engineering have found results that are precisely the opposite**. See

Clayson's meta-analysis (2009) [1] and Uttl et al.'s meta-analysis of meta-analyses (2017) [2]. Perhaps initially, students could accurately evaluate their learning, and then this changed as the teaching system evolved. Thus, experiment after experiment has shown that students are poor at estimating their own learning. Under normal conditions, students consistently overestimate their learning, resulting in generally negative reactions when their actual performance is communicated to them. Not only that. Students may believe they have learned more than they have if the instructor speaks fluently and enthusiastically. And they may believe they have learned less than they have if the instructor speaks less fluently and lacks enthusiasm. The literature on this matter is vast. In addition to Clayson and Uttl et al., we have Papamitsiou and Economides (2014) [3], Deslauriers et al. (2019) [4], Clauss and Geedey (2010) [5], Stroebe [6], and many more references included within these articles. Uttl et al. is a meta-analysis of previous meta-analyses.

A very interesting result is found in Claus et al., where the difference between perceived learning and assessed learning reached a value of up to 36 on the Kruskal-Wallis test, when the limit is 8 [5]. The problem with this study is the small sample size. The sample contains about 80 evaluations when the minimum sample size is 384. More robust and interesting results are found in Papamitsiou et al., where, using pre- and post-tests, the difference between perceived learning both pre and post, compared against the measurement of actual learning, is up to 6.56 standard deviations, whereas the limit for statistical significance is approximately 2 standard deviations [3]. **This means that students are not bad at assessing their own learning.**

What does this mean for the evaluation of professors? We will discuss this in the following section.

3 Professor Evaluation and Actual Learning

The first article on student evaluations of their professors (SET) was published over 80 years ago, in 1927. In the 1960s and 1970s, the use of student evaluations expanded significantly in the United States, and currently, nearly all North America employs SET to assess the effectiveness of professors. Such evaluations are used for hiring, salary increases, promotions, selection for special projects, classification, department and school evaluations, among other purposes.

There are several reasons for utilizing SET. Among them, a) it is a quick, relatively inexpensive, and numerical way to evaluate teaching, b) it allows students to have a voice in the evaluation of their professors and academic departments, and c) students are the primary witnesses to how professors conduct teaching [1] [2].

Although there is abundant literature defending SET as a good indicator of learning, this literature is subject to **publication bias**, in the sense that it mainly comes from the field of pedagogy and typically has methodological issues such as lack of statistical

rigor in data treatment and small sample sizes [1] [2]. A vast array of studies, including a couple of meta-analyses that reassess the conclusions of several previous studies [1] [2] [3] [4] [7] [8], demonstrate that **the correlation between professor evaluation and actual learning is at best very low, commonly null, and even occasionally negative**. Uttl et al. found that **SET can only explain 1% of measured learning and that if the student's prior learning is considered, the relationship between SET and learning is not statistically different from zero** [2]. The positive correlation, typically low, occurs on rare occasions and is highly dependent on disciplinary area, professor, course level, and multiple other factors [1]. And in general, **the more objectively learning is assessed, for example, in analytical classes using standardized exams, the relationship between learning and SET disappears** [1]. To clarify, a negative correlation means that a professor with certain speaking characteristics gives a false sense of learning but teaches less than a professor who is rated worse. This will be discussed in the following section.

4   The Illusion of Learning: A Definition of Teaching Fluency

If students are poor at estimating their learning within a class, what are we evaluating? Actually, **SETs measure student satisfaction with the course**, which, while a genuinely useful measurement, is not a measurement of learning. SET is based on the student's perception of their professor's characteristics, such as their level of subject knowledge, clarity of expression, organization, enthusiasm, likability, fairness, availability, accessibility, sense of humor, personal treatment, etc. [2]. But there are many other factors, all outside the control of the professor, that influence student satisfaction with the subject matter and therefore influence SET. For example, whether the student has received the grades they believe they deserve, the student's interest in the topics of the subject, whether the subject is elective or the student is forced to take it, whether the course is analytical or quantitative or not, the level of the class in the student's career profile, the class time and duration, whether the student has had any disputes with the professor, whether the student finds the professor's accent strange, whether the professor is male or female, and even the well-known effects of height and attractiveness on perceived competence [2] [8].

As we have already mentioned, students are poor at estimating their own learning. Derived from this, multiple studies have shown that students feel they have learned more when the lesson is taught by a professor who achieves fluency in presentation than when the lesson is taught by a professor who disrupts fluency for active learning. Even when it is the same professor. Carpenter et al call this the "**Illusion of Learning**".

There are a couple of experiments by Carpenter et al. [7] and Deslauriers et al. [4]. In Carpenter's experiment, the same professor taught two classes. In the first group, the professor gave the class with great fluency, and in the second group with little fluency. In both experiments, a fluid class is defined as one in which the presenter stands upright, maintains eye contact with the students, does not consult their notes, and is expressive

in their gestures. Additionally, the professor must speak clearly, and the presentation must be well-organized with a defined structure. In the non-fluid class, the professor had a slouched posture, somewhat hiding behind furniture, did not maintain eye contact with the students, consulted their notes constantly, spoke in a monotone, and was not expressive in their gestures. The students gave a much better evaluation to the fluent version of the professor and expressed that they felt they learned a lot with this professor, while the students of the non-fluid version rated the professor much lower and expressed having learned little. Surprisingly, the **exams showed that the learning of both groups was the same.**

In Deslauriers' experiment, two classes were set up again. One was a fluid class, without interruptions for activities, taught by an excellently well-rated star professor, and the other class was taught using active learning considering the best practices of the discipline. This learning included difficult activities that students should perform within the classroom. Again, the **students of the fluid class rated the professor up to 20% higher**. However, when tests were conducted to assess learning, it was found that the students in the active class achieved higher grades, as expected. In both experiments, the results passed the test of statistical significance.

There is another experiment conducted in 1973 [8], called the "Dr. Fox" experiment, and its results are called the "Dr. Fox Effect." In this experiment, Naftulin, Ware, and Donnelly offered a course on the applications of game theory to human behavior that would be taught by the expert Dr. Myron L. Fox. Dr. Fox was a Hollywood actor who knew nothing about the topic. Additionally, the actor was asked to make the presentation less precise, including vague material, contradictory statements, redundancies, and some humor unrelated to the topic. Dr. Fox delivered the presentation in an enthusiastic and passionate manner. How was the presentation evaluated? 90% of the feedback stated that the presentation was well-organized, interesting, and contained interesting examples.

It is evident that students prefer a professor with fluent speech over one who engages students with activities. It is also evident that most students do not learn much in class and prefer to be entertained. There is a statement that, although an exaggeration, summarizes what we have seen. In 1965, journalist Donald R. P. Marquis said, "If you make people think they're thinking, they'll love you. If you actually make them think, they'll hate you" (quoted by Morley & Everett, 1965, p. 237). In short, the **best teacher is not the one who produces the best learning**, indicating that the problem of teaching and learning is truly complicated.

5   Factors Biasing Professor Evaluation. Honesty in Evaluation.

Apart from fluency and enthusiasm, there are numerous factors that influence professor evaluation that are unrelated to learning. Some of these factors may be controlled by the professor, while others are not.

We have already mentioned some factors. For example, whether the student has received the grades they believe they deserve, the student's interest in the subject topics, whether the subject is elective or mandatory, whether the course is analytical or quantitative, the level of the class in their career profile, class time, and duration, whether the student has had any disputes with the professor, whether the student finds the professor's accent strange, whether the professor is male or female, and even the well-known effects of height and attractiveness on perceived competence [1] [2] [8] [9].

In Carpenter et al. [8], studies on biases in SET are mentioned containing many bibliographic references that we will not mention here. The following biases have been found:

1. One of the most documented biases is **gender**. Most of the time, men are better evaluated than women, although occasionally it is the other way around. It all depends on what is being asked. If the question is exclusively about effectiveness, men are better evaluated. However, in interactions, women are usually better evaluated.
2. **Leniency and severity** in grading also affect SET. We will address this topic in the following section.
3. The **age** of the professor is negatively correlated with SET. That is, younger professors are usually better evaluated than older ones.
4. The **appearance** of the professor affects SET. More attractive professors receive better evaluations than those perceived as less attractive.
5. **Race and accent**. Multiple studies in the United States show that Caucasian professors or those who speak without an accent are better evaluated than non-Caucasians or those who speak English with the accent of a non-native English speaker.
6. And although Carpenter et al. do not document it we also have **height**, it is known that taller people in general are perceived with more authority and competence.
7. Even **chocolate**. In an experiment conducted by Youmans and Jee in 2007, where students taking two statistics courses and one research methods course were enrolled in either of two discussion groups on course topics. Halfway through the course, it is evaluated using the discussion groups. Just before evaluating the course, one discussion group received chocolate and the other did not. **The students in the discussion group that received chocolate rated the professors better than those that did not receive chocolate**. There was no statistically significant difference in grades between discussion groups in each course [8].

SET is partly a memory exercise. Students must remember relevant aspects that have happened in the class over a period of several months. And it is a well-known fact that feelings influence memory [10]. Facts with emotional connotations are remembered more easily. Therefore, students will mainly remember facts that emotionally affected them positively or negatively. If there was no emotional connotation, it is likely that students will not remember enough of the class to conduct an objective SET.

In another experiment mentioned in [8], conducted by Uijtdehaage and O'Neal in 2015, in a medical course normally taught by several professors, **fictitious professor names**

were included. And although students could select **"Not Applicable,"** only **34% of students selected this option**. The rest, **66%, provided creative evaluations of the fictitious professors**. In an additional experiment conducted in 2008 by Brown, it was found that some students used SET to **retaliate** against professors for not receiving the grade they expected, that **36.5% of students personally reported knowing other students who made evaluations with false comments** about the professor because they didn't like them, and it is estimated in the same study **that 30% of SETs contain deliberately false information introduced by students**.

6   Persistence of Professor Evaluation: Correlations Over Time and Across Groups

SET exhibits persistence within the same group and across groups, as well as over time. This means that professors are often evaluated similarly at the beginning and end of the course. If they teach other courses, they are also likely to be evaluated similarly. It is rare for a professor to receive a low evaluation in one course and a high evaluation in another. The latter can only occur if there are significant differences in the courses. For example, students who take a course voluntarily tend to rate their professors better than those who are compelled by the academic program. Students in advanced semesters tend to rate their more demanding professors better, etc. As mentioned, SET depends on many factors that cannot be controlled by the professor. However, this persistence in evaluation indicates that **SET is based on relatively stable aspects and therefore difficult for the professor to change**. Evidence indicates that **each professor has a baseline evaluation related to their appearance, background, and personality**, which has nothing to do with student learning. Therefore, some professors may need to make significant changes to alter their baseline evaluation, or they may need to exert more effort than other professors to compensate for their inherent disadvantages. Ewing found that in SET, under the same conditions, simply changing the professor altered the intercept of regression models, i.e., the baseline evaluation. This is known in statistics as a **fixed-effect parameter**, indicating that the instructor has a **predetermined average evaluation** [9].

7   The Relationship Between Professor Evaluation and Their Level of Rigor

One of the oldest questions that has been raised is the relationship between a professor's rigor and Student Evaluation of Teaching (SET). Is it true that more demanding professors are evaluated worse than more lenient ones?

The quick answer to this question is yes. The **correlation between a professor's rigor and SET is negative**. This can be observed in [1], [8], [9], and [11], along with many references to experiments contained in these articles. Generally, when a professor is perceived as demanding in grading and in designing activities (creating more

challenging activities and grading them rigorously), they will receive a lower SET than a less demanding professor in the same circumstances (and with the same students).

Furthermore, the effects of rigor and lack of fluency in teaching can combine to create the perfect storm. Under certain circumstances, for example, in advanced elective courses where students voluntarily enroll, students may accept rigor as fair and concede that it enhances learning. But if students do not perceive significant learning, as would be the case with a professor who teaches with little fluency, they will consider high grading rigor very unfair and evaluate this professor poorly.

Now we know that rigor affects SET, but to what extent? This is difficult to estimate due to the challenge of measuring rigor in evaluation while keeping all other factors constant. However, it has been established that the negative correlation between rigor and SET is clear and has been demonstrated through various experiments. However, a relationship that is much more documented is the relationship between SET and grades. Since grades are easily available, there are many studies exploring this relationship.

8   The Relationship Between Learning, Professor Evaluation, and Grades

Multiple studies show a positive relationship between SET and student grades. Not only that, but it has also been proven that there is a relationship between SET and **the grade that students expect to receive in the course**. And this expected grade serves as a proxy for rigor. In other words, students expect lower grades with demanding professors, and they tend to give them a lower evaluation.

This relationship between the grade and the expected grade and SET is discussed in [1], [8], [9], [12], [13], and other studies mentioned within these articles show a **positive relationship, ranging from low to moderate, between grades and SET**. However, Ewing and Clayson and others have found **a strong relationship** between the grade expected by the student and SET [1], [9], as well as Lindsay [12] and Eiszler [13].

This association between grades and SET, which has been widely documented, is like the chicken or the egg problem. The relationship has been extensively discussed, and two hypotheses have been proposed. A) **Reciprocity**. Better-evaluated professors are better teachers, and therefore, students achieve better grades. Students feel they learn more and reward the professor with a better evaluation. B) **Leniency**. Students reward professors who grade leniently because the grade they expect will be higher, closer to the grade they believe they deserve. Both effects, reciprocity and leniency, blend and are difficult to separate. However, there is a moment in time when the effects do not blend, and that is the future. Both Clayson [1], Uttl et al [2], and Carpenter and Witherby [8] found that students who took introductory courses with professors with high SET, getting high grades on average, significantly lowered their grades in subsequent courses. So, there is evidence of leniency, although it cannot be said that all

professors with high SET are lenient. There is also evidence, at least to some extent, that the grade students expect influences SET positively. And there is compelling evidence that rigor negatively affects SET. The same articles mentioned also found evidence that **students who participated in classes where the assessment was rigorous in introductory courses performed better in subsequent courses** on the same topic, even if SET was not as high.

9 Does Teacher Evaluation Encourage Poor Teaching Practices? Grade Inflation and Student Work Deflation

The use of SET has its positive aspects. As previously mentioned, it is a numerical instrument from which information is easily obtainable directly from the main participants of the process. The cost of obtaining this information is moderate and allows students to express and reflect on their experiences in the teaching-learning process. It is a quality control instrument intended to improve teachers' classroom performance.

However, despite the advantages of SET, the imperfections of the instrument discussed earlier **lead to a series of problems in the teaching-learning process**. These are:

1. **Promoting enthusiastic and entertaining lessons that have little influence on students' true learning, reducing student engagement during class, and being generous in evaluation** [4] [8]. Students favor teachers who give fluid, entertaining, and enthusiastic classes. This way of teaching promotes passivity in students, a fact that has been proven in multiple studies. Remember that passive learning, even if interesting and entertaining, leads to inferior learning compared to learning obtained in a class where students have periods of active learning. Naturally, giving a fluid class is not bad and can be good, if this oratory is accompanied by effective learning activities in which students actively apply themselves to their learning.

2. **Promoting learning at the lower levels of Bloom's taxonomy at the expense of higher levels** [5]. Clauss et al. observed that students were more widely mistaken in their perception of their level of learning at the intermediate (and perhaps higher) levels of Bloom's taxonomy. This means that as teachers raise the cognitive level, and therefore the effort required to learn, students may feel that they are being asked too much and that they are not being graded properly [5]. This could lower teachers' evaluations and incentivize them to reduce cognitive demand to the lower levels of Bloom's taxonomy. Uttl et al. (pp. 40) report that **teachers who teach courses with more quantitative content receive lower evaluations on average than teachers who teach courses without quantitative content** [2]. While Butcher et al. report that departments teaching humanities and social sciences courses generally receive higher SETs than departments teaching economics and sciences [14].

3. **Poor student decisions regarding study habits**. Carpenter et al. report that students who took classes with fluent-speaking professors believe they have learned more, therefore, they dedicate less time to study [7]. While Babcock reports that the

expected grade significantly affects study time. Students who took classes where the average grade was A, dedicated 50% less time to studying than students who took classes in courses where the average grade was C [15].

4. **Lower performance in advanced courses**. Uttl et al.'s meta-study reports that the limited follow-up given to students' performance in advanced courses after taking basic courses indicates that students who took classes with fluent-speaking professors continuously had lower performance than students who took courses with less fluent professors that included more activities in the classroom [1] [2] [8].

5. **Deliberate selection of courses where a higher grade can be obtained with less effort**. Multiple studies indicate that students are much more likely to select courses where the expected grade is higher [1] [9] [14]. This is a bad study habit that can have consequences for student performance in more advanced courses.

6. **Grade inflation**. Grade inflation is a phenomenon that occurs in almost all higher education institutions in North America primarily but has also been observed in Europe. It is a widely studied phenomenon, which is why it is proven and documented. Grade inflation has multiple causes, and the relationship between SET and grades as a probable cause has been studied since the first SETs began. There is strong evidence that SET is, at least in part, a cause of grade inflation [1] [2] [8] [9] [12] [16] [17]. The degree of this relationship is unclear. However, although the relationship between SET and grades can be debated, there is no doubt that both **teachers and students believe that there is a positive correlational relationship** between them [1]. This belief alone can be one of the most powerful causes of grade inflation. Undoubtedly, grade inflation has multiple causes beyond SETs, and consequences, which will be studied in a separate document. But we can advance consequences, for example, a) loss of confidence that grades reflect students' preparation, b) reduction of time invested by students in their studies, c) promotion of bad study habits, d) false sense of achievement, e) unhealthy competition between departments and institutions, etc.

10   The Role of Social Media

An aspect not addressed in the literature is the influence of social media on teacher evaluation. As mentioned earlier, SET is persistent across courses and over time. Regarding social media and SET, the **first effect** they have is to **serve as a long-term social memory**. It is entirely possible for students to enter with an already made idea of what to expect from the teacher's teaching style and the grade they can expect based on the multiple comments that can be read on social media. This will make the persistence over time of teacher evaluation even greater. The teacher will have to work for a long period to ensure that a negative collective memory about their teaching style is gradually erased, replaced by a positive memory.

It is known that social media has negative effects on young people. These effects have been acknowledged by the companies that created these networks [18][1]. Social media is present in SET through real-time discussions among active students of the teacher. In this discussion held on social media, all aspects of the teaching-learning process will be discussed, but the teacher's opinion will not be necessary. In general, the social network consists only of students. Why is this important? This is the **second effect** of social media on education. We are talking about the **"influencer" effect** on SET. We define "influencer" as a person with the ability to communicate and influence decisions on specific topics of other people using social media. In Ding et al., and in many other easily accessible references, it is reported that an "influencer" can radicalize the opinions of others to one side or the other [19]. Ding et al.'s article presents research focused on politics. But it is easy to see how their conclusions extend to the educational field. Multiple studies show that social media can affect people's behavior, beliefs, and engagement in various areas of society. There is no evidence that education is different.

Another important aspect related to the **possibility of changing opinions or their perseverance**, which would be the **third effect**. This can be read in Lee et al. [20] and in Swale et al. [21]. In these studies, it is shown that when people are novices in a certain topic, that is, without a fixed opinion, they seek information about the topic and naturally consult the social media at their disposal; for example, a social network explicitly formed to exchange information about a class or a teacher. In this situation, their opinion can be changed by "influencers". However, after a certain time, the opinion will solidify and will not change because the person will only accept information that reinforces their opinions. Both effects can be good or bad for teacher evaluation.

Let's talk about the negative effect of the "influencer" on teacher evaluation. In the case of opinions that can be changed because students are not sure of their opinion, for example, when the teacher is new or at the beginning of their career, only a negative opinion from an "influencer" within the classroom is required to change the opinion of multiple students in the same direction. Hence, removing evaluations with extreme numerical values (i.e., eliminating the lowest and highest evaluations) in an SET is useless. The "influencer" moves multiple opinions, possibly many, not just the lowest or highest. And in the case a well-known teacher, and who has had previous experiences with the students of his current class or who is somehow known (for example, in the memory left in social networks), if the students' opinion regarding the teacher's teaching style is negative, this opinion will hardly change even if the teacher does a good job. Students with solidified opinions will selectively choose facts during the class that reinforce their opinion of the teacher. Hence the false or exaggerated comments that teachers often read when receiving their evaluations.

Naturally, if the influencer's opinion is positive, it will sway multiple opinions of their followers in the same direction. And if a person holds a positive opinion of a teacher

---

[1] Social media also have positive effects on young adults and adolescents. See the advisory on social media from the Surgeon General USA and the American Psychological Association.

that is already established, the teacher will have to have significant flaws for that opinion to change. This is a well-known effect of popularity, called the "Halo effect."

A **fourth aspect** is the teacher's involvement in social media. Students prefer their teacher to have the characteristics of an influencer: to be an active presence on social media, to have positive opinions on these platforms, to receive many likes, to be followed by many people, and to be seen interacting with other entities, individuals, and recognized companies. It is entirely possible that, in the future, young people will obtain both their learning and their news through social media, bypassing entirely traditional educational environments and media sources. It is not a bad idea for teachers to increase their presence on social media platforms.

11   Alternatives for Assessing Teacher Performance in the Classroom: A Tec Model

The SET is a survey, not a teacher evaluation. If it is intended to make SET a teacher evaluation, several aspects of its design must be changed.

1. Currently, the **SET cannot answer the following questions**: **a)** Why would a teacher with 20 years of experience and several periods of teaching the class receive a score of 0 or 1? **b)** What behavior or lack of skills corresponds to a rating lower than 50%? **c)** Why can a teacher who teaches the same subject to different groups, with the same syllabus, materials, and teaching techniques, receive statistically significant differences in evaluations? As long as these questions remain unanswered, the SET is too limited a tool to function as a classroom performance evaluation.
2. An even more intriguing question. It has already been mentioned that specific questions within the SET are not related to the final evaluation. **Why can a teacher receive high ratings in knowledge, fairness, and even teaching method, but receive a global low rating?** This point, along with the previous one, suggests that the evaluation may be based on emotional rather than strictly subjective factors.
3. **Reverse mapping**. Assume a teacher receives a low rating in the teaching method or any other variable in the SET. How can she correct this? What actions can be taken? The answer to this question is crucial to help teachers with poor evaluations.
4. Above all, if the SET is an evaluation, there must be **accountability**. When a student receives a rating of 67%, the teacher must account for each point missing from 100%. **Why can a student give a rating of 0% or 10% to a teacher without justifying anything?** If the SET is an evaluation, then it should be confidential, not anonymous. If a student makes a claim about the teacher, the student should be able to say when and where the incident occurred, and it should be possible to verify it with recordings or witnesses. **How is it possible for a student to lie in SET with impunity, affecting the reputation and career of the teacher without consequences?**
5. **Comparing teacher evaluation with the evaluation of other professionals**. Who evaluates a doctor? Other doctors. Certainly, the opinion of a patient is considered;

however, the performance evaluation of a doctor is done by tracking the progress of patients over time and considering the opinion of other doctors. How are engineers evaluated? Similarly, by reviewing the quality of their products over time and considering the opinion of other engineers.

However, just like in other professions, it is **universally observed that the primary article in the evaluation of professionals is the outcomes**. And, above all, **long-term outcomes**. Therefore, outcome measurement should be as objective as possible, although in education it is truly difficult to achieve due to the role that feelings and ideologies play in each of the instruments that can be used.

Below are several instruments proposed for measuring teacher performance in the classroom in addition to the SET. No evaluation instrument is free of biases. What is proposed here is a proposal that hopes that by using multiple instruments to evaluate the teacher, the set of instruments will decrease the uncertainty caused by the biases that each instrument has.

Other alternatives for teacher evaluation:

1. **On-site evaluation or through recordings carried out by entities responsible for pedagogy**. **Bias**: Pedagogy specialists typically are not experts in the topic of the class and therefore find it difficult to assess the effectiveness of the method used by the teacher.
2. **On-site evaluation or through recordings carried out by other teachers in the same area of specialization**. **Bias**: Each teacher has their teaching method and generally believes that their method is the best. A teacher may disagree with another's methods. To determine if one method is superior to another, a statistical test with strict control of variables must be conducted.
3. Since grades are not a solid instrument for evaluating student learning, **using standardized tests**. **Bias**: In the US, standardized tests have been used for 10 years at the high school level, and although the vast majority of teachers are evaluated very well, students have not learned more [22]. Why? Because teachers know the content of standardized tests and prioritize that content over the rest of the material.
4. **Evaluate multiple times in the academic period**. It has been observed that, although the SET is highly correlated over time, there are variations depending on when the evaluation is conducted. At 20% of the school period, the student already knows if the teacher presents information clearly and fluently. This estimation is free of teacher-student conflicts. At 80% of the school period, the student now has an estimate of what their final grade will be and may have had conflicts with their teacher. And at 100% or more, the student can map the efficiency of the teacher's method with their result. Three evaluations are proposed at the mentioned time periods.
5. **Reviewing the teacher's portfolio**. This is probably the method with the least bias. It involves reviewing notes, presentations, examples, exercises, exams, etc., generated by the teacher. **Bias**: It's necessary to see how teachers apply these resources.

6. **Student surveys**. The student surveys always provide valuable information that can show serious errors and give an idea of students' course satisfaction.

12   Conclusions

Assessing a teacher's performance in the classroom is complex, and there is no single solution. Students are very poor at evaluating their own learning, so teacher evaluation by students is far from being a reliable tool for assessing a teacher's performance in the classroom. This evaluation, although it is a numerical instrument that provides valuable information for improving teaching practice, is affected by multiple biases that are sometimes difficult to overcome. Furthermore, it tends to favor fluent classes that are not necessarily the most effective for teaching complex concepts. Evaluating teachers' performance in the classroom using only student evaluation can incentivize bad teaching practices, grade inflation, and a mistaken idea in students about their own learning. A method is proposed to improve teacher evaluation for use in the current environment, and a multidimensional and comprehensive model of teacher evaluation is proposed.

References

[1]   D. E. Clayson, "Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature," *Journal of marketing education,* vol. 31, no. 1, pp. 16-30, 2009.

[2]   B. Uttl, C. White and D. W. Gonzalez, "Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related," *Studies in Educational Evaluation,* no. 54, pp. 22-42, 2017.

[3]   Z. Papamitsiou and A. Economides, "STUDENTS'PERCEPTION OF PERFORMANCE VS. ACTUAL PERFORMANCE DURING COMPUTER-BASED TESTING: A TEMPORAL APPROACH," in *INTED 2014 Proceedings*, Valencia (Spain), 2014.

[4]   L. Deslauriers, L. McCarty, K. Miller and G. Kestin, "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom," *Proceedings of the National Academy of Sciences,* vol. 39, no. 116, pp. 19251-19257, 2019.

[5]   J. M. Clauss and C. K. Geedey, "Knowledge surveys: Students ability to self-assess," *Journal of the Scholarship of Teaching and Learning,* vol. 10, no. 2, pp. 14-24, 2010.

[6]   W. Stroebe, "Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis," *Basic and Applied Social Psychology,* vol. 42, no. 4, pp. 276-294, 2020.

[7]   S. K. Carpenter, M. M. Wilford, N. Kornell and K. M. Mullaney, "Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning," *Psychonomic bulletin & review,* no. 20, pp. 1350-1356, 2013.

[8] S. K. Carpenter, A. E. Witherby and S. K. Tauber, "On students'(mis) judgments of learning and teaching effectiveness," *Journal of Applied research in Memory and cognition,* vol. 2, no. 9, pp. 137-15, 2020.

[9] A. M. Ewing, "Estimating the impact of relative expected grade on student evaluations of teachers," *Economics of Education Review,* vol. 31, no. 1, pp. 141-154, 2012.

[10] J. E. Zull, Zull, J. E. (2020). The art of changing the brain: Enriching the practice of teaching by exploring the biology of learning, Stylus Publishing, LLC, 2020.

[11] S. K. Carpenter, A. E. Witherby and S. K. Tauber, "On students'(mis) judgments of learning and teaching effectiveness: Where we stand and how to move forward," *Journal of Applied Research in Memory and Cognition,* vol. 9, no. 2, pp. 181-185, 2020.

[12] T. K. Lindsay, Combating the "other" inflation: Arresting the cancer of college grade inflation, Texas Public Policy Foundation, 2014.

[13] C. F. Eiszler, "College students' evaluations of teaching and grade inflation," *Research in Higher Education,* vol. 43, pp. 483-501, 2002.

[14] K. F. Butcher, P. J. McEwan and A. Weerapana, "The effects of an anti-grade inflation policy at Wellesley College," *Journal of Economic Perspectives,* vol. 28, no. 3, pp. 189-204, 2014.

[15] P. Babcock, "Real costs of nominal grade inflation? New evidence from student course evaluations," *Economic inquiry,* vol. 48, no. 4, pp. 983-996, 2010.

[16] T. Bar, V. Kadiyali and A. Zussman, "Grade information and grade inflation: The Cornell experiment," *Journal of Economic Perspectives,* vol. 23, no. 3, pp. 93-108, 2009.

[17] B. Park and J. Cho, "How does grade inflation affect student evaluation of teaching?," *Assessment & Evaluation in Higher Education,* vol. 48, no. 5, pp. 723-735, 2023.

[18] "Social Media's Concerning Effect on Teen Mental Health," The Annie E. Casey Foundation, 10 08 2023. [Online]. Available: https://www.aecf.org/blog/social-medias-concerning-effect-on-teen-mental-health. [Accessed 19 01 2024].

[19] C. Ding, W. Jabr and H. Guo, "Ding, C., Jabr, W., & Guo, H. (2023). ELECTORAL COMPETITION IN THE AGE OF SOCIAL MEDIA: THE ROLE OF SOCIAL MEDIA INFLUENCERS," *MIS Quarterly,* vol. 47, no. 4, 2023.

[20] W. Lee, S. G. Yang and B. J. Kim, "The effect of media on opinion formation," *Physica A: Statistical Mechanics and its Applications,* vol. 595, p. 127075, 2022.

[21] F. I. Swalé, "Inception: Social Media's Influence On Your Opinion," Forbes, 27 02 2023. [Online]. Available: https://www.forbes.com/sites/forbeseq/2023/02/27/inception-social-medias-influence-on-your-opinion/?sh=77aa7db9cf13. [Accessed 19 01 2024].