Check for updates

Measuring Teaching Effectiveness: A comprehensive analysis of assessment tools, criteria, and standardization

Nice John [1] * , Dr. Gopal Krishna Bhardwaj [2]

1. Research Scholar, Department of Education, University of Technology, Jaipur, Rajasthan, India
nicejohn508@gmail.com ,

2. Professor, Department of Education, University of Technology, Jaipur, Rajasthan, India

Abstract: Students' social and attitude development is said to be influenced by effective instructors just as much as their academic performance. "What makes a good teacher?" is an old and well-studied topic. Teacher effectiveness (TE) has been measured using a variety of approaches, each with its own set of advantages and disadvantages. In order to measure how successful school instructors are, many turn to the Teacher Effectiveness Scale (TES-KU) created by Kulsum (2011). The majority of research participants viewed the administration of this 60-item measure as tedious and time-consuming. Therefore, a condensed version of this scale was attempted. The Principle Component technique was used to factor analyze responses from 200 school instructors. It would be ideal if schools and teachers could agree on a common set of criteria to evaluate classroom performance. Also, it would be great if these measures were congruent with what we know about how to evaluate teachers' efficacy in the classroom. Sadly, most institutions are currently not operating at this optimum level. The purpose of this interview research was to gather information from 72 physics teachers on the methods used by their schools to determine the efficacy of their lessons.

Keywords: Assesment, Teacher Effectiveness, TES Scale, Tools

- X -

INTRODUCTION

Structured classroom observations, teachers' roles in students' increased success, and students' views of their teachers' efficacy and the classroom environment are the three most popular evaluation tools in the US. National Council on Teacher Quality data shows that as of 2019, evaluation methods for teachers include classroom observations in 44 states, assessments of student performance progress in 33, and surveys of students in 7 states. The use of student surveys in the assessment of teachers is legal in 24 more states. Teachers' involvement in professional development, mentoring, or committees, as well as instructional artifacts like lesson plans and assignments, as well as teacher self-reporting like instructional logs, and feedback from parents, peers, or administrators, can all be used to infer teaching effectiveness.

Nonetheless, academics have not looked at these metrics as much as they have test-based and observational approaches. For that reason, they are not often used in systems that assess educators. There is no universally accepted definition of successful teaching, which is one of the key obstacles to quantifying teaching effectiveness. The expectations, criteria, and standards for the quality of instruction might vary between fields, organizations, and stakeholders. Not only that, but the traits and requirements of each student cohort, in addition to the course's setting, material, and goals, may affect how successful a teacher is. Recognizing the range and complexity of teaching circumstances and having a clear and common

knowledge of the aims and consequences of teaching are thus necessary for assessing teaching effectiveness. A lot of people have started to look at SETs with suspicion as a way to gauge how successful teachers are. Not only are there valid issues about gender bias in the findings, but SETs are notoriously hard to understand and rely on student opinions that don't always correspond to how well a teacher is doing their job. Furthermore, it is evident that SETs are just one of many data types needed for a thorough evaluation of teaching. When used in conjunction with other measures of teaching effectiveness, they are most useful for summative and formative evaluations, as well as for professional and personal development decisions.

LITERATURE REVIEW

Shawn R. Simonson et.al (2021) Institutional systems for evaluating teaching are often insufficient, erroneous, and fail to directly enhance teaching or encourage teachers to develop their craft. This is usually due to the fact that successful instruction is hard to measure, and the majority of evaluation instruments either fail to do so or do not have clear criteria to work by. This could make teachers resistant to change or make them oblivious to the fact that they need to alter their pedagogical approach. We create a tool that takes into account all aspects of teaching and can adapt to diverse techniques, modes, and contexts, and we lay out a framework for what constitutes good teaching.

Charles R Henderson et.al (2014) It would be ideal if schools and teachers could agree on a common set of criteria to evaluate classroom performance. Also, it would be great if these measures were congruent with what we know about how to evaluate teachers' efficacy in the classroom. Sadly, most institutions are currently not operating at this optimum level. The purpose of this interview research was to gather information from 72 physics teachers on the methods used by their schools to determine the efficacy of their lessons. The results indicate that when evaluating the performance of a teacher, most or all of the weight is given to student feedback. Conversely, teachers rely heavily or exclusively on students' test scores and informal formative evaluations to determine how well they are meeting learning outcomes. The academic literature suggests assessment procedures, but few institutions and teachers actually use them. Teachers tend to have much more favorable things to say about the ways they personally assess their own teaching compared to the ways their schools do the same. By relying more on systematic formative assessment and using standardized measures based on student learning, both institutions and instructors might benefit from expanding the sources of evaluation they utilize to evaluate the success of instruction.

Ezequiel Molina et.al (2020) The absence of freely accessible classroom observation tools that are practical to implement, validated in their context, and can be integrated into national monitoring systems is a common barrier to assessing the quality of teaching practices among primary school teachers in low- and middle-income nations. In response to this disparity, researchers in low- and middle-income nations created Teach, an open-access classroom observation tool, to evaluate the effectiveness of elementary school teachers' pedagogical strategies. This article assesses Teach's validity using data collected from Punjab, Pakistan. The results demonstrate strong inter-rater reliability, internal consistency, and enough information to distinguish between low- and high-quality teaching methods in the Teach scores. Moreover, student outcomes improved in correlation with better Teach scores.

John Michael Del et.al (2022) In order to achieve SDG4, which is providing excellent education, teachers

play a crucial role. Their impact on pupils is substantial depending on the quality of instruction they provide. This PAR asserts that there are deficiencies in the current state of teaching effectiveness and proposes three solutions: (a) create a theoretical framework based on shared assumptions that guided the research; (b) find a way to measure this framework to make sure students are getting a good education and learning the skills the country needs; and (c) create an evaluation tool. Drawing on the findings of bottom-up evaluation focus groups with key informants in selected mid-level administrators (coordinators), instructors, students, and alums from selected sectarian universities, this study developed a theoretical framework on the efficacy of instruction. In order to evaluate the efficacy of educators, the researchers created indicators and selected test questions for a standardized evaluation method. Key participants in a learning organization include the pedagogy, content, and knowledge of the teacher; the academic performance of the learners; and the adaptation of institutions in outcomes-based education, according to the theory of successful teaching. Based on the results, effective teachers demonstrate pedagogy, content, and knowledge; they have good qualities; they take a humanistic and professional approach to their work; they incorporate real-world examples into their lessons; they use a variety of technology and online tools to help students learn; and they base their lessons on the school's VMGO, program, course outcomes, and activities. This aids in the education of the students. This study's findings suggest that future studies should make advantage of the PAR's indicators, particularly when developing an evaluation instrument to gauge the efficacy of classroom instruction.

Madhu Gupta et.al (2022) A standardized observational scale to evaluate educators' pedagogical efficacy is the focus of this article. Planning, creating first and second drafts of items, writing and analyzing the items, finishing the items, scoring, reliability, validity, and establishing of norms were all part of the process to design and standardize the scale. Lesson preparation, lesson delivery, classroom management, teacher professional and personal competency, and lesson closure were the five components of the first scale formulation. Each dimension included 90 items. For this scale, we enlisted the help of twenty specialists in the domains of language, sociology, psychology, and education. In order to evaluate the items' likelihood, the investigators also monitored 15 educators. The second draft kept 75 elements based on the unanimous consensus. One hundred secondary and senior high school educators in Haryana were selected at random to review the final version. Using t-test calculation, the entries were finally selected. At the 0.05 or 0.01 threshold of significance, only the items that were determined to be significant were kept. As a result, 56 things (important items) were kept for the final draft, whereas 19 items were eliminated. The internal consistency technique, which varies from 0.195 to 0.555, test-retest reliability (0.727), and split-half reliability (0.970) were used to establish the dependability of the scale. There is strong validity as the coefficient of correlation between the measures of teaching effectiveness falls within the range of 0.353 to 0.688. z-Score criteria have been developed to assess the efficacy of instruction.

METHODOLOGY

There has to be an easy way to measure TE in schools for researchers. An improved substitute for observation and principal's evaluations is standardized scales. Creating a condensed version of TES was the primary goal of this research. According to Kulsum (2011), the original scale was characterized as a self-anchored striving measure. The five domains covered by the sixty-item scale are as follows: (PTP) Lesson planning and preparation; (CRM) Classroom management; (KSM) Subject matter knowledge and

its presentation; (TC) Teacher characteristics; and (IPR) Interpersonal relations. There is good validity, a test-retest reliability of 0.63, and a split-half reliability of 0.68 for the scale. This scale was used as part of a broader research project that was carried out in the rural and urban educational districts of Bangalore. The data presented in this research only relate to the TES's alteration.

Fifty schools in Bangalore's rural areas and fifty schools in the city's metropolitan areas made up the sample. In the fall of 2008, a group of randomly chosen American physics teachers were asked to fill out an online survey regarding their pedagogical aims and methods, as well as their familiarity with and application of a collection of twenty-four instructional strategies in the context of teaching quantitative physics to beginners. Teachers from both four- and two-year colleges and universities participated in the survey. The survey research had an overall response rate of 50.3%. In order to conduct an accompanying interview research, a subset of survey participants was intentionally selected. The interviewees were chosen to include current and past users as well as informed bystanders of two research-based pedagogical approaches: Workshop Physics and Peer Instruction. We also made sure that the interviewees were a mix of male and female professors from both two- and four-year institutions. Only 72 (or 72% of the 100 teachers we contacted) were willing to take part in the interviews. There was around a 50/50 split between instructors who declined to participate and those who did not reply to our requests for interviews. The interviews were place over the phone and typically lasted about one hour. As an incentive for taking part in the interview, each interviewee received $75.

Basic details of the sample are given in Table 1

Table 1: Sample Characteristics

Gender	Rural	Urban	Total
Male	35	17	52
Female	65	83	148
Total	100	100	200

The data clearly shows that the sample consisted of more female instructors than male ones. As a whole, the subjects were mostly married. The average age of the sample was around 20 years, while the mean age of the participants was roughly.

Analysis: In this work, the TES was reduced in length using factor analysis. We used factor analysis to rate the 200 instructors' TES responses. One way to break down big datasets into their component variables is by factor analysis. It is referred to be "a queen of analytic methods" (p.659) by Kerlinger (1973). As it simplifies a lot of different measurements, it helps the cause of scientific parsimony, in his opinion. What this shows us is which metrics are related, or at least measure the same thing to a certain

extent. Reducing the length of surveys by removing unnecessary items might be rationalized using factor analysis. Most of the time, principal components are used as the default approach for extraction. It finds linear combinations of variables that aren't linked and offers the first component the most explained variance. The following variables are uncorrelated with one another and explain progressively less of the variation. This approach is thought to be suitable for situations when data reduction is the objective.

RESULT

The original TES was re-tested with 200 educators, and the 25 items with the greatest loading on factor 1 were kept after factor analysis. Information is provided in the tables below.

Table 2: Summary of Factor Analysis

Factor	1	2	3	4	5
SS Loadings	19.645	11.774	5.785	4.915	1.692
Proportion Var	0.327	0.196	0.096	0.082	0.028
Cumulative Var	0.327	0.524	0.62	0.702	0.73

The item distribution on the five subscales was compared between the original and condensed scales in Table 4. Subscale codes include Class Room Management (CRM), Knowledge of Subject Matter (KSM), Teacher Characteristics (TC), and Interpersonal Relations (IPR).

Table 3: Showing the Number of Items in the Original and Short Scale

Subscales	PTP	CRM	KSM	TC	IPR	Total
Original Scale	11	14	7	17	11	60
Short Scale	5	7	2	8	3	25

The original scale ranked TC highest, CRM second, and KSM last in terms of item count. The tendency is the same even on the smaller scale. In contrast to PTP, things on IPR are much smaller. It was believed that using these items, we could quickly and easily gather data on teachers' performance across all five dimensions of the scale.

Scoring and Interpreting the TES-S

The TES-S asks participants to read each statement and rate how effective they are right now from zero to ten. Unless otherwise specified, all explanations and instructions will be TES-style. There is a possible score range of 0–250. The 'future' rating was not included into the score in the original scale either. This is why it is not included. To determine if a group of individuals is highly or poorly effective on the scale, it is best to take the mean score of the group plus or minus one standard deviation. The distribution's Median score is another option. The following would be the interpretation for individual administration:

Description	Score on TES	Score on TES-S
Average Teacher	> 320	> 133
Most Effective Teacher	>435	>181
Most Ineffective Teacher	< 252	< 105

If your TES-S score is less than 133, it means that your efficacy is below average. Two or three entries from the same category could appear consecutively in Table 3. To make sure that references to the five sections are spread out more equitably, things were moved about. Appendix A provides the revised and final version for the benefit of users in the future. The next stage was to provide this condensed version to a subset of the initial sample and compare their results. Unfortunately, none of the participants were accessible because of the COVID-19 pandemic, therefore this could not be done. This sample does not lend itself to online testing or phone-in responses. We may consider this for a future project when the timing is right.

Overall attitudes towards assessment

First, we take a look at how respondents overall felt about the effectiveness of teacher and school evaluation strategies (Fig. 1). When asked about their experiences evaluating their own teaching, the majority of the 72 educators who participated in the survey reported feeling good about it (72%). Unfortunately, only 15% of teachers reported feeling good about how their school evaluated their performance in the classroom. A large percentage of teachers (32% to be exact) are reportedly unhappy with the way their school evaluates their performance in the classroom. Not only that, but we couldn't generalize about the teachers' attitudes towards institutional evaluation techniques for 53% of them.
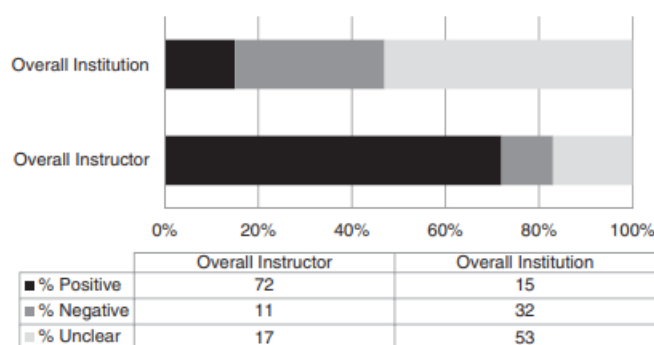
	Overall Instructor	Overall Institution
■ % Positive	72	15
■ % Negative	11	32
□ % Unclear	17	53

Figure 1. Interviewees' overall attitudes of how well they and their institutions are able to assess teaching effectiveness.

According to Figure 1, teachers are more sure of their own evaluation methods than they are of their school's. Most of the time, there wasn't a clear and concise quote to back up the conclusions drawn regarding the respondents' sentiments since such conclusions were based on data collected during the whole interview. The examples below show how some teachers made their opinions known in relatively brief quotations. In terms of general attitudes, the two most prevalent ones were faculty members' favorable evaluations of their own teaching success and students' negative evaluations of their school's teaching efficacy.

Use of assessment strategies

Instructors' and institutions' reported usage of various evaluation procedures is summarized in Figure 2, which also includes a numerical count of these strategies. In this study, "use" was defined as using in order to evaluate instruction. Some teachers may have alluded to something being part of their class but failed to specify how it would be evaluated. So, in this case, we would code the student evaluations as being used by the institution but not by the instructor. This is because, for instance, if a teacher said that collecting evaluations at the end of the semester was part of their institution's official evaluation system, but they didn't say that they were using the evaluations to inform their own teaching.
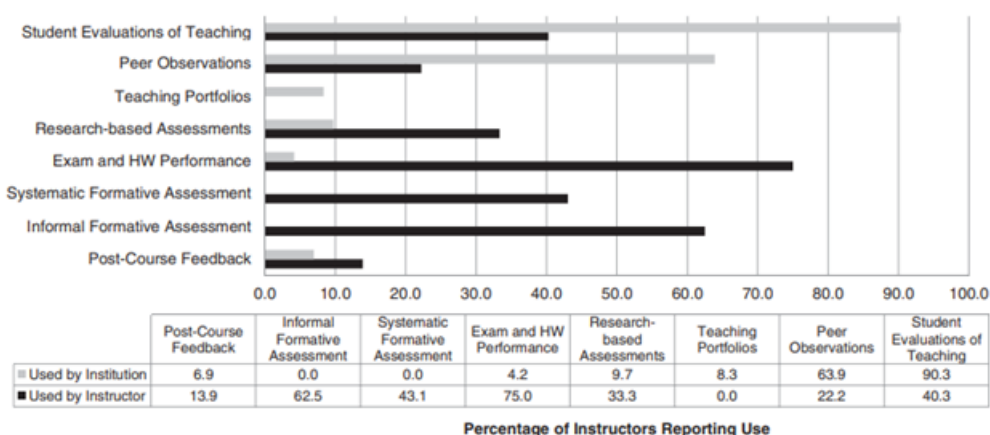

	Post-Course Feedback	Informal Formative Assessment	Systematic Formative Assessment	Exam and HW Performance	Research-based Assessments	Teaching Portfolios	Peer Observations	Student Evaluations of Teaching
▨ Used by Institution	6.9	0.0	0.0	4.2	9.7	8.3	63.9	90.3
■ Used by Instructor	13.9	62.5	43.1	75.0	33.3	0.0	22.2	40.3

Percentage of Instructors Reporting Use

Figure 2. Reported use of various sources of assessment information by instructor and by institutions in judging teaching effectiveness

Instructors and institutions supposedly employ different sources for evaluation information, as seen in the image. When evaluating a teacher's performance, most institutions rely on student assessments (90%) and peer observations (64%), according to instructors. Official formative assessment (such as tests and quizzes) accounts for 75% of teachers' evaluations, while informal forms account for 63%. In order to determine if there were any variations in the evaluation procedures used, we looked at instructor and institutional factors. The gender of the teacher and the kind of school were the recognized characteristics. These included two-year colleges, four-year colleges with a physics B.A. as the highest physics degree, and four-year colleges with a physics graduate program. We conducted four tests to rule out possible outcomes based on these two features: 1) The types of information used to evaluate teaching effectiveness do not differ between male and female instructors. 2) The types of information used to evaluate teaching effectiveness by instructors' institutions do not differ between male and female instructors. 3) The types of information used to evaluate teaching effectiveness by instructors at the three different types of institutions do not differ. 4) The types of information used to evaluate teaching effectiveness by instructors at the three different types of institutions do not differ from one another.

CONCLUSION

A condensed version of the 60-item TES was created due to the discovered time-consuming nature of the original. Two hundred elementary school teachers who took the first TES had their data factored using the principal component analysis. If we are serious about moving away from traditional classroom methods and toward ones grounded on research, this is a crucial discovery. Without measuring and using intended results in evaluating instruction, it is very unlikely that teaching methods would improve.

References

1. Simonson, S. R., Earl, B., & Frary, M. (2021). Establishing a Framework for Assessing Teaching Effectiveness. *College Teaching*, *70*(2), 164–180. https://doi.org/10.1080/87567555.2021.1909528

2. Henderson, Charles & Turpen, Chandra & Dancy, Melissa & Chapman, Tricia. (2014). Assessment of teaching effectiveness. Physical Review Special Topics - Physics Education Research. 10. 10.1103/PhysRevSTPER.10.010106.

3. Ezequiel Molina et.al "Measuring the quality of teaching practices in primary schools: Assessing the validity of the *Teach* observation tool in Punjab, Pakistan" Teaching and Teacher Education Volume 96, November 2020, 103171

4. Del, John & Aquino, Rosario & Chavez, Maria Asela. (2022). Teaching effectiveness: Basis for the development of assessment tool. Journal of Research, Policy & Practice of Teachers & Teacher Education. 12. 10.37134/jrpptte.vol12.2.4.2022.

5. Gupta, Madhu & Verma, Gunjan. (2022). Construction and Standardization of Teaching Effectiveness Observation Scale (TEOS). The International Journal of Indian Psychology. 10. 586-595. 10.25215/1001.058.

6. Teddlie, C., & Reynolds, D. (2000). The international handbook of school effectiveness research.

London, UK: Falmer Press.

7. Torrance, H., & Pryor, J. (1998). Investigating formative assessment: teaching, learning and assessment in the classroom. Buckingham, UK: Open University Press.

8. Visscher, A. J., & Coe, R. (2002). School improvement through performance feedback Rotterdam: Swets & Zeitlinger.

9. Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: what will it take to make it work? In C.A. Dwyer (Ed.), The future of assessment: shaping teaching and learning (pp. 53-82). Mahwah, NJ: Lawrence Erlbaum Associates.

10. Wiliam, D., Lee, C., Harrison, C., & Black, P.J. (2004). Teachers developing assessment for learning: Impact on student achievement. Assessment in Education: Principles Policy and Practice, 11(1), 49– 65.