

---

**RESEARCH ARTICLE**

**Teacher Performance Assessment: Models and Practices- A Narrative Review**

**Basim Nadhreen**

Assistant Professor, English Language Institute, Umm Al-Qura University, Makkah, Saudi Arabia

**Corresponding Author:** Basim Nadhreen, **E-mail:** [bmanadhreen@uqu.edu.sa](mailto:bmanadhreen@uqu.edu.sa)

---

**ABSTRACT**

Teacher performance assessment (TPA) in Saudi Arabian higher education has evolved from a narrow, compliance-focused review to a multidimensional system that includes evidence of teaching quality, scholarly productivity, and service contributions. In English language teaching (ELT) units within universities such as English Language Institutes (ELIs), TPA is further complicated by the need to evaluate classroom interactional competence, assessment literacy, curriculum alignment, and student learning evidence alongside research and service expectations. This narrative review synthesizes major models and practices of TPA published between 2000 and 2025, with a focus on portfolios, student feedback, and peer observation, and considers additional evidence sources, including research metrics, supervision and mentoring, professional development, and community engagement. The review proposes a framework based on *Umm Al-Qura University's guide for Faculty Member Performance Evaluation and Development instrument*, which allocates 40 points to teaching performance, 45 to scholarly output, and 15 to community service. This framework offers indicators tailored for ELT contexts and combines formative development with summative evaluations. Recommendations for implementation at the ELI across various faculty ranks, including instructors, lecturers, teaching assistants, and faculty members, are also provided, covering training, policy adjustments, and administrative procedures.

**KEYWORDS**

Teacher performance assessment; faculty evaluation; teaching portfolios; student evaluations; ELT; English Language Institutes; Saudi Arabia; Umm Al-Qura University; narrative review.

**ARTICLE INFORMATION**

**ACCEPTED:** 01 February 2026

**PUBLISHED:** 19 March 2026

**DOI:** 10.32996/ijllt.2026.9.3.15

---

**1. Introduction**

Teacher performance assessment (TPA) is a significant practice at universities because it influences several functions, such as contract renewal, promotion, professional development, and institutional quality assurance. This (TPA) has shifted, obviously, from a single-assessment dominated by student evaluations to multi-source frameworks aimed at increasing accuracy and developmental value (Seldin, 2006; Berk, 2005; Spooen et al., 2013). This transformation prompts many institutions to take practical steps to reduce excessive reliance on student input and imprecise performance criteria (Benton & Cashin, 2014; Marsh, 2007; Ory & Ryan, 2001).

In Saudi Arabia, the transformation of university accreditation has promoted interest in evidence-based evaluation systems. Thus, in English Language Institutes (ELIs), evaluation is expected to focus on academic issues confined to language teaching, such as instructional interaction, communicative task design, learner engagement, assessment for learning, and alignment with course specifications. As mentioned earlier, at Umm Al-Qura University (UQU) in Makkah, the Faculty Member Performance Evaluation and Development (hereafter, the UQU Guide; Umm Al-Qura University, 2022) defines a division of marks into Teaching Performance (40 points), Scientific Productivity (45 points), and Community Service (15 points).

This paper presents a narrative review of TPA models with a special focus on Applied Linguistics and ELT instructors at UQU, while remain sensitive to institutional-level comparability in terms of ranks (teaching assistants, lecturers and faculty). The goal of the paper is not only to summarize research, but also to suggest an integrated framework for operationalizing the UQU Guide in a more effective manner, and to strengthen the evidence base in support of fair and development-oriented evaluation.

**Review Question:**

How can the current UQU Guide be enhanced through an integrated framework suitable for the ELI context?

**2. Literature Review**

According to Nunan (1988), the terms assessment and evaluation are often used synonymously in the literature. For example, Geddes and Marks (1997, 206) state program administrators should "assess objectively the job performance of each employee... and must establish criteria and a formal plan for supervising and evaluating employee performance". On the one hand, they recommend assessing job performance; on the other, they see the issue as employee performance being evaluated, with no distinction in meaning intended. Palomba and Banta (1999, 7) write, "These categories are often referred to as formative and summative evaluations...Formative assessment is... Summative assessment is..." The term evaluation is replaced in the following sentence by assessment, again with no apparent change in meaning. A fundamental question in TPA regards whether evaluation is basically for accountability (summative decisions) or for development (formative improvement). Many researchers recommend explicitly dividing (or at least differentiating) these objectives: using low-stakes feedback to improve, and merging different graded measures for high-stakes outcomes (Centra, 1993; Berk, 2005; Seldin, 2006).

Nunan (1999:85) states that: evaluation is the collection and interpretation of information... for decision making purposes. Assessment refers to the tools, techniques, and procedures for collecting and interpreting information about what learners can and cannot do". In other words, evaluation is used to make decisions, and assessment consists of the various ways which information can be collected on a language learner's achievement in order to improve instruction (Fradd, et al., 1994; Brindley, 2001). In contrast, Hecht et al., (1999) seem to give the two terms the opposite relationship when they define summative and formative evaluations under the heading Faculty Assessment. This view of the relationship between the two terms is exemplified further when Hecht et al. (1999, 241) state that "the most common instrument for assessing (bold added) faculty teaching is the student course evaluation form." To them, an evaluation tool is used as part of a larger assessment of teacher performance, which appears to be in direct contrast to the view held by Nunan.

Palomba and Banta (1999) do, however, distinguish between the two terms in another section of their book by using assessment to refer to student learning and evaluation to refer to faculty members. To them, student learning is assessed, whereas faculty members are evaluated. Making this distinction is important in order to lessen the fear many faculty members have when their performance is evaluated specially if it has the potential to be used against them in the faculty review process. Genesee (2001, 144) relates assessment to learners and evaluation to programs, defining evaluation in TESOL settings as "a process of collecting, analyzing, and interpreting information about teaching and learning in order to make informed decisions that enhance student achievement and the success of educational programs".

The key to the above discussion is that evaluation can be taken to refer to collecting information for the purpose of making decisions, especially as they relate to faculty retention, promotion, and salary (Darling-Hammond, 2013).

One of the main reasons for undertaking such a task is to manage the quality of teaching and learning in a particular language program. By actually observing faculty teaching, meeting with them to discuss their performance and progress in the program, eliciting and analyzing student feedback on the teaching and learning process, and so on, program administrators gain insights into the effectiveness of instruction in the program AlmuBarak et al. (2025). Make no mistake about it; evaluating faculty is an information gathering activity, and this information can later be used to make the program stronger by preparing and enabling instructors to better meet the language learning needs of the students.

Another reason for evaluating faculty is to recognize and reward teaching excellence in the hope that it will enhance both individual performance and team morale (Song Jiao, et al., 2026). Recognizing outstanding performance can be achieved by presenting certificates and awards to deserving faculty members, while rewarding them for their achievements can involve providing pay increases based on merit (Arreola, 2007).

An additional reason to evaluate faculty is to provide guidance to teachers in need of professional development. At face value, the opportunity for supervisors and teachers to discuss and explore together problematic issues would seem to be one of the greatest

benefits of the faculty review process in that it provides opportunities for constructive feedback (Richards & Farrell, 2005). Unfortunately, some faculty may interpret such feedback as demeaning, and they may resent being evaluated in the first place, even if the aim is to improve their classroom performance (Fanselow, 1990).

The evaluation system can be used for contract renewal; faculty members being evaluated on their performance in three categories: (i) teaching (which is the most important area for the English language teaching faculty) (ii) professional development (which included participation at conferences, workshops, and seminars); and (iii) service to the unit college, university, and the wider professional community. These categories are typically cited in the literature as ones to focus on in the evaluation of language program faculty (Gedded & Marks, 1997; Hecht, et al., 1999).

Lastly, faculty evaluation may be a requirement for accreditation purposes, which is not a trivial matter for the host institution. In this case, the impetus for the evaluation would be beyond the language program's control and would come from higher levels within a university's or institution's administrative hierarchy. The evaluation itself would provide documentation of faculty work that could be reviewed by accrediting agencies as evidence of faculty evaluation itself would provide documentation of faculty work that could be reviewed by accrediting agencies as evidence of faculty output and its relevance to the mission of the particular university or institution under review (Penny, 2003; Stake, 2004).

### **3. Methodology**

This study adopted a narrative review design to bring together evidence from multiple sources across different educational domains, including teaching evaluation, ELT teacher development, program evaluation, and quality assurance. A narrative approach was selected because the research literature uses different methods and involves multiple units of analysis. This strategy helped a comprehensive overview directed towards informing an evaluation framework aligned with the UQU Guide.

#### **3.1 Search strategy and sources**

The search was done using a concept-based approach. Fundamental databases integrated, namely, Scopus, Web of Science, ERIC and Google Scholar. To take advantage of regional importance, more searches were applied in accessible Saudi digital archives and local university databases.

Search terminologies were combined and used on different platforms targeting faculty and teaching evaluation (e.g. teaching effectiveness, faculty evaluation, student evaluations of teaching, teaching portfolio, teaching assessment), quality assurance (e.g. promotion criteria, quality assurance) and other indicators of performance (e.g. research productivity measures, community service evaluation). Regional contexts including Saudi Arabia and Gulf higher education were used to identify context-specific resources. These themes were then used to determine practical implications for the development of an integrated evaluation approach that is aligned with the UQU Guide.

#### **3.2 Inclusion and exclusion criteria**

The review included peer-reviewed journal articles, academic publications and significant research papers on teacher evaluation in higher education, particularly publications on multi-evidence assessment. Since the study is based on an English Language Institute setting, ELT-related literature was included where it provided relevant methodologies or ideas (e.g. teaching portfolios, teacher competencies), and when regional policy or quality assurance documents were available.

#### **3.3 Key evidence domains reviewed**

The analysis focused on criteria that are commonly used in faculty evaluations. Teacher portfolios have been identified as an essential resource because they could record instructional setting, assessment practice, reflective improvement and consistency with learning objectives, in particular when guided by clear criteria and standardized assessment (Wolf, 1991; Hutchings, 1998).

Student feedback was reviewed as one of the extensively used tools for providing scalable insights into student experiences with the recognition of its known limitations and the importance of careful interpretation and contextual reporting in ELT settings. The review also examined course files and course reports as quality assurance artifacts, supervision and mentoring as both under-defined but important contributions and research productivity and bibliometrics as common indicators which require role-sensitive interpretation. Finally, supporting domains of professional development and university/community service was included, with a focus on the specification of evidence standards for improved fairness and transparency.

#### 4. Discussion

This narrative review highlights five points that are directly relevant to the UQU Guide and evaluation practices in the ELI context. First, teaching portfolios offer the most complete picture of teaching quality when they are evaluated using clear rubrics and when portfolio evidence is linked to continuous improvement, rather than documentation (Seldin, 2006). Second, student feedback is still useful as it is a reflection of student experiences, but this feedback should be interpreted in the context of the course and used together with other evidence to mitigate factors that can skew ratings (Yaho & Grady, 2005; Marsh, 2007; Spooen et al., 2013). Third, assessment of research should be focused on accurate expectations of function and the identification of ELT-relevant outcomes, such as scholarship of teaching and learning, and practical findings if possible, rather than giving complete weight to exclusive publishing indicators. Fourth, service should be assessed according to the quality of contribution and the results obtained, instead of merely through participation or membership.

For UQU, the guide form could potentially be improved without changing its current scales by standardizing what qualifies as evidence and how it is assessed within the 40-point teaching aspect- in order to make sure that outcomes are more comparable, fair, and helpful for decision-making and progression.

#### 5. Results

Across the literature, the main conclusion is clear: teaching performance cannot be accurately judged with a single measure. Effective evaluation systems first define what “good teaching” means in a specific context, then gather evidence from multiple sources, and use clear rubrics with trained evaluators. The literature also suggests separating formative feedback from high-stakes scoring whenever possible, and promoting transparency through clear procedures, equity safeguards, and appeal options (Berk, 2005; Seldin, 2006; Marsh, 2007; Benton & Cashin, 2014). At UQU’s ELI, this involves creating consistency by connecting the most meaningful evidence—especially portfolios, student feedback, and peer observations—to the UQU Guide, and clearly explaining how each indicator will be documented, scored, and used for growth.

Based on these findings, this paper proposes an integrated framework that maintains the same UQU Guide structure (40/45/15) but enhances it through clearer indicator definitions, standardized evidence requirements, shared scoring rubrics, and straightforward triangulation rules. Within the 40-point teaching component, the most effective model for an ELI relies on three interconnected sources: a rubric-based teaching portfolio as the primary evidence, student feedback that includes key context (such as course level and class size) and serves as supporting—not decisive—evidence, and peer observation conducted with a common ELI rubric involving at least two visits (one developmental and one evaluative). These sources can be linked directly to the existing UQU Guide teaching criteria by using portfolio alignment and course reports to demonstrate coverage of course specifications, observation and portfolio evidence to support “teaching efficiency,” and logs or access records to verify office hours.

The outcome indicators for the 45-point scientific productivity criterion highlight the importance of job clarity in ensuring fairness within an ELI. Research-active faculty can be assessed using traditional markers such as indexed publications, funding, and innovation outputs, while teaching-focused staff should be permitted to provide specified ELT-relevant outputs (e.g., materials development, field studies, and peer-reviewed conference contributions) with clear quality and publication guidelines. Lastly, the literature advocates for an impact-oriented approach to the 15-point community service indicator, where committees and meetings are recognized for their documented accomplishments and outputs, and university/community activities are evaluated using concise impact narratives supported by evidence (such as deliverables, letters, attendance records, and feedback). Overall, the UQU framework improves assessment evidence to be more consistent, verifiable, and beneficial for professional development, while maintaining its official classifications and ratings.

#### 6. Implementation Plan for the ELI at UQU

The ELI at UQU demands the implementation of a transparent, consistent, and realistic timeframe, as well as effective ethical security measures for the assessment process. The ELI should form an Evaluation Committee whose members are trained to apply standard ratings. To ensure that judgements reflect the adequate teaching setting, the committee should include representatives of all academic ranks and program levels. Additionally, it should monitor clear challenges and review procedures for better consistency and confidence in conclusions. In order to improve consistency, the ELI should conduct regular calibration and training activities. This includes yearly calibration sessions for peer observers and portfolio reviewers in order to ensure that scoring standards are the same between evaluators and over time (Weller, 2009; Steinberg & Garrett, 2016). Short and practical training is also necessary in interpreting student feedback, with attention to the most common sources of bias and misinterpretation (Benton & Cashin, 2014).

The evaluation system should work through a predictable data cycle. A semester-based formative procedure can be implemented for instruction and feedback that permits teachers to act upon results while still teaching. A distinctive yearly (or biannual)

summative cycle might thus be used for more significant decisions that are based on numerous sources of data rather than just one. Finally, the strategy should be directed by ethical and practical factors, such as preserving student feedback and observation records confidential, not using single observations as comprehensive documentation, and incorporating staff evaluation time (reviewing portfolios, observing classes, and committee work) into workload distribution.

### **7. Limitations**

As a narrative review, this paper does not claim exhaustive coverage nor a formal quality appraisal of all included studies. Additionally, local Saudi-specific empirical studies on ELI evaluation may be underrepresented in international databases and may require targeted institutional access. Some references are placeholders pending final source verification, as permitted by the project requirements.

### **8. Conclusion**

Teacher performance assessment (TPA) in higher education is most conclusive and useful when it is multi-source, explicit, and adjusted with both institutional goals and achievable objectivity. Faculty performance evaluation can lead to improved instruction, promoted student learning, and a more successful program. The ELI at Umm Al-Qura University (UQU Guide) stipulates a good structural groundwork (40/45/15). Implemented carefully, such a framework can enhance fairness, strengthen teaching quality, and support strategic performance improvement across ranks. Conversely, it can lead to additional stress for both the program administrators and the faculty, who may feel threatened by the entire process. Moreover, over-evaluating faculty can produce feelings of being undervalued, while under-evaluating them can result in problems being overlooked. The appropriate balance should be in line with the mission of the institution and should lead to more efficient attainment of instructional objectives.

**Funding:** This research received no external funding

**Conflicts of Interest:** The author declares no conflict of interest.

**ORCID iD:** Basim Nadhreen (<https://orcid.org/0009-0004-9257-4366>)

### **References**

- [1] Almubarak, A., Alhalabi, W., Albidewi, I., & Alharbi, E. (2025). *An AI-powered framework for assessing teacher performance in classroom interactions: A deep learning approach*. *Frontiers in Artificial Intelligence*, 8, 1553051. <https://doi.org/10.3389/frai.2025.1553051>
- [2] Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators* (3rd ed.). Anker.
- [3] Benton, S. L., & Cashin, W. E. (2014). Student ratings of teaching: A summary of research and literature. IDEA Center. <https://www.ideaedu.org/idea-papers/>
- [4] Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48–62. <https://www.isetl.org/ijtlhe/pdf/IJTLHE8.pdf>
- [5] Brindley, G. (2001). Assessment. In Carter, R., & Nunan, D. (Eds.). *Teaching English to Speakers of Other Languages* (pp. 137-143). Cambridge: Cambridge University Press.
- [6] Cashin, W. E. (1995). Student ratings of teaching: The research revisited. IDEA Center. <https://www.ideaedu.org/idea-papers/>
- [7] Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. Jossey-Bass.
- [8] Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.
- [9] Darling-Hammond, L. (2013). Getting teacher evaluation right: What really matters for effectiveness and improvement. *Teachers College Record*, 115(12), 1–58.
- [10] Fanselow, J.F. (1990). "Let's see": contrasting conversation about teaching, In Richard, J.C., & Nunan, D. (Eds.). *Second Language Teacher Education* (pp. 183-197). Cambridge: Cambridge University Press.
- [11] Fradd, S.H., McGee, P.L., & Wilen, D.K. (1994). *Instructional Assessment: An Integrative Approach to Evaluating Student Performance*. Reading, MA: Addison Wesley.
- [12] Geddes, J.M., & Marks, D.R. (1997). Personnel matters. In Christison, M.A. & Stroller, F.L. (Eds.). *A Handbook for Language Program Administrators* (pp.199-218) Burlingame, CA: Alta Book Center.
- [13] Genesee, F. (2001). Evaluation. In Carter, R., & Nunan, D. (Eds.). *Teaching English to Speakers of Other Languages* (pp. 144-150). Cambridge: Cambridge University Press.
- [14] Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education*, 32(5), 603–615. <https://doi.org/10.1080/03075070701573773>

- [15] Hecht, I.W. D., Higgerson, M.L., Gmelch, W.H., & Tucker, A. (1999). *The Department Chair as Academic Leader*. Phoenix, AZ: American Council on Education and Oryx Press.
- [16] Hutchings, P. (Ed.). (1998). *The course portfolio: How faculty can examine their teaching to advance practice and improve student learning*. AAHE.
- [17] Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319–383). Springer.
- [18] Nunan, D. (1988). *The Learner Centered Curriculum: A Study in Second language Teaching*. Cambridge: Cambridge University Press
- [19] Nunan, D. (1999). *Second language Teaching and Learning*. Boston, MA: Heinle & Heinle.
- [20] Ory, J. C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? In M. Theall, P. C. Abrami, & L. A. Mets (Eds.), *The student ratings debate: Are they valid? How can we best use them?* (pp. 27–44). Jossey-Bass.
- [21] Palomba, C.A. and Banta, T.W. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*, San Francisco, CA: Jossey- Bass.
- [22] Penny, A. R. (2003). Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education*, 8(3), 399–411. <https://doi.org/10.1080/13562510309396>
- [23] Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers: Strategies for teacher learning*. Cambridge University Press.
- [24] Seldin, P. (2006). *Evaluating faculty performance: A practical guide to assessing teaching, research, and service* (3rd ed.). Anker.
- [25] Song Jiao, S., Sururi, S., Aan Komariah, & Kurniatun, T. C. (2026). *Integrating artificial intelligence into teacher performance evaluation: Evidence from undergraduate institutions in Anhui Province*. *Journal of Sustainable Competitive Intelligence*, 16, e0606. <https://doi.org/10.37497/eagleSustainable.v16i.606>
- [26] Stake, R. E. (2004). *Standards-based & responsive evaluation*. SAGE.
- [27] Steinberg, M. P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, 38(2), 293–317. <https://doi.org/10.3102/0162373715615943>
- [28] Spooen, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598–642. <https://doi.org/10.3102/0034654313496870>
- [29] Umm Al-Qura University, University Vice-Presidency for Academic Affairs, Deanship of Development and Quality. (2022, August 25). *Faculty member performance development guide* (Version 11) [PDF]. Umm Al-Qura University. <https://uqu.edu.sa/hrdept/132798>
- [30] Weller, S. (2009). What does 'peer observation of teaching' mean? A literature review. *International Journal of Teaching and Learning in Higher Education*, 21(1), 25–35.
- [31] Wolf, K. (1991). The schoolteacher's portfolio: Issues in design, implementation, and evaluation. *Phi Delta Kappan*, 73(2), 129–136.
- [32] Yao, Y., & Grady, M. L. (2005). How do faculty make formative use of student evaluation feedback? A multiple case study. *Journal of Personnel Evaluation in Education*, 18(2), 107–126. <https://doi.org/10.1007/s11092-006-9000-9>