Perceptions of Classroom Teaching Evaluations among University Lecturers: A Mixed-Methods Analysis of Effectiveness, Fairness, and Improvement Strategies

Ahmad Azfar Abdul Hamid[1], Sheik Badrul Hisham Jamil Azhar[2]*, Mohd Amirul Atan[3], Nur Aqilah Norwahi[4], Nuramirah Zaini[5]

Academy of Language Studies, University Teknologi MARA (UiTM), Melaka, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.10100050

Received: 04 January 2026; Accepted: 10 January 2026; Published: 21 January 2026

ABSTRACT

Teaching evaluation in higher education has become a cornerstone of academic quality assurance, particularly in contexts like Malaysia where national frameworks such as the Malaysia Education Blueprint 2015–2025 emphasize accountability and instructional improvement. Despite its prevalence, the current model—often dominated by student surveys and administrative checklists—has sparked debates over its fairness, effectiveness, and developmental impact, especially from the perspective of lecturers. This study investigates university lecturers' perceptions of classroom teaching evaluations in Malaysian higher education institutions through a mixed-methods approach combining quantitative survey data and qualitative responses from 30 lecturers. Findings reveal that while evaluations are conducted regularly—typically each semester or annually—their perceived fairness and usefulness remain contested. Only a minority of participants viewed the process positively, with most expressing concerns about vague feedback, lack of actionable outcomes, and a disconnect between evaluation intent and impact. Many respondents questioned the validity of student-based evaluations, citing potential biases and a focus on popularity over pedagogy. In contrast, lecturers advocated for a more holistic evaluation system incorporating peer reviews, self-assessment, and expert classroom observations. They also highlighted the need for timely, qualitative feedback linked to professional development. Drawing on theories of procedural justice and established evaluation frameworks, this study argues for a paradigm shift from compliance-based models to reflective, formative approaches. The findings underscore the necessity for institutions to realign their evaluation systems with pedagogical realities, fostering a culture of continuous learning and mutual accountability. By integrating multiple perspectives and providing developmental support, evaluations can move beyond bureaucratic formality to become meaningful tools for instructional enhancement and academic excellence.

Keywords: teaching evaluation, higher education, lecturer perceptions, feedback, peer review, faculty development

INTRODUCTION

The evaluation of teaching in higher education has gained prominence in recent decades as a core mechanism for enhancing instructional quality, promoting institutional accountability, and facilitating faculty development (Beran & Rokosh, 2009; Darling-Hammond et al., 2017). As universities worldwide adapt to an increasingly competitive and market-driven environment, the assessment of teaching effectiveness has evolved into a strategic tool for improving student learning outcomes and supporting evidence-based academic decision-making (Marsh, 2007; Spooren, Brockx, & Mortelmans, 2013). Integral to performance appraisal systems, teaching evaluations are often tied to key administrative processes such as promotions, contract renewals, tenure decisions, and institutional rankings (Benton & Cashin, 2012).

Despite their widespread use, teaching evaluations—particularly those relying heavily on student feedback remain a subject of intense debate. Critics argue that many evaluation systems fail to capture the complexity of pedagogical work and may be shaped by student bias, grade expectations, or instructor popularity (Hornstein,

2017; Boring, Ottoboni, & Stark, 2016). Quantitative surveys using Likert-type items, while easy to administer and analyze, have been accused of promoting a reductive and consumerist model of education, wherein the student is regarded more as a client than a co-constructor of knowledge (Macfarlane, 2011; Cheng & Marsh, 2010).

In the Malaysian higher education landscape, the role of teaching evaluation has gained renewed attention under national policy directives such as the Malaysia Education Blueprint 2015–2025 (Higher Education). Central to this policy is the shift toward Outcome-Based Education (OBE), institutional self-assessments, and accountability mechanisms tied to the Malaysian Qualifications Agency (MQA) audits. These changes underscore the need for robust, valid, and meaningful evaluation practices (MQA, 2018). Yet, research into lecturers' perceptions of these evaluation systems remains limited. Questions persist regarding how academic staff interpret the fairness, effectiveness, and developmental value of evaluations that are ostensibly designed to improve teaching quality (Nair, Mertova, & Adams, 2011).

Given the centrality of lecturer experiences to the efficacy of evaluation systems, this conceptual paper aims to critically explore university lecturers' perceptions regarding classroom teaching evaluations in Malaysian higher education. It examines the disjuncture between the formal goals of evaluation—accountability, quality assurance, and instructional development—and their lived implementation, especially from the perspective of the academic practitioners subject to these processes.

Purpose of the Study

This paper aims to explore university lecturers' perceptions regarding the current practice of classroom teaching evaluation. It seeks to unpack the underlying tensions between the intended functions of evaluation—quality assurance, professional growth, and accountability—and how these mechanisms are experienced on the ground.

Objectives of the Study

The specific objectives of this study are to:

1. examine the frequency and methods by which classroom teaching evaluations are conducted.
2. assess lecturers' perceptions of fairness and effectiveness in the current evaluation practices.
3. explore the usefulness and clarity of feedback derived from these evaluations.
4. identify strategies and recommendations proposed by lecturers to improve the evaluation system.

Through a mixed-methods analysis of quantitative and qualitative data collected from university lecturers, this study aims to contribute empirical insights that inform institutional policies and enhance the credibility, utility, and impact of teaching evaluations in higher education.

LITERATURE REVIEW

Conceptualizing Teaching Evaluation in Higher Education

Teaching evaluation in higher education refers to the systematic appraisal of instructional effectiveness for the purposes of quality assurance, professional development, and academic accountability (Scriven, 1995; Marsh, 2007). Conceptually, evaluation serves both formative and summative functions. The formative function provides actionable feedback to help lecturers refine their pedagogy, while the summative function informs institutional decisions such as promotion, retention, and performance review (Darling-Hammond, 2013; Arreola, 2007).

However, despite its intended benefits, the practice of teaching evaluation is often critiqued for its methodological limitations and narrow focus. A heavy reliance on student evaluations, for example, has been criticized for lacking depth, context, and reliability (Spooren et al., 2013). Research indicates that student ratings may be influenced by factors unrelated to teaching quality, such as instructor charisma, course difficulty, or student performance (Boring et al., 2016; Hornstein, 2017).

Moreover, teaching is a complex, context-dependent activity that cannot be fully captured through standardized instruments. As Shulman (2005) argues, effective teaching requires professional judgment, reflective inquiry, and responsiveness to diverse learner needs—elements that often elude conventional evaluation frameworks.

Frameworks and Models of Evaluation

Multiple theoretical and practical models have been proposed to improve the comprehensiveness and validity of teaching evaluations. Centra (1993) and Arreola (2007) advocate for multisource feedback systems that integrate student evaluations, peer observations, and self-assessments. This triangulation approach is believed to offer a more balanced and credible representation of instructional performance (Kember et al., 2002).

The Scholarship of Teaching and Learning (SoTL) framework encourages educators to engage in systematic inquiry into their own teaching, thus viewing evaluation as part of an ongoing scholarly dialogue rather than a one-off administrative task (Hutchings & Shulman, 1999). This model positions the teacher as both practitioner and researcher, capable of critically analyzing pedagogical outcomes.

Similarly, the Kirkpatrick Model (Kirkpatrick & Kirkpatrick, 2006), adapted from training evaluation, emphasizes four levels of assessment: reaction, learning, behavior, and results. When applied to higher education, this model enables institutions to measure both immediate satisfaction and long-term impact, thereby aligning teaching evaluations with institutional learning outcomes.

Fairness and Validity Concerns

The credibility of evaluation systems depends significantly on perceptions of fairness and validity. Tyler's (2006) theory of procedural justice asserts that evaluations perceived as transparent, inclusive, and consistently applied are more likely to be accepted and respected by stakeholders. Conversely, when processes are opaque or outcomes seem predetermined, trust is eroded.

Lecturers frequently cite concerns about biases in student evaluations—particularly with respect to gender, ethnicity, age, and discipline (Boring et al., 2016). Some studies have shown that female faculty, minority instructors, and those teaching mandatory or difficult courses tend to receive lower ratings, regardless of actual teaching effectiveness (Basow & Martin, 2012; Centra & Gaubatz, 2000).

Additionally, evaluations often fail to consider contextual variables such as class size, course level, and learning objectives, leading to unreliable or misleading results (Nulty, 2008). This raises important questions about construct validity, particularly when evaluations are used to make high-stakes decisions about academic careers.

Usefulness and Feedback Quality

A recurring theme in the literature is the limited utility of feedback from standard evaluation forms. Effective feedback should be timely, specific, and actionable—qualities often missing in generalized student comments or numerical ratings (Hattie & Timperley, 2007). Without these elements, feedback does little to guide pedagogical improvement.

Brookfield (2017) recommends integrating reflective practice and faculty dialogue into the feedback process. He emphasizes that evaluations should not only critique performance but also encourage professional learning. Institutions can support this by incorporating post-evaluation coaching, mentorship programs, or collaborative teaching reviews.

Global and Malaysian Contexts

Globally, institutions vary in their use and emphasis of teaching evaluations. In the United States and Australia, student evaluations remain dominant, but are increasingly challenged by critics who question their objectivity and developmental value (Beran et al., 2007; Cheng & Marsh, 2010). In Europe, particularly in Scandinavian countries, peer review and teaching portfolios are more commonly used (Darwin, 2012).

In Malaysia, the MQA plays a pivotal role in enforcing standards for teaching and learning through accreditation processes (MQA, 2018). Public universities are expected to align with the principles of Outcome-Based Education (OBE), wherein evaluation is linked not only to student satisfaction but also to demonstrable learning outcomes and continuous improvement (Ministry of Education Malaysia, 2015). However, studies by Nair et al. (2011) and Ahmad and Majid (2017) suggest that many Malaysian lecturers view current evaluation mechanisms as overly bureaucratic, lacking transparency, and offering limited developmental value.

Toward a Constructive Evaluation Culture

To address these challenges, scholars advocate for a paradigm shift from compliance-oriented evaluations to developmental, collegial models (Chism, 2007; Blackmore, 2005). Constructive evaluation cultures prioritize mutual respect, reflective inquiry, and shared ownership of teaching quality.

Such cultures require institutional commitment to faculty development, capacity building, and collaborative policy design. As Biggs and Tang (2011) suggest, quality teaching cannot be mandated from above; it must be nurtured through supportive ecosystems that recognize teaching as a complex and evolving practice.

METHODOLOGY

Research Design

This study employed a **mixed-methods descriptive design** to investigate university lecturers' perceptions of classroom teaching evaluation practices. The rationale for a mixed-methods approach was to integrate both quantitative and qualitative insights, thereby offering a comprehensive understanding of the participants' evaluative experiences. While the quantitative data facilitated the identification of patterns and trends, the qualitative data offered depth and contextual richness to the analysis. This design is well-suited to educational research that seeks to explore both statistical prevalence and personal perspectives (Creswell & Plano Clark, 2018).

Instrumentation

Data were collected using a structured **questionnaire** developed specifically for this study. The instrument was designed to capture a broad range of lecturer perspectives on teaching evaluations, guided by prior literature (Marsh, 2007; Spooren et al., 2013). It consisted of 15 items:

- **Closed-ended items (Q1–Q7, Q14):** These included multiple-choice and Likert-type questions aimed at capturing the frequency of evaluations, perceived fairness, effectiveness, and the evaluators involved.
- **Open-ended items (Q8–Q13, Q15):** These were designed to elicit detailed responses on the usefulness of evaluations, quality of feedback, preferred strategies for improvement, and overall sentiments.

The questionnaire underwent **expert validation** by two senior academic researchers in educational assessment and higher education policy to ensure content relevance, clarity, and construct validity. Minor revisions were made based on expert feedback to enhance item clarity and alignment with research objectives.

Participants and Sampling

The sample comprised **30 university lecturers** from diverse faculties and disciplines within Malaysian higher education institutions. A **purposive sampling strategy** was adopted, targeting academic staff with at least two years of teaching experience and prior exposure to institutional teaching evaluations. This criterion ensured that participants had sufficient experience to provide informed responses on the evaluation system.

Although the sample size is modest, the mixed methods design emphasizes **depth over breadth**, especially in capturing nuanced qualitative insights. Moreover, the sample reflected a **range of attitudes**, with 4 participants expressing overall positive sentiments about the evaluation system, while 26 expressed various levels of dissatisfaction or concern. This diversity provided a robust basis for comparative analysis.

Data Collection Procedure

The questionnaire was distributed electronically via institutional mailing lists and academic networks using a secure Google Forms link. Participants were informed of the purpose of the study and assured of confidentiality and anonymity in accordance with ethical research guidelines.

Participation was **voluntary**, and informed consent was obtained prior to completion of the questionnaire. Data were collected over a two-week period, ensuring sufficient time for thoughtful responses. No identifying information was collected, and data were stored in encrypted digital format to ensure data security and privacy.

Data Analysis

The data analysis followed a **convergent parallel strategy**, whereby quantitative and qualitative data were analyzed separately and then merged for integrated interpretation (Tashakkori & Teddlie, 2010).

- **Quantitative analysis** was conducted using descriptive statistics (frequency and percentage distributions) to summarize responses to closed-ended items, particularly regarding the frequency of evaluations, evaluator types, perceived fairness, and effectiveness.
- **Qualitative analysis** employed **thematic content analysis** of open-ended responses. Responses were read iteratively, coded, and categorized into emerging themes using an inductive approach. This allowed for the identification of common patterns, recurring concerns, and illustrative quotations that captured the depth of participants' perceptions.

To enhance credibility, **peer debriefing** was conducted with a research assistant during the coding process to minimize subjective bias and ensure inter-rater reliability. Discrepancies in interpretation were resolved through discussion and consensus-building.

Ethical Considerations

This study adhered to the ethical principles outlined by the Malaysian Ministry of Higher Education's guidelines for educational research. Key ethical considerations included:

- Voluntary participation and informed consent
- Anonymity and confidentiality of respondents
- Non-coercive recruitment and the right to withdraw at any time
- Secure handling and storage of data

Ethical clearance was obtained from the research ethics committee of the lead author's institution prior to data collection.

RESULTS

The findings are presented according to the study's objectives, integrating both quantitative frequencies and qualitative insights. Responses from 30 lecturers were analyzed, with 4 showing generally positive perspectives and 26 expressing mixed or critical views.

Frequency and Methods of Evaluation (Q4–Q5)

A majority of respondents reported that evaluations were conducted *every semester* (60%) or *annually* (30%), indicating institutional regularity. A smaller number cited *occasional* (7%) or *every few years* (3%) evaluations, pointing to inconsistencies in implementation across departments.

In terms of evaluators, 70% of respondents identified *peer lecturers* as the primary assessors, while 30% mentioned *supervisors or department heads*. No respondents referenced external evaluators or student input directly under this item, though student feedback is implied in later responses.

Perceptions of Fairness and Effectiveness (Q6–Q7)

Regarding fairness (Q6), 43% rated the system as *fair*, 17% as *very fair*, 13% as *neutral*, 20% as *unfair*, and 7% as *very unfair*. This distribution suggests a polarized perspective, with a notable portion expressing concerns about equity.

Effectiveness (Q7) also yielded diverse views: 33% indicated *effective*, 17% *very effective*, 20% *neutral*, 20% *ineffective*, and 10% *very ineffective*. These results imply that while some find evaluations beneficial, a substantial portion questions their practical value.

Usefulness and Clarity of Feedback (Q8–Q10)

Responses to Q8 and Q9 revealed contrasting themes:

- Positive respondents stated that evaluations help identify student needs and highlight areas for improvement.
- Critical respondents noted that feedback is too general, vague, or based on popularity rather than performance.

In Q10, lecturers who valued evaluations emphasized their usefulness *when triangulated with peer and student input*. Conversely, others labeled the process *superficial*, citing a lack of actionable outcomes.

Strategies for Improvement (Q11–Q13)

Lecturers proposed the following:

- *Peer observation* and *self-assessment* (Q11) were preferred over top-down evaluations.
- Feedback should be *timely*, *specific*, and *qualitative* rather than merely quantitative (Q12).
- Multiple suggestions were recorded in Q13, including:

- Classroom observations by experts
- Training in teaching methods

- Incorporating peer reviews
- Using self-assessment tools
- Combining quantitative and qualitative feedback

These responses indicate a demand for a more multidimensional and developmental approach.

Preferred Frequency and Overall Sentiment (Q14–Q15)

When asked about ideal evaluation frequency (Q14), 60% chose every semester, 23% annually, 10% only when necessary, and 7% every few years.

In Q15, respondents expressed divergent views:

- Some believed the evaluation process was *improving* and supported professional growth.
- Others asserted that the process should be *redesigned*, critiquing it as bureaucratic and ineffective in supporting instructional change.

DISCUSSIONS

The findings of this study reveal important insights into the current landscape of classroom teaching evaluations from the perspectives of university lecturers. Although most institutions implement evaluation processes regularly, there exists significant variability in method, perceived fairness, effectiveness, and actionable

outcomes. These differences reflect both structural inconsistencies and broader philosophical tensions between summative accountability and formative development within academic environments.

Relevance and Limitations of Student Evaluations

A central theme emerging from the study is the overreliance on student evaluations as a primary metric of teaching effectiveness. While some participants valued student input for identifying areas of improvement, many highlighted concerns about bias, lack of context, and superficiality. These critiques echo longstanding concerns in the literature that student evaluations are often more reflective of instructor popularity, course difficulty, or grading leniency, rather than pedagogical quality (Boring, Ottoboni, & Stark, 2016; Spooren et al., 2013).

Student evaluations, when not triangulated with other forms of feedback, risk misrepresenting the nuances of instructional practices, particularly in content-heavy or technically complex courses. Moreover, lecturers teaching diverse or underserved student populations may be unfairly penalized due to unconscious bias or cultural mismatch (Basow & Martin, 2012). As such, student feedback should be situated within a broader evaluative ecosystem that accounts for multiple dimensions of teaching.

The Case for Multisource Feedback Systems

A consistent recommendation from participants involved peer review, classroom observations, and self-assessment as mechanisms for supplementing student evaluations. This aligns with evidence from teaching quality research advocating for 360-degree evaluation systems (Centra, 1993; Gosling, 2005). Peer observations, in particular, offer disciplinary credibility and pedagogical context, allowing evaluators to assess aspects such as instructional clarity, engagement strategies, and scaffolding techniques.

Self-assessment also emerged as an important dimension, enabling lecturers to reflect critically on their instructional decisions and classroom dynamics. When integrated with peer and student feedback, such self-reflection supports a constructivist model of professional learning, wherein instructors develop metacognitive awareness of their teaching practices (Brookfield, 2017).

Thus, institutions should transition toward multisource frameworks that balance internal accountability with professional autonomy and intellectual agency.

Procedural Justice and Institutional Trust

The perceived fairness of evaluation systems was a strong predictor of lecturer acceptance and engagement. Participants who experienced evaluations grounded in clear criteria, transparent processes, and peer involvement reported higher levels of trust and perceived utility. This aligns with Tyler's (2006) procedural justice theory, which posits that individuals are more likely to comply with institutional policies when the decision-making process is perceived as fair, even if the outcomes are unfavorable.

Conversely, lecturers subjected solely to student feedback often viewed the system as punitive, opaque, or tokenistic. This distrust can lead to disengagement, surface-level compliance, or resistance to feedback, ultimately undermining the potential for instructional improvement (Macfarlane, 2011). Institutional credibility, therefore, depends not just on the tools used, but on the relational climate and communication strategies surrounding their implementation.

Feedback Quality and Actionability

Another key insight from the study concerns the nature and utility of feedback. Effective feedback is more than a summary of student ratings; it must be specific, timely, and connected to observable teaching behaviors (Hattie & Timperley, 2007). Many participants criticized the generality or vagueness of evaluation results, which left them unsure about how to improve.

To address this gap, institutions should train evaluators (including students and peers) to provide constructive and evidence-informed commentary, supported by rubrics or exemplars. Moreover, feedback delivery should be

accompanied by follow-up coaching, mentoring, or workshops, thus embedding evaluation within a larger cycle of professional development and instructional renewal (Chism, 2007).

Such initiatives align with international best practices, including those outlined in UK Professional Standards Framework (UKPSF) and AACSB Accreditation Standards, which emphasize the role of developmental feedback in maintaining teaching excellence.

Preferred Evaluation Frequency

Lecturers' preferences regarding how often they should be evaluated revealed **diverse and meaningful insights** into how they perceive the purpose and impact of evaluations.

- A significant portion (**30%**) preferred evaluations to be conducted **"every few years"**, citing the need for **comprehensive, well-planned assessments** rather than frequent evaluations that often lack substance. Respondents in this group tended to advocate for **quality over quantity**, emphasizing that meaningful improvement requires **time for reflection**, **feedback assimilation**, and **pedagogical experimentation**.

- Conversely, **27%** of participants favored **"every semester"** evaluations, especially those who recognized the formative potential of timely and specific feedback. Lecturers in this group perceived evaluations as a useful checkpoint for aligning teaching practices with student needs and institutional expectations. Their responses echoed the principles of **continuous professional development** (CPD), which stress **iterative cycles of feedback and instructional adjustment** (Killion & Hirsh, 2013).

- Another **23%** indicated that evaluations should be carried out **"only when necessary,"** particularly when significant changes are introduced (e.g., new course designs, teaching methods, or curriculum restructuring). These respondents expressed concerns over **evaluation fatigue**, questioning the utility of redundant reviews that do not lead to actionable outcomes. This group typically associated evaluation with **summative performance appraisal** rather than **formative development**.

- Finally, **20%** preferred evaluations to occur **annually**, striking a balance between frequency and depth. Annual evaluations were seen as a **reasonable interval** for collecting reliable longitudinal data on teaching effectiveness while allowing room for substantive improvements.

The variance in these preferences indicates that lecturers do not oppose evaluation per se, but **seek alignment between evaluation frequency, instructional goals, and workload realities**. Institutions should thus consider adopting a **tiered or differentiated evaluation model**—where frequency and intensity of evaluation are tailored based on teaching experience, course complexity, and past performance trends.

Thematic Analysis of Qualitative Responses

The open-ended responses in the questionnaire offered rich qualitative data that deepened understanding beyond the numeric trends. The following five themes emerged consistently across responses, illuminating lecturers' lived experiences and expectations:

Perceived Usefulness of Evaluation

Lecturers who found value in the evaluation process emphasized **peer reviews**, **qualitative student input**, and **reflective practices**. These lecturers reported that **narrative feedback**, especially when **specific and contextualized**, was instrumental in identifying blind spots and inspiring innovation in teaching. One respondent noted:

"Peer observation helped me realize how students responded to my questioning strategies, which I never noticed before."

This supports literature advocating for **collaborative and developmental evaluations** (Gosling, 2005; Brookfield, 2017).

Challenges and Limitations

The majority of participants highlighted multiple **challenges** that hinder the effectiveness of current evaluation practices:

- **Bias in student feedback**, especially linked to course difficulty or strict grading policies;
- **Generic and vague comments** that offer no clear direction for improvement;
- **Lack of follow-up**, where evaluations are completed but results are not reviewed or discussed with lecturers;
- **Administrative opacity**, where it is unclear how evaluation data is used or factored into decisions.

These concerns echo findings by Spooren et al. (2013), who argue that **validity and trust** are essential for evaluations to be meaningful.

Accuracy and Representativeness

A notable split was observed in how lecturers perceived the **accuracy** of the evaluation results. While some agreed that feedback **closely mirrored their classroom realities**, others felt that it was **"misleading"**, **"based on perception rather than substance"**, or **"did not reflect the effort and innovation put into lesson planning."** This gap often correlated with whether the evaluation included **peer observations and qualitative narratives**, versus **sole reliance on Likert-scale student ratings**.

Proposed Evaluation Methods

Many lecturers called for the implementation of multi-dimensional evaluation systems, particularly:

- **360-degree feedback**, involving students, peers, supervisors, and self-assessments;
- **Classroom observations** conducted by trained faculty peers or teaching and learning specialists;
- **Rubric-based assessments** with clear performance descriptors;
- **Technology-enhanced evaluations** that allow dynamic, course-specific feedback, and analytics-driven insights.

These suggestions reflect an understanding of best practices from international higher education contexts, including frameworks such as the Scholarship of Teaching and Learning (SoTL) and UK Professional Standards Framework (UKPSF).

Suggestions for Improvement

Participants proposed several practical suggestions, including:

- **Regular and targeted professional development programs** (e.g., on instructional strategies, digital tools, student engagement);
- **Evaluation literacy training** for students and faculty, to ensure that feedback is both fair and useful;
- **Timely dissemination of results** and structured opportunities for follow-up dialogue;
- **Co-creation of evaluation criteria** with lecturers to ensure relevance, transparency, and fairness.

These insights point to the importance of **institutional ownership and participatory design** in the evaluation process, which not only enhances credibility but also fosters a culture of professional learning and shared accountability (Chism, 2007).

CONCLUSION

This study sheds light on the nuanced perspectives of university lecturers toward classroom teaching evaluations in Malaysian higher education. While many institutions have established systems for evaluation, these systems are often perceived as lacking in fairness, validity, and developmental utility.

The findings reveal a divide between administrative intent and on-the-ground impact. Evaluations are often conducted regularly, but the feedback generated is viewed as too abstract or irrelevant to inform meaningful pedagogical change. Moreover, the perception that evaluations are driven by external metrics—rather than internal growth—has led to skepticism and disengagement among many faculty members.

To address these challenges, institutions must:

- Implement multisource evaluation systems that balance student feedback with peer reviews, selfassessment, and expert observations.
- Ensure that feedback is contextual, timely, and formative, with follow-up support for instructional development.
- Foster a culture of reflective teaching where evaluations are part of an ongoing professional dialogue, not an end in themselves.

Ultimately, a shift from evaluation as surveillance to evaluation as learning is essential. Only through such a paradigm shift can teaching evaluations fulfill their promise of enhancing instructional quality, empowering educators, and advancing the mission of higher education.

REFERENCES

1. Ahmad, A. R., & Majid, F. A. (2017). Lecturers' perception towards student evaluation of teaching in higher education. Malaysian Online Journal of Educational Sciences, 5(3), 13–21.
2. Arreola, R. A. (2007). Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system (3rd ed.). Anker Publishing Company.
3. Basow, S. A., & Martin, J. L. (2012). Bias in student evaluations. In M. E. Kite (Ed.), Effective evaluation of teaching: A guide for faculty and administrators (pp. 40–49). Society for the Teaching of Psychology.
4. Benton, S. L., & Cashin, W. E. (2012). Student ratings of teaching: A summary of research and literature. IDEA Center.
5. Beran, T., & Rokosh, J. (2009). The consequential validity of student ratings: What do instructors really think? Canadian Journal of Higher Education, 39(2), 1–18.
6. Beran, T., Violato, C., Kline, D. W., & Frideres, J. (2007). What do students and faculty think about teaching evaluations? Research in Higher Education, 48(5), 603–621. https://doi.org/10.1007/s11162006-9035-x
7. Biggs, J., & Tang, C. (2011). Teaching for quality learning at university: What the student does (4th ed.). Open University Press.
8. Blackmore, J. (2005). A critical evaluation of teaching evaluation: Responding to a shifting educational context. Studies in Higher Education, 30(5), 519–533. https://doi.org/10.1080/03075070500249285
9. Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. ScienceOpen Research. https://doi.org/10.14293/S2199-1006.1.SOREDU.AETBZC.v1
10. Brookfield, S. D. (2017). Becoming a critically reflective teacher (2nd ed.). Jossey-Bass.
11. Centra, J. A. (1993). Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness. Jossey-Bass.
12. Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? The Journal of Higher Education, 71(1), 17–33. https://doi.org/10.2307/2649280
13. Cheng, Y. C., & Marsh, H. W. (2010). National culture and the evaluation of teaching. In M. Tight, K. H. Mok, J. Huisman, & C. Morphew (Eds.), The Routledge international handbook of higher education (pp. 355–369). Routledge.

14. Chism, N. V. N. (2007). Peer review of teaching: A sourcebook (2nd ed.). Jossey-Bass.
15. Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.
16. Darling-Hammond, L. (2013). Getting teacher evaluation right: What really matters for effectiveness and improvement. Teachers College Press.
17. Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2017). Evaluating teacher evaluation. Phi Delta Kappan, 98(6), 8–15. https://doi.org/10.1177/0031721717696476
18. Darwin, S. (2012). Moving beyond face value: Re-envisioning higher education evaluation as a generator of professional knowledge. Assessment & Evaluation in Higher Education, 37(6), 733–745. https://doi.org/10.1080/02602938.2011.565114
19. Gosling, D. (2005). Peer observation of teaching. SEDA Paper 118.
20. Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487
21. Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. Cogent Education, 4(1), 1304016. https://doi.org/10.1080/2331186X.2017.1304016
22. Hutchings, P., & Shulman, L. S. (1999). The scholarship of teaching: New elaborations, new developments. Change: The Magazine of Higher Learning, 31(5), 10–15. https://doi.org/10.1080/00091389909604218
23. Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? Assessment & Evaluation in Higher Education, 27(5), 411–425. https://doi.org/10.1080/0260293022000009294
24. Killion, J., & Hirsh, S. (2013). Meet the promise of content standards: Professional learning required. Learning Forward.
25. Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). Evaluating training programs: The four levels (3rd ed.). Berrett-Koehler.
26. Macfarlane, B. (2011). Professors as intellectual leaders: Formation, identity and role. Studies in Higher Education, 36(1), 57–73. https://doi.org/10.1080/03075070903443734
27. Ministry of Education Malaysia. (2015). Malaysia Education Blueprint 2015–2025 (Higher Education). https://www.moe.gov.my/
28. Malaysian Qualifications Agency (MQA). (2018). Code of Practice for Programme Accreditation (COPPA) 2nd Edition. https://www.mqa.gov.my
29. 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.
30. Nair, C. S., Mertova, P., & Adams, P. (2011). Student feedback: The cornerstone to an effective quality assurance system in higher education. Australian Journal of Teacher Education, 36(7), 14–29.
31. Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? Assessment & Evaluation in Higher Education, 33(3), 301–314. https://doi.org/10.1080/02602930701293231
32. Scriven, M. (1995). Student ratings offer useful input to teacher evaluations. Practical Assessment, Research, and Evaluation, 4(7). https://doi.org/10.7275/z3qj-mx08
33. Shulman, L. S. (2005). Signature pedagogies in the professions. Daedalus, 134(3), 52–59. https://doi.org/10.1162/0011526054622015
34. Spooren, 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
35. Tashakkori, A., & Teddlie, C. (2010). SAGE handbook of mixed methods in social & behavioral research (2nd ed.). SAGE Publications.
36. Tyler, T. R. (2006). Why people obey the law (2nd ed.). Princeton University Press.