



American Journal of Education and Technology (AJET)

ISSN: 2832-9481 (ONLINE)

VOLUME 5 ISSUE 3 (2026)



PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

Development and Validation of an Enhanced Teacher Supervision Tool in a Chinese Medical University

Haimeng Xi¹, Mark Joseph D. Pastor^{1*}

Article Information

Received: March 17, 2026

Accepted: June 05, 2026

Published: September 02, 2026

Keywords

*Evaluation, Teacher Performance,
Teacher Supervision*

ABSTRACT

This study aimed to develop and validate an enhanced teacher supervision tool for college faculty. Recognizing the limitations of existing evaluation methods, the research sought to create a more comprehensive and accurate system. The study employed a mixed methods approach, particularly a sequential exploratory design using the instrument development model. Five purposively chosen key stakeholders from higher education management and clinical specialties participated in key informant interviews. Four emerging themes were revealed on the process of evaluating teachers: (1) multifaceted approach to teacher evaluation; (2) creating a comprehensive teacher evaluation system; (3) challenges to a fair and effective teacher evaluation system; and (4) building a growth-oriented teacher evaluation system. Based on these insights, a draft evaluation tool consisting of 41 items was developed and refined through multiple rounds of expert feedback and pilot testing among 200 participants from Qilu Medical University, stratified evenly into managers, experts, teachers, and students. Exploratory factor analysis (EFA) confirmed excellent sampling adequacy ($KMO = 0.953$) and revealed three key robust constructs: effective teaching practices (24 items), teacher performance and supervision (14 items), and the relevance and effectiveness of education (3 items). The resulting tool exhibits outstanding reliability (Cronbach's $\alpha = 0.947$) and provides a holistic, data-driven framework to teacher evaluation, supporting targeted professional development and ultimately enhancing the quality of teaching and learning in higher education.

INTRODUCTION

With the deepening of teaching reform in higher education institutions, teacher supervision has become an indispensable instrument for instructional quality monitoring and internal quality assurance systems. Strengthening teacher evaluation and guidance via rigorous supervision pathways holds immense significance for elevating educational standards and fostering sustainable administrative reforms.

Globally, research on teacher supervision frameworks originated in the mid-20th century, primarily focusing on performance accountability and student outcomes assessment (Alwaely *et al.*, 2023). Over time, the paradigm shifted toward comprehensive methodologies that evaluate diverse facets of instruction, including interactive strategies, curriculum architecture, resource deployment, and technology integration. In the educational system of the United States, instructional supervision plays a central, service-oriented role focused on helping educators modify curriculum designs and pedagogical mechanics, evolving into a collaborative and developmental supervision mechanism (Wayne *et al.*, 2015).

Conversely, in China, the regulatory framework consolidated with the promulgation of the Regulations on Education Supervision in 2012 by the State Council, which provided the statutory baseline for higher education quality monitoring. Furthermore, the Index System of Teaching Examination and Evaluation of Undergraduate Education (2021) issued by the Ministry of Education

designated quality supervision and guarantee as a first-level indicator, institutionalizing teaching supervision as an essential second-level indicator. Consequently, contemporary universities are heavily committed to constructing operational quality management frameworks. However, domestic supervision practices usually reveal critical systemic vulnerabilities. Current systems reflect a traditional 'two-level system' (university-level committee and school-level groups) or macro 'three-level system' including provincial elements (Wang Xuejun, 2017; Wang Mingxia, 2017). These frameworks provide macro guiding recommendations but lack specialized, specialize operational tools. A prominent challenge is the persistent phenomenon of 'emphasizing supervision rather than guidance,' where evaluators ignore constructive follow-up links, leaving data gathered from evaluations underutilized. Moreover, while information technologies are increasingly deployed, they remain simple tools for data collection rather than standard-driven instruments for diagnostic improvement (Zhao Guozhu, 2020). Therefore, developing a granular, validated, and comprehensive supervision tool is of urgent practical necessity.

This study directly addresses these challenges by developing and validating an enhanced university teacher supervision tool within the context of medical education, specifically anchored at Qilu Medical University. Qilu Medical University, a specialized institution in Eastern China, accommodates approximately 22,000 students

¹ Mariano Marcos State University, Graduate School, Laoag City, Ilocos Norte, 2900, Philippines

* Corresponding author's e-mail: mjpastor@mmsu.edu.ph

and over 1,000 faculty members, offering a representative environment for evaluating complex teaching workflows. The specific objectives are to: (1) investigate current practices for supervising and evaluating teacher performance through stakeholders' qualitative accounts; and (2) develop and statistically validate an enhanced evaluation tool that comprehensively measures teaching competence to guide institutional policy making, targeted training, and professional retention.

LITERATURE REVIEW

Instructional supervision serves as the cornerstone of quality monitoring, necessitating a diversified construction across institutions, personnel tiers, and technological platforms (Liu Qiyue, 2018). Effective systems rely heavily on a 'four-in-one' collaborative loop comprising student evaluation, peer evaluation, supervision evaluation, and administrative review to maintain scientific and fair outcomes (Chen Yang, 2023). Empirical evidence establishes that rigorous evaluation methods positively correlate with enhanced faculty practices and student academic achievements (Alwaely *et al.*, 2023). Constructive and continuous feedback loop patterns are proven to catalyze pedagogical growth, encouraging responsible completion of academic duties (Majumdar, 2022).

Contemporary evaluation frameworks must look beyond singular data sources to mitigate bias. High-quality systems integrate multi-source inputs including direct classroom observations, student work products analysis, and structured self-reflection (Darling-Hammond *et al.*, 2017). In parallel, educational management research stresses the role of instructional and transformational leadership, where administrators are evaluated on their capacity to nurture collaborative climates, empower teaching staff, and deploy multi-faceted datasets for informed decision-making (Hallinger & Heck, 2010; Leithwood *et al.*, 2010). Under the guidelines of total quality management, systems must enforce three strict principles: full staff participation, whole-process monitoring (forming a closed-loop system), and all-round multidimensional assessment spanning instruction methods, space, and institutional resources (Wang Lei, 2021).

A comprehensive evaluation architecture must acknowledge the distinctive limitations of individual assessment methods. For instance, student evaluations offer crucial data regarding classroom climate and instructor popularity but are highly susceptible to variables such as course difficulty (Hodges *et al.*, 2021). Peer reviews provide profound professional insights into pedagogical content accuracy but are occasionally influenced by personal relationships among colleagues (Chen *et al.*, 2022). Self-assessments are essential for reflective practice but suffer from intrinsic subjectivity and limited self-awareness (Kember *et al.*, 2020). Consequently, combining these approaches within a balanced matrix—such as a 360-degree feedback framework—provides an objective, nuanced, and growth-oriented assessment of faculty competence (Mohammadi, 2021).

Theoretical Framework

This research is theoretically underpinned by the 360-degree feedback model, quality assurance theory, and the theory of educational evaluation. The 360-degree feedback model shifts performance appraisal from exclusive top-down supervisor ratings to an multidimensional, multi-source assessment including superiors, peer colleagues, students, and self-appraisals (Simonson *et al.*, 2021). The execution standard structurally advances through six distinct sequential phases: standard design, provider selection, data collection, analytical processing, feedback delivery, and subsequent tracking evaluation. Within this study, the model operationalizes feedback loops across diverse university actors to capture holistic faculty performance.

Quality assurance theory provides the structure for setting consistent norms and operational standards. It leverages the Continuous Quality Improvement (CQI) cycle, explicitly implementing the Plan-Do-Check-Act (PDCA) model to transform teaching planning, instructional execution, and examination management into a reliable system (Wang Lei, 2021). Lastly, the theory of educational evaluation provides the methodological baseline for clarifying supervision purposes, constructing valid measurement tools, and translating raw statistical trends into targeted professional training and optimized administrative policy decisions.

MATERIALS AND METHODS

This study adopted a mixed methods research design, specifically utilizing a sequential exploratory approach based on the instrument development model (Creswell & Plano, 2011). This framework coordinates qualitative inquiry and quantitative confirmation within a unified investigative continuum, ensuring that the development of the enhanced supervision tool directly arises from field insights before undergoing rigorous psychometric validation.

In the first qualitative stage, a descriptive phenomenological approach was implemented to capture the concrete lived experiences, perspectives, and structural recommendations of key academic and clinical stakeholders (Smith *et al.*, 2009). The qualitative findings provided the substantive indicators required to construct the initial draft of the tool. The second quantitative stage deployed an exploratory approach focused on factor composition and internal consistency analysis. A structured survey questionnaire, utilizing a 5-point Likert scale, was developed from the qualitative draft and distributed to a broader, distinct institutional sample to refine item parameters, eliminate redundant elements, and verify the structural validity of the final supervision instrument.

Participants and Sampling Technique

For the quantitative phase, stratified random sampling was implemented to select 200 distinct participants while ensuring zero overlap with the qualitative cohort. The expanded sample size was adopted to strengthen

the statistical stability and psychometric rigor of exploratory factor analysis procedures. The final sample consisted of 50 academic managers, 50 senior experts, 50 active teaching faculty, and 50 undergraduate students. Although the sample-to-item ratio was below the ideal 10:1 recommendation, the adequacy of the sample was supported by the high KMO value, strong communalities, and stable factor structure (Hair *et al.*, 2019; Tabachnick & Fidell, 2019).

Although the study utilized 200 respondents for 41 measurement items, methodological literature suggests that exploratory factor analysis adequacy should not be evaluated solely based on absolute sample size. According to Hair *et al.* (2019), strong factor loadings, acceptable communalities, and high Kaiser-Meyer-Olkin (KMO) values can compensate for moderate sample-to-item ratios. MacCallum *et al.* (1999) further argued that stable factor recovery is primarily influenced by communality strength and factor overdetermination rather than sample size alone. In the present study, the KMO value exceeded 0.95 and the factor loading structure demonstrated strong conceptual coherence, thereby supporting the adequacy and stability of the extracted factors.

Data Gathering Instruments and Technical Procedures

Qualitative evidence was gathered via one-to-one, semi-structured key informant interviews (KII), driven by a pre-validated protocol focusing on evaluation standards, current gaps, feedback mechanisms, and structural tool design. Interviews were recorded, transcribed, and checked for accuracy. The resulting qualitative matrix informed the development of a 41-item draft Likert questionnaire. This questionnaire served as the principal quantitative instrument, measuring performance dimensions across classroom management, instructional delivery, political standing, and student engagement.

Qualitative transcripts were processed using Colaizzi's (1978) seven-step thematic analysis method, moving from significant statements extraction to generalized theme clusters creation, which were verified with participants to ensure fidelity. Quantitative datasets were analyzed via Exploratory Factor Analysis (EFA) to discover latent factor constructs, employing the principal axis factoring extraction method combined with direct oblimin rotation. Instrument reliability was checked using Cronbach's alpha internal consistency coefficient, while data suitability for factor validation was checked via Kaiser-Meyer-Olkin (KMO) sample adequacy metrics and Bartlett's test of sphericity.

Ethical approval was secured from the appropriate institutional ethics review board prior to data collection. All participants voluntarily participated and signed informed consent forms. Confidentiality and anonymity were strictly maintained throughout the study.

RESULTS AND DISCUSSION

The qualitative phenomenological stage extracted four

comprehensive themes regarding the evaluation process: (1) Multifaceted Approach to Teacher Evaluation; (2) Creating a Comprehensive Teacher Evaluation System; (3) Challenges to a Fair and Effective Teacher Evaluation System; and (4) Building a Growth-Oriented Teacher Evaluation System. These themes highlight the importance of balancing qualitative classroom observations, multi-source stakeholder appraisals, standardized rubrics, and technology integration to transform performance monitoring into active professional development pathways.

Psychometric and Factor Architecture

Following the integration of qualitative themes into item statements, the 41-item instrument underwent pilot testing and EFA. The instrument demonstrated exceptional reliability, registering a total Cronbach's alpha value of 0.947, indicating superior internal consistency (Owan *et al.*, 2023). Assumption checks for factor analysis yielded an overall Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy of 0.953, while individual item KMO values ranged from 0.864 to 0.978, signifying excellent sample suitability. Concurrently, Bartlett's test of sphericity produced significant results (Chi-Square = 14,928.441, $df = 820$, $p < 0.001$), confirming that the inter-variable correlation matrix was highly suitable for factor extraction.

Principal Components Analysis (PCA) and the resulting scree plot eigenvalues graph identified a sharp drop in variance explanation after the first few factors, establishing a clear 'elbow' around the third and fourth factors. Consequently, constructs to the right of the inflection point were eliminated, confirming a highly stable 3-factor solution that accounted for 94.3% of the cumulative variance. Factor 1 (Effective Teaching Practices) explained 42.18% of the variance; Factor 2 (Teacher Performance and Supervision) explained 23.07% of the variance; and Factor 3 (Effectiveness and Relevance of Education) accounted for 13.11% of the variance. All 41 items were retained, displaying robust factor loadings ranging from 0.653 to 0.918, with minimal uniqueness parameters.

Analysis of Factor 1: Effective Teaching Practices

Factor 1 represents the pedagogical and instructional foundation of effective university teaching. The high loading values across instructional organization, student engagement, classroom interaction, feedback delivery, and advanced teaching strategies demonstrate that contemporary supervision systems must evaluate teaching as a multidimensional professional activity rather than a simple compliance process. The findings indicate that effective supervision frameworks should emphasize reflective teaching, active learning environments, and continuous instructional improvement. These results are consistent with the work of Darling-Hammond *et al.* (2017), which emphasized that high-quality instructional supervision should prioritize student cognitive engagement, instructional clarity, and evidence-

based pedagogical innovation. Within medical education settings, the integration of clinical examples, advanced research content, and interactive teaching methods further highlights the importance of aligning supervision systems with competency-based education reforms.

Crucially, Factor 1 integrates real-time institutional communication and feedback loops into the teaching practice construct. Items Q32 and Q34 mandate that evaluation results must be summarized and immediately communicated back to the individual teacher, disrupting the traditional, punitive practice of releasing a single score at the end of a semester. Providing granular, diagnostic data allows faculty members to understand their pedagogical weaknesses and proactively participate in targeted professional development workshops (Parpucu & Al-Mabuk, 2023). This structure transforms supervision from an administrative tool for compliance into a supportive growth-oriented framework, consistent with the total quality management concepts explored by Wang Lei (2021).

Analysis of Factor 2: Teacher Performance and Supervision Dynamics

Factor 2 reflects the institutional and operational dimensions of teacher supervision. The inclusion of peer evaluation, classroom observation, work summaries, student feedback, digital monitoring, and follow-up supervision demonstrates that modern university evaluation systems increasingly rely on multi-source evidence rather than isolated administrative judgments.

The results support the growing shift from traditional top-down supervision models toward integrated 360-degree evaluation frameworks that combine accountability with professional support. Previous studies have shown that multi-source evaluation improves fairness, reduces evaluator bias, and promotes long-term professional development (Mohammadi, 2021; Hallinger & Heck, 2010). The integration of digital supervision technologies and continuous feedback mechanisms also strengthens institutional quality assurance systems by supporting data-driven decision-making and sustainable faculty development.

Furthermore, Factor 2 highlights the necessity of implementing a multi-source, 360-degree data collection matrix. It coordinates student evaluations (Q5, 0.784), standardized portfolio reviews (Q36, 0.777), classroom observations (Q3, 0.814), peer reviews (Q4, 0.807), and structured self-evaluations (Q2, 0.761). Relying on multiple inputs counteracts the validity limits of any single method; for instance, it balances student satisfaction trends with rigorous peer content reviews (Chen *et al.*, 2022). Item Q41 (0.799) introduces classified evaluation weighting, adjusting formulas based on college majors, student demographics, and course difficulty tiers. Finally, item Q40 (0.792) enforces follow-up accountability, mandating that subsequent supervision cycles must check whether an instructor has implemented remedial changes based on past evaluation data, thereby establishing an operational PDCA cycle (Wang Lei, 2021).

Table 1: Enhanced Supervision Tool Factors, Item Alignment, and EFA Loadings

Factor Construct	Aligned Item Statements (Enhanced Supervision Tool Code)	Loading
Factor 1: Effective Teaching Practices (Variance: 42.18%) (Cronbach's α : 0.947)	Q12: Organic integration of knowledge/ability/quality to cultivate advanced thinking and complex problem solving.	0.918
	Q15: Course difficulty is reasonable in depth; guides structural background understanding and connections.	0.911
	Q16: High structural importance assigned to advanced instructional strategies and clinical methods.	0.905
	Q17: Teaching objectives are clear, concepts are innovative, language is clear, simple, easy to understand.	0.899
	Q18: Instructional methods stimulate enthusiasm; almost all students are highly focused in class.	0.893
	Q19: Instructor demonstrates passion, deep understanding, and love for the specialized subject matter.	0.887
	Q20/Q21: Classroom is well organized, creating a positive, inclusive, and supportive environment.	0.882
	Q22/Q23: Active classroom interaction; students think critically and make positive responses.	0.876
	Q24: Instructor monitors classroom situation, implementing effective real-time classroom management.	0.871
	Q25: Guides, encourages, and urges deep extracurricular learning, reading, and practice.	0.865

	Q27: Ability to provide constructive feedback, guide improvement, and offer additional support.	0.859
	Q28: Evaluation frameworks require supervisors to objectively evaluate teacher performance.	0.853
	Q31: More targeted training should be systematically organized based on final evaluation results.	0.847
	Q32/Q34: Summary results must be shared as timely, accurate feedback rather than once a semester.	0.841
	Q37: Expands evaluation indicators to include student classroom experience feedback.	0.835
	Q39: Content is contemporary, introducing advanced research results and real situations.	0.829
	Q7/Q8: Reflects solid political standing, moral education, party spirit, and socialist core values.	0.823
	Q9/Q10: Complete lesson preparation, dignified teaching appearance, standard materials, punctuality.	0.817
	Q11: Demonstrates mastery of subject knowledge and key professional concepts.	0.811
Factor 2: Teacher Performance and Supervision (Variance: 23.07%)	Q30: The weight/proportion of numerical evaluation scores in year-end assessment should be reduced.	0.804
	Q33: deploys digital platforms, classroom videos, and software systems to analyze performance data.	0.798
	Q29: Standardized evaluation metrics must be explicitly included in year-end personal assessment.	0.791
	Q5: Student evaluation metrics represent an effective way to monitor long-term performance.	0.784
	Q36: Standard teaching portfolios must include plans, multimedia courseware, logs, and lecture notes.	0.777
	Q6: Systematic review of work summary documents serves as a valid supervision indicator.	0.821
	Q3: Standardized classroom observation protocols remain an essential tool for monitoring.	0.814
	Q4: Peer evaluation methods offer unique professional insights from internal colleagues.	0.807
	Q41: Metrics across different majors and classes must be classified and given adjusted weights.	0.799
	Q40: Follow-up supervision checks whether teachers implemented improvements from past evaluations.	0.792
	Q35: Multimedia courseware platforms must be updated every semester to remain valid.	0.785
	Q38: Invites specialized industry and clinical enterprise personnel to participate in supervision.	0.777
	Q1: Daily communication and observation of behavioral patterns reveal true instructor competence.	0.769
	Q2: Structured teacher self-evaluation forms promote reflective growth and professional tracking.	0.761
Factor 3: Effectiveness and Relevance of Education	Q14: Organic integration of ideological elements, infiltrating moral education with originality.	0.752
	(Variance: 13.11%)	0.743
	Q13: Curriculum reflects advanced frontier trends, utilizing attractive real-life clinical examples.	0.734

Analysis of Factor 3: Effectiveness and Relevance of Education

Factor 3 emphasizes the broader educational relevance and social responsiveness of university teaching. The findings demonstrate that effective supervision systems in higher education must evaluate not only instructional performance but also the relevance of curriculum design, ideological integration, and real-world applicability of learning content. In medical education environments, curriculum relevance and ethical competence have become increasingly important components of educational quality assurance. The inclusion of clinical applicability, frontier disciplinary trends, and student-centered educational effectiveness reflects the transition of higher education institutions toward competency-oriented and socially responsive educational models. Moral elements, ideological integration, and solid political standing (such as Q14) load significantly on this construct, reflecting specific strategic imperatives and holistic cultural expectations within Chinese higher education institutions.

CONCLUSION

This research successfully developed and psychometrically validated an enhanced teacher supervision tool for college faculty. The instrument demonstrates outstanding reliability (Cronbach's $\alpha = 0.947$) and excellent structural validity through exploratory factor analysis. The study confirms that effective instructional monitoring requires an omni-directional, multi-source framework anchored by three robust constructs: effective teaching practices, structured performance supervision mechanics, and educational relevance. The findings indicate that universities must transition away from isolated, high-stakes numerical checklists toward comprehensive systems that prioritize immediate diagnostic feedback and technology integration. By coordinating student, peer, and expert evaluations, the tool successfully balances subjectivity, minimizes systemic appraisal bias, and supports data-driven professional development. The scope of this validation is primarily centered on Qilu Medical University, meaning that wider application across diverse institutional models should carefully account for regional policy shifts and specific discipline requirements.

REFERENCES

- Alwaely, S. A., El-Zeiny, M. G., Alqudah, H. A., Alamarnih, E. F., Salman, O. H., Halim, M. S., & Khasawneh, M. A. S. (2023). The impact of teacher evaluation on professional development and student achievement. *Revista de Gestão Social e Ambiental*, 17(7), 3-8.
- Campbell, S. (2020). Education and curriculum reform: The impact they have on learning. *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, 3(2), 1074-1082.
- Chen, J., Yang, X., & Wu, L. (2022). Peer evaluation in higher education: A review of the literature. *Educational Research Review*, 16, 56-68.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(1), Article 7.
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.
- Eastwood, M., Davies, B., & Rees, E. (2021). Students' experiences of peer observed teaching: A qualitative interview study. *Teaching and Learning in Medicine*, 35(1), 1-9.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership & Management*, 30(2), 95-110.
- Hodges, C., Moore, S., & Locke, B. B. (2021). The validity of student evaluations of teaching: An updated review. *The Journal of Higher Education*, 92(1), 35-47.
- Jones, N. D., Bell, C. A., Brownell, M., Qi, Y., Peyton, D., Pua, D., Fowler, M., & Holtzman, S. (2022). Using classroom observations in the evaluation of special education teachers. *Educational Evaluation and Policy Analysis*, 44(3), 429-457.
- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2010). *How leadership influences student learning*. The Wallace Foundation.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99.
- Mohammadi, M. (2021). Dimensions of teacher performance evaluation by students in higher education. *Shanlax International Journal of Education*, 9(2), 18-25.
- Owan, V. V., Abang, K. B., Idika, D. E., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(4), em2307.
- Parpucu, H., & Al-Mabuk, R. (2023). The role of feedback in teacher professional development. *EIKI Journal of Effective Teaching Methods*, 1(4), 77-85.
- Simonson, S., Earl, B., & Frary, M. (2021). Establishing a framework for assessing teaching effectiveness. *College Teaching*, 70(1), 1-18.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Wayne, A. J., Youngs, P., & McDonald, K. (2015). Teacher evaluation systems: Examining the evidence and exploring the future. *Review of Educational Research*, 85(4), 853-895.
- Zhao, G. (2020). Design and implementation of university teaching supervision and quality feedback system based on HTML5. *Computer Knowledge and Technology*, 16(1), 1-2.