International Journal of Evaluation and Research in Education (IJERE)
Vol. 15, No. 6, December 2026, pp. 5171~5186
ISSN: 2252-8822, DOI: 10.11591/ijere.v15i6.40286

A multidimensional instrument to measure teaching quality in blended learning environments in higher education

Quang X. Tran[1], Huong Tran Thi Thu[2], An T. Thuy Nguyen[2], Lam Le[3]
[1]Faculty of Teacher Education, VNU University of Education, Hanoi, Vietnam
[2]Faculty of Quality Management, VNU University of Education, Hanoi, Vietnam
[3]College Directorate, Dai Viet Sai Gon College, Ho Chi Minh, Vietnam

Article Info

Article history:

Received May 6, 2026
Revised Aug 6, 2026
Accepted Sep 18, 2026

Keywords:

Blended learning
Quality assurance
Scale development
Teaching quality
Validation

ABSTRACT

The rapid expansion of blended learning in higher education has created an urgent need for robust instruments to evaluate teaching quality in environments that integrate face-to-face and online instruction. However, existing approaches often focus on isolated aspects, such as learning outcomes or online components, and lack a comprehensive, multidimensional perspective. This study aims to develop and validate a multidimensional instrument for assessing teaching quality in blended learning environments in higher education. A sequential mixed-methods design was employed. First, relevant literature and existing frameworks were synthesized to construct an initial set of criteria and indicators. Expert consultation using the Delphi method was then conducted to establish content validity. Subsequently, survey data collected from lecturers and students at a university in Vietnam were analyzed to test reliability and construct validity. The findings resulted in a validated instrument comprising multiple dimensions reflecting context and objectives, input resources, teaching-learning processes, and learning outcomes, with strong evidence of reliability and validity. The proposed instrument provides a systematic and practical tool for evaluating teaching quality in blended learning and offers important implications for quality assurance and improvement in higher education.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Huong Tran Thi Thu
Faculty of Quality Management, VNU University of Education
144 Xuan Thuy Street, Cau Giay, Hanoi, Vietnam
Email: huong.tran@vnu.edu.vn

1. INTRODUCTION

The rapid advancement of digital technologies has fundamentally transformed higher education systems worldwide, reshaping teaching practices, learning environments, and institutional governance. In the context of the fourth industrial revolution, higher education institutions are increasingly required to adopt flexible, technology-enhanced learning models that support lifelong learning and the development of digital competencies. In Vietnam, digital transformation has been identified as a national priority, as reflected in policies such as decision No. 749/QĐ-TTg [1], which emphasizes the modernization of teaching and learning through digital platforms. Within this broader transformation, blended learning has emerged as a key pedagogical approach for enhancing flexibility, accessibility, and personalization in higher education. Recent studies have consistently identified blended learning as one of the dominant instructional models in post-pandemic higher education due to its flexibility, scalability, and ability to combine the strengths of face-to-face and online learning environments [2]–[4].

Blended learning is commonly defined as the purposeful integration of face-to-face and online instruction to optimize learning effectiveness [5]. Rather than representing a single instructional model, blended learning encompasses a spectrum of approaches that vary in the degree and manner of integration between online and in-person components. These approaches include models such as flipped classrooms, hybrid courses, and technology-enhanced learning environments, each emphasizing different configurations of interaction, content delivery, and learner engagement. Previous literature has shown that blended learning models vary considerably in terms of instructional integration, learner autonomy, technological mediation, and interaction patterns, reflecting the complexity and contextual diversity of blended learning implementation across higher education systems [6], [7]. Despite these variations, a shared characteristic across blended learning models is the reliance on digital technologies to extend learning beyond traditional classroom boundaries while maintaining essential elements of direct interaction.

A growing body of research has explored the effectiveness of blended learning in higher education. Empirical studies have demonstrated that blended learning can improve academic performance, learner engagement, and satisfaction compared to traditional teaching approaches [8], [9]. These positive outcomes include improvements in learner engagement, collaborative learning, academic achievement, and learner persistence, although the magnitude of effectiveness often depends on course design quality and institutional readiness [8]–[10]. Meta-analytic evidence suggests moderate to strong positive effects on learning outcomes, highlighting the potential of blended learning as an effective instructional approach. However, these outcomes are highly dependent on the quality of instructional design, implementation, and the interaction between pedagogical, technological, and contextual factors. As a result, understanding and evaluating teaching quality within blended learning environments has become an increasingly important area of inquiry.

Despite the rapid adoption of blended learning, many higher education institutions still lack systematic and validated instruments for evaluating teaching quality in blended learning environments. In practice, the absence of robust evaluation tools creates significant challenges for institutional quality assurance and instructional improvement. Universities may face difficulties in monitoring the consistency and effectiveness of blended teaching across courses and faculties, while lecturers often receive fragmented or overly general feedback that does not adequately support pedagogical improvement. Previous studies have similarly argued that quality assurance in blended learning remains fragmented because institutions often lack integrated evaluation frameworks capable of connecting instructional quality, learner support, technological infrastructure, and institutional management processes [11]–[13]. In addition, without standardized evaluation criteria, institutions encounter obstacles in developing coherent policies, benchmarking teaching practices, and implementing evidence-based quality assurance mechanisms. From the learners' perspective, inconsistent teaching quality across blended courses may lead to unequal learning experiences, reduced engagement, and disparities in learning outcomes. These challenges highlight the urgent need for a multidimensional and empirically validated instrument that can support systematic evaluation and continuous quality improvement in blended learning contexts.

To address the growing need for evaluating blended learning quality, several theoretical frameworks and evaluation models have been proposed in prior research. One of the most influential frameworks is the community of inquiry (CoI) model, which conceptualizes effective blended and online learning through the interaction of teaching presence, social presence, and cognitive presence [5]. The CoI framework has been widely applied to examine interaction quality and learner engagement in blended learning environments. Recent applications of the CoI framework in blended learning research further confirm its importance in explaining learner interaction and collaborative knowledge construction, particularly in language learning and digitally mediated instructional settings [14], [15]. However, although CoI provides a strong explanation of communication and learning presence, it focuses primarily on interactional processes and learner experiences, while giving limited attention to institutional conditions, technological infrastructure, policy support, and broader organizational factors influencing teaching quality.

Another widely adopted framework is the technological pedagogical content knowledge (TPACK) model [16], which emphasizes instructors' capability to integrate technology, pedagogy, and disciplinary knowledge effectively. TPACK is particularly valuable for understanding teacher readiness and digital pedagogical competence in technology-enhanced learning environments. Nevertheless, the model mainly concentrates on teacher knowledge integration rather than offering a comprehensive framework for evaluating teaching quality outcomes, learner support systems, instructional management, or institutional quality assurance processes. In blended learning environments, instructors' technological pedagogical competence has been identified as a critical success factor influencing instructional effectiveness, learner engagement, and technology integration quality [17].

From an evaluation perspective, the context–input–process–product (CIPP) model provides a more systemic approach by examining contextual conditions, resource inputs, implementation processes, and educational outcomes [18]. The CIPP model is especially useful for quality assurance and continuous

improvement because it connects evaluation with decision-making processes. However, although the model offers a comprehensive evaluation logic, it remains relatively general and requires further operationalization into measurable indicators specifically adapted to blended learning environments. Similar concerns have been raised in previous blended learning quality studies, which noted that broad evaluation frameworks often lack sufficiently operationalized indicators capable of supporting systematic quality measurement in real institutional settings [19], [20].

In addition to these theoretical models, several blended learning quality assurance frameworks have been developed, such as quality matters, blended learning quality assurance, and other online learning quality models [13], [21]. These frameworks identify important dimensions including course design, technological infrastructure, assessment, instructor competence, and learner support. Recent systematic reviews have also emphasized that quality assurance interventions in blended learning increasingly require evidence-based measurement instruments capable of supporting institutional decision-making and continuous improvement processes [13]. However, many existing frameworks either focus primarily on online learning components or emphasize institutional quality assurance procedures without sufficient empirical validation across multiple stages of instrument development. Moreover, most existing models have been developed in Western higher education contexts, while evidence from developing and transitional educational systems remains limited.

Although these frameworks were developed from different theoretical perspectives, they are highly complementary for evaluating teaching quality in blended learning environments. The CoI framework explains the quality of learning interactions through teaching, social, and cognitive presence. TPACK focuses on instructors' ability to integrate technological, pedagogical, and content knowledge effectively. The CIPP model provides a comprehensive evaluation logic by connecting contextual conditions, educational inputs, implementation processes, and learning outcomes, while the technology acceptance model (TAM) explains how users' perceptions of usefulness and ease of use influence technology adoption in educational settings. Considered together, these frameworks capture complementary dimensions of blended learning quality, ranging from pedagogical interaction and technological competence to institutional readiness, evaluation processes, learning outcomes, and technology acceptance. Therefore, integrating these frameworks provides a stronger conceptual foundation for developing a multidimensional instrument capable of evaluating teaching quality comprehensively rather than focusing on isolated aspects of blended learning.

Recent systematic reviews have further highlighted these limitations. Several studies [4], [22], [23] indicated that blended learning research remains fragmented across pedagogical, technological, and organizational dimensions. Furthermore, emerging technologies such as artificial intelligence are rapidly reshaping blended learning ecosystems, creating additional challenges for evaluating instructional quality, learner support, and digital engagement in higher education environments [24]. Existing studies frequently examine isolated aspects such as student engagement, interaction quality, or technological readiness, rather than developing multidimensional instruments capable of evaluating teaching quality comprehensively across blended learning ecosystems. Furthermore, rigorous validation procedures integrating expert consensus, pilot testing, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA) are still relatively rare in blended learning quality measurement research.

Consequently, several important gaps remain in the literature. First, although numerous studies have examined blended learning effectiveness, most existing evaluation approaches focus on isolated dimensions such as student satisfaction, online interaction, technological readiness, or learning outcomes, rather than capturing teaching quality as a multidimensional construct. Second, although existing theoretical models provide valuable perspectives on blended learning quality, they usually focus on specific dimensions such as interaction quality (CoI), teacher competence (TPACK), or systemic evaluation logic (CIPP)-without integrating these perspectives into a unified and operational measurement framework suitable for blended learning environments. Third, many quality assurance frameworks provide conceptual guidance but lack rigorous psychometric validation procedures across multiple stages of instrument development. Finally, empirical evidence from developing higher education contexts, particularly Vietnam and similar transitional educational systems, remains limited despite the rapid expansion of blended learning under national digital transformation policies. These gaps indicate the need for a multidimensional, contextually relevant, and empirically validated instrument capable of supporting both research and institutional quality assurance practices in blended learning environments.

Unlike previous studies that examined teaching quality from a single theoretical perspective, such as CoI, TPACK, CIPP, or TAM, the present study integrates these complementary frameworks into a unified multidimensional evaluation model. Furthermore, the proposed instrument was developed and validated through a rigorous sequential mixed-methods process, including literature synthesis, Delphi expert consultation, pilot testing, EFA, and CFA. This integrated conceptual and methodological approach represents the principal novelty of the study and provides a more comprehensive basis for evaluating teaching quality in blended learning environments.

In response to these challenges, this study aims to develop and validate a multidimensional instrument for assessing teaching quality in blended learning environments in higher education. The study is guided by three research questions:

- What dimensions define teaching quality in blended learning environments in higher education? (RQ1)
- How can a valid and reliable instrument be developed to measure these dimensions? (RQ2)
- To what extent is the proposed instrument applicable for evaluating teaching quality in practice? (RQ3)

To address these questions, the study synthesizes theoretical and empirical foundations, develops a structured set of criteria and indicators, and refines the instrument to ensure its relevance and applicability within a Vietnamese higher education context.

The principal novelty of this study lies in the development of the first multidimensional instrument that integrates CIPP, CoI, TPACK, and TAM into a unified evaluation framework for assessing teaching quality in blended learning environments. First, it develops a multidimensional instrument that integrates pedagogical, technological, organizational, and outcome-oriented dimensions of teaching quality within a unified evaluation framework. Second, the study synthesizes and operationalizes concepts from multiple theoretical perspectives, including CoI, TPACK, CIPP, and blended learning quality assurance models, thereby addressing the fragmentation observed in prior research. Third, the instrument was validated through a rigorous multi-stage process involving literature synthesis, Delphi expert consultation, pilot testing, EFA, and CFA, which strengthens both content and construct validity. Fourth, the study contributes empirical evidence from Vietnam, an emerging higher education context undergoing rapid digital transformation but still underrepresented in blended learning quality research. Finally, the validated instrument provides a practical tool that can support internal quality assurance (IQA) systems, accreditation activities, curriculum improvement, and evidence-based decision-making in higher education institutions.

Building on the theoretical foundations and identified research gaps, this study proposes a multidimensional evaluation framework for teaching quality in blended learning environments. The framework integrates key elements from existing models while providing a coherent and operational structure suitable for empirical measurement. Specifically, it consists of four interrelated dimensions: context and objectives, input resources, teaching-learning process, and learning outcomes. These dimensions collectively capture the complexity of teaching quality in blended learning and provide a systematic basis for developing measurable indicators. The conceptual integration described is operationalized into the multidimensional evaluation framework presented in Table 1. The framework translates the complementary theoretical perspectives into four interrelated dimensions that serve as the conceptual basis for instrument development.

Table 1. Proposed framework for evaluating teaching quality in blended learning environments

Dimension	Description	References
Context and objectives	Examines alignment between course outcomes, institutional goals, and learner needs, as well as contextual conditions for implementing blended learning.	[18], [25]
Input resources	Focuses on readiness of technological infrastructure, learning materials, instructor competencies, and institutional support.	[7], [18], [19], [26]
Teaching-learning process	Evaluates instructional delivery, interaction, communication, pedagogy, classroom management, and assessment practices.	[5], [18], [27]
Learning outcomes	Measures academic performance, learner satisfaction, retention, and improvement outcomes.	[18], [25]

2. RESEARCH METHOD

2.1. Research design

This study adopted a sequential mixed-methods design to develop and validate a multidimensional instrument for assessing teaching quality in blended learning environments in higher education. A sequential design was considered appropriate because the study aimed not only to conceptualize the dimensions of blended teaching quality but also to empirically validate the resulting measurement instrument. The research was conducted in two consecutive and interconnected phases to ensure both theoretical rigor and psychometric robustness.

The first phase employed qualitative methods, including literature synthesis, expert consultation, and the Delphi technique, to identify, construct, and refine the dimensions, criteria, and indicators of teaching quality in blended learning environments. This phase primarily addressed the first research question concerning the conceptual dimensions of teaching quality and contributed to establishing content validity and contextual appropriateness of the instrument. The second phase employed quantitative methods, including pilot testing, large-scale survey administration, and multivariate statistical analyses, to evaluate the

psychometric properties of the instrument. Specifically, reliability analysis, EFA, and CFA were conducted to examine internal consistency, construct validity, convergent validity, and discriminant validity. This phase addressed the second and third research questions by validating the measurement structure and examining the practical applicability of the instrument in higher education settings.

The integration of qualitative and quantitative phases allowed the study to combine theoretical foundations, expert judgment, and empirical evidence into a coherent scale-development process. Such an approach is widely recommended in educational measurement research because it strengthens both the conceptual relevance and statistical validity of newly developed instruments. To improve the transparency of the research design, each stage of the study was explicitly aligned with the corresponding research question. The qualitative phase, including literature synthesis and Delphi consultation, addressed RQ1 by identifying and refining the dimensions, criteria, and indicators of teaching quality in blended learning environments. The quantitative validation phase addressed RQ2 through reliability analysis, EFA, and CFA to evaluate the psychometric properties of the proposed instrument. Finally, RQ3 was addressed through the refinement of the validated instrument and the proposal of an implementation procedure demonstrating its applicability for evaluating teaching quality in blended learning practices. To illustrate the overall research design and validation procedure, the sequential mixed-methods process adopted in this study is summarized in Figure 1.

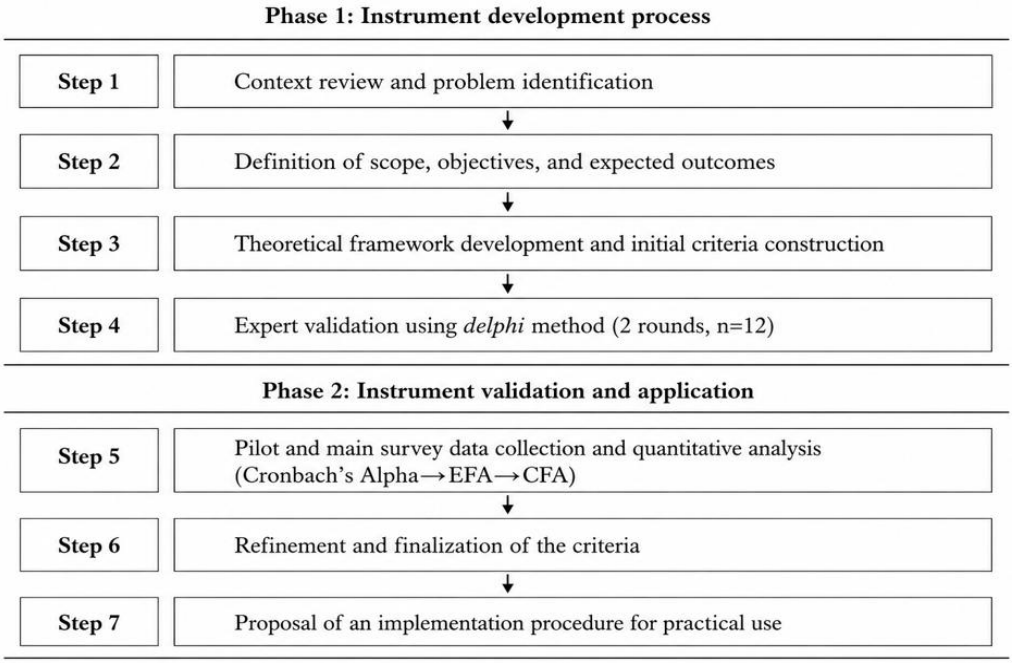


Figure 1. Sequential mixed-methods research design and validation procedure

2.2. Instrument development process

The development of the instrument was grounded in an extensive synthesis of prior literature and established evaluation frameworks, including the CIPP model and blended learning quality models. Based on this synthesis, an initial framework comprising four main dimensions was proposed: i) context and objectives; ii) input resources; iii) teaching-learning process; and iv) learning outcomes. Each dimension was operationalized into a set of criteria and measurable indicators. Initially, a total of 81 indicators were constructed to capture key aspects of teaching quality, including instructional design, technology integration, learner engagement, assessment practices, and institutional support. These indicators were subsequently refined through expert consultation and empirical testing.

2.3. Delphi study

The Delphi study was conducted as the second qualitative step to review and refine the preliminary criteria for evaluating teaching quality in blended learning environments. The Delphi method was selected because it is a widely used expert-consensus technique that allows researchers to collect, synthesize, and refine professional judgments through iterative rounds of consultation. In this study, it was used to examine the relevance, clarity, feasibility, and contextual appropriateness of the proposed standards, criteria, and indicators before conducting the large-scale quantitative survey.

The Delphi questionnaire consisted of three sections. The first section collected general information about the experts to describe their professional background and suitability for participation. The second section provided consultation instructions, including the purpose of the expert review, response procedures, and evaluation principles. The third section required experts to evaluate each criterion and indicator using a 5-point Likert scale, where 1=not appropriate and 5=highly appropriate. Experts were also invited to provide qualitative comments, including suggestions for rewording, adding, merging, or removing indicators.

A panel of 12 experts was invited to participate. All experts were university lecturers with experience in implementing blended learning. They were selected based on three main criteria: experience and competence in teaching blended courses, research or professional understanding of blended learning, and willingness to participate in multiple rounds of feedback until consensus was reached. To ensure objectivity, experts responded independently and anonymously, without access to one another's responses.

The expert panel size was considered appropriate for Delphi-based scale development studies, where panels ranging from 10 to 20 experts are commonly regarded as sufficient to achieve consensus while maintaining manageability of iterative feedback processes. The selected experts possessed diverse expertise. Specifically, the panel included specialists in educational quality assurance, blended learning pedagogy, instructional design, educational technology, educational measurement and evaluation, and higher education management. This diversity ensured that the proposed indicators were reviewed from complementary theoretical and practical perspectives, thereby enhancing the content validity of the instrument.

To ensure the quality and credibility of expert judgments, the study applied several selection criteria. Experts were required to: i) have at least five years of experience in higher education teaching or educational management; ii) possess direct experience in implementing or evaluating blended learning courses; iii) demonstrate research or professional expertise related to educational quality, blended learning, or assessment; and iv) express willingness to participate in multiple Delphi rounds.

In addition to quantitative ratings, qualitative comments provided by experts were analyzed using thematic synthesis. Researchers independently reviewed expert feedback, categorized suggestions into themes such as wording clarity, conceptual overlap, feasibility, contextual appropriateness, and indicator redundancy, and then compared interpretations through iterative discussion. Indicators were retained when experts demonstrated strong consensus regarding relevance and clarity. Indicators were revised when qualitative feedback suggested ambiguity or conceptual overlap, and indicators were removed when they were considered redundant, difficult to operationalize, or insufficiently relevant to blended learning quality evaluation. The complete Delphi consultation procedure adopted in this study is summarized in Table 2.

Table 2. Procedure for expert consultation using the Delphi method

Stage	Procedure	Output
Step 1	Develop the Delphi questionnaire based on the preliminary framework and indicators.	Initial expert consultation form
Step 2	Invite experts and provide guidance on the purpose, rating scale, and response principles.	Confirmed expert panel
Step 3	Conduct Delphi round 1: experts independently rate and comment on each criterion and indicator.	Initial ratings and qualitative feedback
Step 4	Analyze expert responses using statistical indicators and synthesize comments.	Revised criteria and indicators
Step 5	Conduct Delphi Round 2: experts re-evaluate the revised criteria and indicators.	Consensus-based evaluation results
Step 6	Apply decision rules to retain, revise, or remove indicators.	Refined criteria set for quantitative survey

As shown in Table 2, the Delphi process consisted of six sequential stages, beginning with questionnaire development and ending with the refinement of indicators for quantitative validation. To analyze the level of expert consensus, the study applied the knowledge acquisition for multiple experts with time scales (KAMET) principle proposed by Chu and Hwang [28]. This approach evaluates consensus for each factor or indicator $q_i$using statistical indicators such as the mean score M_{q_i}, median Md_{q_i}, quartile deviation Q_{q_i}, and variance V_{q_i}, where variance reflects the proportion of experts who changed their evaluation. The quartile deviation was used to assess dispersion in expert responses. It is calculated as:

$$Q_{q_i} = \frac{Q_3 - Q_1}{2}$$

where Q_3 and Q_1 are the third and first quartiles, respectively. A smaller quartile deviation indicates lower dispersion and stronger agreement among experts. In this study, lower Q_{q_i}values reflected higher consensus regarding the appropriateness of a criterion or indicator.

The decision rules were as: an indicator was accepted when $M_{q_i} \geq 3.5$, $Q_{q_i} \leq 0.5$, and $V_{q_i} < 15\%$. An indicator was rejected when $M_{q_i} < 3.5$, $Q_{q_i} \leq 0.5$, and $V_{q_i} \leq 15\%$, indicating expert agreement on its inadequacy. If consensus was not reached, further consultation would be required. In this study, after two Delphi rounds, expert responses reached a high level of consensus, and the preliminary instrument was considered suitable for subsequent large-scale survey validation, after necessary revisions.

The Delphi results led to the removal of several indicators. Indicator C.1.4 ("the level of agreement among business partners/students regarding the appropriateness of course objectives") and indicator C.2.2 ("budget allocated for digital content development") was also removed due to difficulties in data collection and evaluation. Indicators C.2.4 and P.4.2 were removed because their content overlapped with indicator I.2.4. Similarly, C.3.1 was removed because it overlapped with C.2.5, while C.3.4 was removed due to duplication with I.2.1. Indicator I.1.4 was excluded because it was conceptually similar to C.1.5, and P.2.5 was removed because it overlapped with P.2.1. These revisions were mainly intended to reduce redundancy and improve the applicability of the instrument in the Vietnamese higher education context.

After synthesizing expert feedback and revising the indicators, the instrument was refined into four standards, 16 criteria, and 73 indicators. The refined structure of the proposed instrument after Delphi validation is presented in Table 3. This refined structure provided the basis for developing the questionnaire used in the subsequent quantitative survey. Overall, the Delphi study strengthened the content validity of the instrument by ensuring that the criteria and indicators were theoretically relevant, practically feasible, and appropriate for evaluating teaching quality in blended learning environments.

Table 3. Detailed criteria and indicators after Delphi validation

Standard	Criteria	Number of indicators
Context and objectives	1.1. Alignment with program objectives and learning outcomes	4
(4 criteria and 15 indicators)	1.2. Institutional policies and support	4
	1.3. Readiness and needs of stakeholders	3
	1.4. Cultural and socio-economic relevance	4
Input resources	2.1. Curriculum and instructional materials design	5
(4 criteria and 20 indicators)	2.2. Technological infrastructure and tools	5
	2.3. Instructor competencies	5
	2.4. Student support services	5
Teaching–learning process	3.1. Instructional organization and learner engagement	5
(4 criteria and 18 indicators)	3.2. Interaction and collaboration	4
	3.3. Assessment and feedback	5
	3.4. Classroom management and adaptability	4
Learning outcomes	4.1. Learning outcomes and academic achievement	5
(4 criteria and 20 indicators)	4.2. Skills development and application	5
	4.3. Learner satisfaction and feedback	5
	4.4. Course completion and learning retention	5

2.4. Survey design and data collection

2.4.1. Sampling and participants

Participants were selected using purposive and convenience sampling techniques based on their direct experience with blended learning environments. To ensure the relevance of responses, participants were required to meet at least one of the following criteria: i) students who had completed at least one blended learning course within the previous academic year; or ii) lecturers who had experience teaching blended courses using both face-to-face and online instructional components.

The study involved both pilot and official survey stages. In the official survey, a total of 620 responses were initially collected through an online questionnaire distributed via institutional learning platforms and academic communication channels. After data screening and cleaning, 472 valid responses were retained for analysis. Invalid responses were excluded due to incomplete answers or straight-lining response patterns. The final sample included 432 students (91.5%) and 40 lecturers (8.5%). Student participants represented multiple academic disciplines, including education, social sciences, information technology, economics, and language studies. The sample also included students from different year levels to ensure variation in blended learning experiences. Lecturer participants varied in gender, teaching experience, and years of involvement in blended learning implementation. The demographic characteristics of the official survey sample are reported in Table 4.

2.4.2. The validation process

In the validation process, the study employed a two-stage survey design. First, a pilot survey was conducted to assess the reliability of the instrument and identify problematic items. The pilot sample included participants from different academic programs and years of study, ensuring initial diversity. Second,

the refined survey was distributed to a large-scale sample to validate. Building on the Delphi-based refinement, a pilot study was conducted to preliminarily evaluate the reliability and structural validity of the instrument prior to large-scale implementation. The questionnaire was developed based on the refined set of criteria and indicators, comprising five sections: respondent background information and four core dimensions, namely context and objectives, input resources, teaching-learning process, and learning outcomes. All items were measured using a five-point Likert scale ranging from 1 (strongly disagree/not appropriate) to 5 (strongly agree/highly appropriate), and a small number of open-ended questions were included to capture additional qualitative feedback.

The pilot survey was administered online via Google Forms, allowing respondents to complete the questionnaire conveniently using digital devices. A total of 160 responses were collected, with an average completion time of approximately 11–15 minutes per respondent. After data cleaning, which involved removing incomplete or inconsistent responses, 115 valid responses (n=115) were retained for analysis. This sample size is considered adequate for preliminary scale validation using EFA [29]. The data analysis followed a sequential procedure including Cronbach's alpha reliability testing and EFA. The results indicated that all 16 initial criteria achieved satisfactory internal consistency, with Cronbach's alpha values ranging from 0.822 to 0.919, and corrected item–total correlations (CITC) exceeding acceptable thresholds. During the pilot validation, a total of 11 items were removed, reducing the instrument from 73 to 62 observed variables. Specifically, two items (Q_C3_5 and Q_I4_3) were first excluded during the reliability analysis because their removal improved the internal consistency of the corresponding scales. Subsequent EFA resulted in the removal of nine additional items: Q_P2_2 and Q_P3_4 due to insufficient factor loadings and/or cross-loading patterns, and Q_Pt3_1, Q_Pt3_2, Q_Pt3_3, Q_Pt3_4, Q_Pt3_5, Q_Pt4_2, and Q_Pt4_3 because of their weak factor convergence, low communalities or factor loadings, cross-loadings, and/or inadequate conceptual alignment with the extracted factors. While the original framework proposed 16 criteria, the factor analysis revealed a more parsimonious structure consisting of nine component factors across the four main standards. The Kaiser-Meyer-Olkin (KMO) values ranged from 0.892 to 0.934, and Bartlett's test was statistically significant ($p < 0.001$), indicating that the data were suitable for factor analysis. The extracted factors explained a substantial proportion of variance (above 50% across all dimensions), demonstrating acceptable construct validity at the exploratory stage. At this stage, nine items were deleted including Q_P2_2, Q_P3_4, Q_Pt3_1, Q_Pt3_2, Q_Pt3_3, Q_Pt3_4, Q_Pt3_5, Q_Pt4_2, Q_Pt4_3.

Importantly, the pilot study revealed that several theoretical sub-dimensions converged into broader empirical factors. For example, in the "input resources" dimension, indicators related to instructional materials and technological infrastructure merged into a single factor reflecting a unified learning ecosystem. Similarly, within the "teaching-learning process" dimension, instructional organization and interaction-related indicators converged into a combined experiential construct. These findings highlight the need to align theoretical models with users' perceptions in blended learning contexts.

Consequently, the refined instrument comprised four overarching standards, nine empirically derived factors, and 62 observed variables, ensuring improved clarity, reduced redundancy, and stronger construct validity. The refined instrument provided a robust foundation for the subsequent large-scale survey and CFA. Overall, the pilot study played a critical role in validating and optimizing the measurement instrument, ensuring that it met established standards of reliability and validity in educational measurement research [14].

2.5. Data analysis

Data were analyzed using SPSS 26.0 and AMOS 24.0 following established procedures in psychometric scale development and validation. Prior to analysis, the dataset was screened for missing values, response inconsistency, and straight-lining behavior. Descriptive statistics, standard deviations, skewness, and kurtosis values were examined to assess data distribution and suitability for multivariate analysis. The assumptions for factor analysis were evaluated using the KMO measure of sampling adequacy and Bartlett's test of sphericity.

Internal consistency reliability was first examined using Cronbach's alpha and CITC. Cronbach's alpha values of 0.70 or above were considered acceptable, while CITC values below 0.30 were considered for item removal. Indicators that reduced overall scale reliability or demonstrated weak contribution to their corresponding construct were excluded from subsequent analyses.

The EFA was then conducted to identify the latent structure of the instrument and reduce overlapping indicators. The study employed principal axis factoring as the extraction method because it is appropriate for identifying latent constructs underlying observed variables. Promax rotation was applied because the dimensions of teaching quality were theoretically expected to correlate with one another. Indicators with factor loadings below 0.50, substantial cross-loadings, or conceptual inconsistency were considered for removal. The retained factor solution was evaluated based on interpretability, theoretical coherence, and total variance explained.

Following EFA, CFA was conducted using AMOS 24.0 to validate the measurement model and assess construct validity. Multiple goodness-of-fit indices were examined, including Chi-square/df (CMIN/df), comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and root mean square residual (RMR). Model fit was considered acceptable when CMIN/df <5, CFI and TLI>0.90, and RMSEA<0.08.

Convergent validity was assessed using standardized factor loadings, composite reliability (CR), and average variance extracted (AVE). Standardized factor loadings above 0.50, CR values above 0.70, and AVE values above 0.50 were considered acceptable indicators of convergent validity. Discriminant validity was evaluated using the Fornell–Larcker criterion, whereby the square root of AVE for each construct should exceed its correlations with other constructs. Together, these procedures ensured that the final instrument satisfied accepted standards of reliability and construct validity in educational measurement research.

3. RESULTS AND DISCUSSION

3.1. Description of the official survey sample

The official survey initially yielded 620 responses. After data screening, responses with incomplete information or straight-lining patterns were excluded. Straight-lining responses were identified by a standard deviation of zero across Likert-scale items, as such responses may distort inter-item correlations and threaten the validity of subsequent factor analyses. After data cleaning, 472 valid responses were retained for quantitative analysis.

As shown in Table 4, the final sample consisted of 432 students (91.5%) and 40 lecturers (8.5%). The student sample represented four academic years and eleven academic disciplines, indicating diversity in learning experience and academic background. Lecturer participants also varied in gender and teaching experience; notably, 52.5% of lecturers had more than 10 years of teaching experience. Although students accounted for the majority of the sample, this distribution is consistent with the learner-centered orientation of the instrument, which emphasizes student perceptions of teaching quality in blended learning environments.

Table 4. Demographic characteristics of the official survey sample

Characteristic	Category	Frequency (n)	Percentage (%)
Participant type	Students	432	91.5
	Lecturers	40	8.5
	Total	472	100.0
Student year level	Year 1	123	28.5
	Year 2	111	25.7
	Year 3	106	24.5
	Year 4	92	21.3
	Total students	432	100.0
Academic discipline	Psychology	32	7.4
	Early childhood education	33	7.6
	School counseling	35	8.1
	Primary education	36	8.3
	Mathematics education	37	8.6
	School administration	39	9.0
	Literature education	39	9.0
	Educational sciences	42	9.7
	History-geography education	43	10.0
	Educational technology management	45	10.4
	Educational quality management	51	11.8
	Total students	432	100.0
Lecturer gender	Male	8	20.0
	Female	32	80.0
	Total lecturers	40	100.0
Lecturer teaching experience	Less than 5 years	6	15.0
	5–10 years	13	32.5
	More than 10 years	21	52.5
	Total lecturers	40	100.0

3.2. Reliability analysis

Reliability analysis was conducted using Cronbach's alpha and CITC. As presented in Table 5, all nine component scales achieved high internal consistency, with Cronbach's alpha values ranging from 0.876 to 0.948. The CITC values ranged from 0.663 to 0.833, exceeding the commonly accepted threshold of 0.30. No indicator was removed during this stage because all items contributed positively to their

corresponding component scales. These results indicate that the instrument has strong internal consistency. The high reliability values also suggest that the refinement process conducted after the pilot study successfully reduced redundancy while preserving conceptual coherence among indicators.

Table 5. Reliability analysis of the nine component scales

Dimension	Component scale	Number of indicators	Cronbach's alpha	CITC range
Context and objectives	CT1. Course design aligned with program standards and practice	4	0.900	0.745-0.795
	CT2. Access conditions and blended learning implementation	10	0.946	0.663-0.827
Input resources	IN1. Digital learning materials and technological infrastructure	10	0.940	0.707-0.796
	IN2. Instructor teaching competence	5	0.918	0.757-0.816
	IN3. Learner support	4	0.876	0.705-0.757
Teaching–learning process	PS1. Assessment, feedback, and classroom management	8	0.942	0.758-0.821
	PS2. Instructional organization, interaction, and collaboration	8	0.948	0.765-0.833
Learning outcomes	PT1. Learning effectiveness and retention	8	0.945	0.782-0.827
	PT2. Competence development and value transfer	5	0.917	0.738-0.813

3.3. Exploratory factor analysis

The EFA was conducted separately for the four main dimensions to examine the empirical structure of the instrument. The KMO values ranged from 0.957 to 0.970, indicating excellent sampling adequacy. Bartlett's test of sphericity was statistically significant for all dimensions ($p<0.001$), confirming that the correlation matrices were suitable for factor analysis. As summarized in Table 6, EFA replicated the structure refined from the pilot study. The four dimensions produced a 2–3–2–2 factor pattern, corresponding to nine component scales. The total variance explained ranged from 64.836% to 68.554%, exceeding the commonly recommended threshold of 50%. Importantly, no observed indicators were removed at this stage, and all 62 indicators were retained for CFA.

Table 6. Summary of EFA results

Dimension	Number of retained indicators	Number of extracted factors	KMO	Bartlett's test	Total variance explained (%)	Decision
Context and objectives	14	2	0.962	p<0.001	65.767	Retained
Input resources	19	3	0.962	p<0.001	64.836	Retained
Teaching–learning process	16	2	0.970	p<0.001	68.554	Retained
Learning outcomes	13	2	0.957	p<0.001	68.513	Retained

The EFA results also provided an important empirical basis for combining several theoretically separate criteria into broader component scales. In the input resources dimension, indicators related to digital learning materials and technological infrastructure converged into one factor. This suggests that respondents perceived learning materials, the learning management system (LMS), digital tools, and connectivity as elements of a unified digital learning ecosystem rather than as separate input conditions. Similarly, within the teaching-learning process dimension, indicators concerning assessment, feedback, and classroom management formed a coherent component scale because these activities jointly represent the mechanism through which blended courses are monitored, regulated, and improved during implementation. In the learning outcomes dimension, learning effectiveness and course retention were grouped together, reflecting the close relationship between perceived learning achievement and students' ability to complete and sustain participation in blended courses. These findings indicate that learners tend to experience blended learning holistically, where technological, pedagogical, and managerial elements are interdependent rather than isolated. The EFA summary demonstrates that the factor structure was stable and theoretically interpretable. All retained indicators had factor loadings above the minimum threshold of 0.50 and no substantial cross-loading.

3.4. Confirmatory factor analysis

The CFA was conducted to verify the measurement model identified through EFA. Table 7 presents the CFA results for the four dimensions. Overall, the measurement models achieved acceptable to good fit. The CMIN/df values ranged from 2.230 to 3.118, RMSEA values ranged from 0.051 to 0.067, and CFI values ranged from 0.958 to 0.980. Other fit indices, including normed fit index (NFI), relative fit index (RFI), incremental fit index (IFI), and TLI, also reached acceptable or good levels, while RMR values were small. These results confirm that the measurement model adequately fits the empirical data. The CFA results

support the multidimensional structure of teaching quality in blended learning environments and provide evidence that the instrument is appropriate for evaluating teaching quality across the four proposed dimensions.

Table 7. CFA results by dimension

Fit index	Context and objectives	Input resources	Teaching-learning process	Learning outcomes
CMIN	194.877	431.580	229.683	199.553
df	76	149	103	64
CMIN/df	2.564	2.897	2.230	3.118
RMSEA	0.058	0.063	0.051	0.067
CFI	0.977	0.958	0.980	0.973
NFI	0.962	0.938	0.965	0.961
RFI	0.955	0.928	0.959	0.952
IFI	0.977	0.958	0.981	0.973
TLI	0.972	0.952	0.977	0.967
RMR	0.016	0.021	0.012	0.013

3.5. Composite reliability, convergent validity, and discriminant validity

CR and AVE were examined to assess construct reliability and convergent validity. As shown in Table 8, CR values ranged from 0.881 to 0.949, exceeding the recommended threshold of 0.70. AVE values ranged from 0.614 to 0.698, exceeding the recommended threshold of 0.50. These results indicate that the component scales demonstrate adequate convergent validity and that the indicators represent their intended constructs.

Discriminant validity was assessed using the Fornell-Larcker criterion. As shown in Table 9, the square root of AVE for each component scale was generally higher than its correlations with other component scales. Some correlations were relatively high, particularly among conceptually related constructs within the same evaluation domain. However, these correlations are theoretically explainable because blended learning quality is inherently systemic, and dimensions such as technology readiness, learner support, assessment, feedback, and instructional interaction are expected to be interrelated.

Table 8. CR and AVE

Dimension	Component scale	CR	AVE	√AVE
Context and objectives	CT1	0.900	0.693	0.832
	CT2	0.947	0.640	0.800
Input resources	IN1	0.941	0.614	0.784
	IN2	0.920	0.696	0.834
	IN3	0.881	0.649	0.806
Teaching–learning process	PS1	0.942	0.671	0.819
	PS2	0.949	0.698	0.835
Learning outcomes	PT1	0.945	0.681	0.825
	PT2	0.917	0.688	0.829

Table 9. Discriminant validity matrix

Construct	CT1	CT2	IN1	IN2	IN3	PS1	PS2	PT1	PT2
CT1	0.832								
CT2	0.721	0.800							
IN1	0.645	0.688	0.784						
IN2	0.620	0.578	0.694	0.834					
IN3	0.563	0.622	0.719	0.678	0.806				
PS1	0.663	0.642	0.768	0.773	0.748	0.819			
PS2	0.575	0.609	0.704	0.650	0.663	0.759	0.835		
PT1	0.559	0.561	0.675	0.620	0.622	0.703	0.634	0.825	
PT2	0.528	0.551	0.676	0.612	0.614	0.680	0.685	0.732	0.829

Note: diagonal values represent the square root of AVE.

3.6. Final validated instrument

Based on the results of reliability analysis, EFA, and CFA, the final instrument consisted of four dimensions, nine component scales, and 62 indicators. Table 10 summarizes the final structure of the validated instrument. Overall, the validation results demonstrate that the instrument satisfies key psychometric requirements. Reliability analysis confirmed strong internal consistency; EFA replicated the refined factor structure and retained all indicators; and CFA confirmed the adequacy of the measurement model. The final instrument therefore provides a reliable and valid basis for evaluating teaching quality in

blended learning environments. Its multidimensional structure enables higher education institutions to report evaluation results at multiple levels, including overall teaching quality, four major dimensions, nine component scales, and individual indicators. This structure is particularly useful for identifying strengths and weaknesses, prioritizing quality improvement areas, and supporting evidence-based decision-making in IQA systems.

Table 10. Final structure of the validated instrument

Dimension	Component scale	Number of indicators
Context and objectives	CT1. Course design aligned with program standards and practice	4
	CT2. Access conditions and blended learning implementation	10
Input resources	IN1. Digital learning materials and technological infrastructure	10
	IN2. Instructor teaching competence	5
	IN3. Learner support	4
Teaching-learning process	PS1. Assessment, feedback, and classroom management	8
	PS2. Instructional organization, interaction, and collaboration	8
Learning outcomes	PT1. Learning effectiveness and retention	8
	PT2. Competence development and value transfer	5
Total	Nine component scales	62

3.7. Discussion

The findings of this study provide strong evidence that teaching quality in blended learning environments should be understood as a multidimensional and systemic construct rather than as a collection of isolated teaching activities. The validated instrument confirmed that teaching quality in blended learning encompasses interconnected dimensions related to contextual alignment, resource readiness, instructional implementation, and learning outcomes. This result supports the broader argument in blended learning research that effective teaching quality emerges from the interaction between pedagogy, technology, institutional conditions, and learner experience rather than from any single component alone [4], [22].

One important contribution of the study lies in the empirical convergence of theoretically distinct criteria into broader experiential component scales during EFA. For example, indicators associated with digital learning materials and technological infrastructure converged into a single component scale, suggesting that learners perceive these elements as part of an integrated digital learning ecosystem. Similarly, assessment, feedback, and classroom management were empirically grouped together, indicating that students experience these processes as interconnected mechanisms through which blended learning courses are organized, regulated, and supported. In the learning outcomes dimension, academic achievement and learning retention also formed a coherent construct, reflecting the close relationship between successful learning experiences and sustained participation in blended courses. These findings suggest that students do not experience blended learning through isolated pedagogical or technological dimensions, but rather through holistic and integrated learning experiences. Similar patterns were identified by Han and Ellis [15], who found that students tend to perceive blended learning quality through integrated learning experiences that combine interaction, collaboration, learning support, and technological usability.

This holistic interpretation extends previous blended learning frameworks in several ways. In particular, the present findings support recent reviews emphasizing that learner interaction and teaching presence remain central to successful blended learning implementation, but they also suggest that these dimensions must be interpreted alongside institutional and technological conditions [14], [22]. The findings are consistent with the CoI framework, which emphasizes the interconnected roles of teaching presence, social presence, and cognitive presence in meaningful learning experiences [5]. However, the present study broadens this perspective by incorporating organizational readiness, technological infrastructure, learner support, and outcome dimensions into a unified evaluation structure. Likewise, while the TPACK framework highlights instructors' technological pedagogical competence [16], the current study demonstrates that teaching competence alone is insufficient for explaining teaching quality in blended environments unless accompanied by institutional support systems, accessible technological conditions, and effective learning management mechanisms.

The findings also reinforce the systemic logic of the CIPP model. Contextual conditions, resource inputs, implementation processes, and learning outcomes were shown to be empirically interconnected, supporting the view that teaching quality should be evaluated as part of a continuous and dynamic educational system rather than through isolated outcome indicators. This finding aligns with previous conceptual discussions suggesting that blended learning quality emerges from the interaction between pedagogical design, learner engagement, institutional readiness, and technological mediation rather than from isolated instructional variables [12], [30]. At the same time, the study operationalizes the CIPP framework

into measurable indicators specifically adapted for blended learning environments, thereby addressing limitations identified in previous research concerning the lack of validated multidimensional measurement instruments [12], [13].

Another important implication concerns the contextual contribution of the study within Vietnamese higher education. The findings highlight the critical role of technological infrastructure, institutional policy support, digital readiness, and learner support systems in shaping perceptions of teaching quality. In rapidly digitalizing higher education systems such as Vietnam, blended learning quality cannot be separated from broader issues of technological access, institutional implementation capacity, and digital competence development. This observation aligns with previous studies indicating that blended learning effectiveness in developing and transitional contexts is strongly influenced by institutional readiness and infrastructural conditions [23]. Similar conclusions were reported in recent global reviews showing that blended learning implementation in Global South and transitional educational systems is often constrained by infrastructure inequality, digital competence gaps, and uneven institutional support [23]. The strong convergence between infrastructure-related indicators and learning process indicators in the present study further suggests that students perceive technology not merely as a support tool, but as an inseparable component of the learning environment itself. This observation is consistent with the TAM [31], which argues that perceptions of usefulness and ease of use strongly influence users' acceptance of technology-mediated learning environments.

The study also offers several practical implications for higher education institutions. First, the validated instrument can support IQA systems by providing a structured mechanism for evaluating teaching quality in blended learning environments. Universities may use the instrument at multiple levels, including course evaluation, program review, faculty-level benchmarking, and institutional monitoring. The multidimensional structure of the instrument is also consistent with recent recommendations calling for integrated quality assurance systems capable of supporting benchmarking, accreditation, curriculum redesign, and data-driven instructional improvement in blended learning environments [13], [21]. Second, the instrument can support evidence-based curriculum improvement by helping institutions identify strengths and weaknesses across contextual conditions, instructional resources, learning processes, and outcomes. Third, the instrument provides a useful framework for lecturer professional development, particularly in areas related to digital pedagogy, learner engagement, assessment strategies, and online classroom management.

In addition, the multidimensional structure of the instrument makes it applicable for accreditation and quality audit activities. Higher education institutions may use the instrument to generate systematic evidence for program evaluation, blended learning policy development, and continuous improvement initiatives. The instrument may also support benchmarking across departments or faculties by enabling comparative analysis of blended teaching quality indicators over time. Furthermore, because the instrument captures dimensions related to technological readiness and learner support, it can help institutions monitor their preparedness for digital transformation and identify priority areas for strategic investment.

Overall, the findings indicate that evaluating teaching quality in blended learning requires moving beyond narrow measures of student satisfaction or academic outcomes toward more integrated approaches that capture the complexity of blended learning ecosystems. By combining theoretical integration with empirical validation, this study contributes to the growing body of research seeking to conceptualize and operationalize quality in digitally mediated higher education environments. Overall, the findings reinforce the growing consensus that blended learning quality evaluation requires integrated, context-sensitive, and empirically validated approaches capable of capturing the complexity of contemporary digital higher education ecosystems [4], [22]. Beyond Vietnam, the findings may also be relevant to other developing higher education systems undergoing rapid digital transformation. Many universities in these contexts face similar challenges, including unequal technological access, variations in digital competence, limited institutional support, and evolving quality assurance mechanisms. Consequently, the multidimensional evaluation framework proposed in this study may provide a useful reference for institutions seeking to establish evidence-based approaches for monitoring and improving blended learning quality under comparable educational conditions.

4. CONCLUSION

This study developed and validated a multidimensional instrument for evaluating teaching quality in blended learning environments in higher education. Drawing upon the integration of major theoretical and quality assurance frameworks, including CIPP, CoI, TPACK, TAM, and blended learning quality models, the study proposed a comprehensive framework consisting of four dimensions, nine component scales, and 62 indicators. The findings confirm that teaching quality in blended learning should be conceptualized as a multidimensional construct shaped by the interaction between contextual conditions, institutional and technological resources, instructional processes, and learning outcomes.

Methodologically, the study contributes to educational measurement research by applying a rigorous sequential mixed-methods design that combines Delphi expert consultation with psychometric validation procedures, including pilot testing, EFA, and CFA. This approach strengthens both the conceptual relevance and empirical robustness of the developed instrument. The study therefore provides an example of how blended learning quality frameworks can be operationalized into measurable and validated evaluation tools.

Practically, the validated instrument offers higher education institutions a feasible tool for supporting IQA, course evaluation, curriculum improvement, lecturer professional development, and accreditation-related activities. The instrument may also support benchmarking across academic units and monitoring institutional readiness for digital transformation initiatives. Because the framework captures both pedagogical and organizational dimensions of blended learning, it may help institutions adopt more evidence-based approaches to improving teaching quality in digitally mediated learning environments.

Despite these contributions, the study has several limitations. First, the official sample was heavily dominated by student respondents, while lecturer participation remained relatively limited, which may have resulted in greater emphasis on learner perceptions. Second, the study was conducted within a single higher education institution, limiting the generalizability of the findings to other institutional contexts. Third, the study primarily relied on self-reported survey data, which may be influenced by respondent perception and contextual bias.

Future research should therefore expand the validation process across multiple universities and institutional contexts to examine the broader applicability of the instrument. Increasing lecturer participation would also allow more balanced comparison between student and instructor perspectives. In addition, future studies may conduct measurement invariance testing across participant groups and institutional settings to examine the stability of the measurement structure. Longitudinal studies are also recommended to investigate how teaching quality evolves over time in blended learning environments. Finally, integrating survey-based evaluation data with learning analytics, LMS behavioral data, and actual learning outcomes may provide a more comprehensive understanding of teaching quality and support data-driven quality improvement in higher education.

ACKNOWLEDGMENTS

We would like to sincerely thank the students at Viet Nam National University, Hanoi and the scholars for their participating in validating the instrument and collecting data process of the research.

FUNDING INFORMATION

This work was funded by the Vietnam National University, Hanoi through the research project number QG.23.46.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Quang X. Tran	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	
Huong Tran Thi Thu			✓	✓	✓		✓	✓	✓	✓	✓			
An T. Thuy Nguyen			✓	✓	✓	✓	✓			✓				
Lam Le	✓	✓					✓		✓			✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [QXT], upon reasonable request.

REFERENCES

[1] Prime Minister of Vietnam, *Decision No. 749/QĐ-TTg: approving the national digital transformation program until 2025, with a vision towards 2030*. Hanoi, Vietnam, 2020. [Online]. Available: https://vanban.chinhphu.vn/?docid=200163&pageid=27160.

[2] A. H. Maarop and M. A. Embi, "Implementation of blended learning in higher learning institutions: a review of the literature," *International Education Studies*, vol. 9, no. 3, pp. 41–52, Feb. 2016, doi: 10.5539/ies.v9n3p41.

[3] M. Thahir, W. Widiawati, and N. Baitillah, "The post-pandemic education: a blended learning approach for teaching and learning in higher education in the new normal era," *International Journal of Ethno-Sciences and Education Research*, vol. 3, no. 3, pp. 99–108, Jul. 2023, doi: 10.46336/ijeer.v3i3.461.

[4] S. McCarthy and E. Palmer, "Defining an effective approach to blended learning in higher education: a systematic review," *Australasian Journal of Educational Technology*, vol. 39, no. 2, pp. 98–114, Jul. 2023, doi: 10.14742/ajet.8489.

[5] D. R. Garrison and N. D. Vaughan, *Blended learning in higher education: framework, principles, and guidelines*. San Francisco, CA: John Wiley & Sons, 2008, doi: 10.1002/9781118269558.

[6] R. Boelens, B. de Wever, and M. Voet, "Four key challenges to the design of blended learning: a systematic literature review," *Educational Research Review*, vol. 22, pp. 1–18, Nov. 2017, doi: 10.1016/j.edurev.2017.06.001.

[7] B. Amarathunga, "Blended learning trends and future directions: a systematic literature review and bibliometric analysis," *International Journal of Information and Learning Technology*, vol. 42, no. 2, pp. 147–164, Mar. 2025, doi: 10.1108/IJILT-04-2024-0072.

[8] B. Means, Y. Toyama, R. Murphy, and M. Baki, "The effectiveness of online and blended learning: a meta-analysis of the empirical literature," *Teachers College Record*, vol. 115, no. 3, pp. 1–47, Mar. 2013, doi: 10.1177/016146811311500307.

[9] H. M. Vo, C. Zhu, and N. A. Diep, "The effect of blended learning on student performance at course-level in higher education: a meta-analysis," *Studies in Educational Evaluation*, vol. 53, pp. 17–28, Jun. 2017, doi: 10.1016/j.stueduc.2017.01.002.

[10] S. Kaur, "Virtual learning and blended learning: a comparative analysis of effectiveness, engagement, and outcomes," *International Journal of Leading Research Publication*, vol. 7, no. 4, Apr. 2026, doi: 10.70528/IJLRP.v7.i4.2053.

[11] P. Mazohl and H. Makl, "Quality assurance in blended learning: A quality framework," in *Blended Learning Quality – Concepts Optimized for Adult Education*, Mazohl Publish, 2016, pp. 21–35.

[12] Y. Blieck *et al.*, "A conceptual model for effective quality management of online and blended learning," *Electronic Journal of e-Learning*, vol. 18, no. 2, pp. 189–204, Feb. 2020, doi: 10.34190/EJEL.20.18.2.007.

[13] M. Kayyali, "Quality assurance in online and blended learning environments," in *Quality Assurance and Accreditation in Higher Education: Issues, Models, and Best Practices*, M. Kayyali, Ed., Cham, Switzerland: Springer Nature Switzerland, 2024, pp. 227–292, doi: 10.1007/978-3-031-66623-0_5.

[14] X. LI, N. A. Sulaiman, and A. A. Aziz, "A systematic review of blended learning in higher education: Second language acquisition through the community of inquiry framework," *International Journal of Learning, Teaching and Educational Research*, vol. 23, no. 10, pp. 226–251, Oct. 2024, doi: 10.26803/ijlter.23.10.11.

[15] F. Han and R. A. Ellis, "Initial development and validation of the perceptions of the blended learning environment questionnaire," *Journal of Psychoeducational Assessment*, vol. 38, no. 2, pp. 168–181, Apr. 2020, doi: 10.1177/0734282919834091.

[16] P. Mishra and M. J. Koehler, "Technological pedagogical content knowledge: a framework for teacher knowledge," *Teachers College Record*, vol. 108, no. 6, pp. 1017–1054, Jun. 2006, doi: 10.1111/j.1467-9620.2006.00684.x.

[17] M. M. Ergo, B. M. Yimer, and A. T. Bishaw, "TPACK implementation in higher education: an assessment of teacher educators' competence," *Research in Learning Technology*, vol. 33, pp. 1–14, Oct. 2025, doi: 10.25304/rlt.v33.3481.

[18] D. L. Stufflebeam and C. L. S. Coryn, *Evaluation theory, models, and applications*, 2nd ed. San Francisco, CA, USA: Jossey-Bass, 2014.

[19] B. M. Gutiérrez-Pérez and A. V. Martín-García, "Psychometric validity of a tool for assessing the educational quality of blended learning systems," *Journal of Educators Online*, vol. 20, no. 4, pp. 1–15, Sep. 2023, doi: 10.9743/JEO.2023.20.4.10.

[20] B. M. Gutiérrez-Pérez and A. V. Martín-García, "Evaluation of quality in blended learning training," in *Blended learning: Convergence between technology and pedagogy*, B. M. Gutiérrez-Pérez and A. V. Martín-García, Cham: Springer International Publishing, 2020, pp. 91–111, doi: 10.1007/978-3-030-45781-5_5.

[21] R. Panigrahi, K. L. Nihar, and N. Singh, "Quality measurement of blended learning model in higher education: scale development and validation," *Higher Learning Research Communications*, vol. 14, no. 1, pp. 86–102, Apr. 2024, doi: 10.18870/hlrc.v14i1.1467.

[22] M. A. Ashraf *et al.*, "A systematic review of systematic reviews on blended learning: trends, gaps and future directions," *Psychology Research and Behavior Management*, vol. 14, pp. 1525–1541, Oct. 2021, doi: 10.2147/PRBM.S331741.

[23] N. Shaya, R. AbuKhait, R. Madani, and V. Ahmed, "Conceptualizing blended learning models as a sustainable and inclusive educational approach: an organizational dynamics perspective," *International Journal of Sustainability in Higher Education*, vol. 26, no. 9, pp. 90–111, Dec. 2025, doi: 10.1108/IJSHE-03-2024-0167.

[24] D. Dang, N. H. S. Alani, W. Nayani, and E. Erturk, "Blended learning design in higher education: a systematic review through TPACK and AI role perspectives (2020–2025)," *Education Sciences*, vol. 16, no. 6, p. 848, May 2026, doi: 10.3390/educsci16060848.

[25] L. Wong, A. Tatnall, and S. Burgess, "A framework for investigating blended learning effectiveness," *Education + Training*, vol. 56, no. 2/3, pp. 233–251, Apr. 2014, doi: 10.1108/ET-04-2013-0049.

[26] B. H. Khan, *Managing e-learning strategies: design, delivery, implementation and evaluation*. Hershey, PA: IGI Global Scientific Publishing, 2005, doi: 10.4018/978-1-59140-634-1.

[27] L. Pham, S. Williamson, and R. Berry, "Student perceptions of e-learning service quality, e-satisfaction, and e-loyalty," in *Research Anthology on Preparing School Administrators to Lead Quality Education Programs*, Information Resources Management Association, Ed., Hershey, PA: IGI Global Scientific Publishing, 2020, pp. 806–830, doi: 10.4018/978-1-7998-3438-0.ch037.

[28] H. C. Chu and G. J. Hwang, "A Delphi-based approach to developing expert systems with the cooperation of multiple experts," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2826–2840, May 2008, doi: 10.1016/j.eswa.2007.05.034.
[29] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate data analysis*, 8th ed. Harlow, U.K.: Pearson, 2019.
[30] Q. Wang, C. L. Quek, and X. Hu, "Designing and improving a blended synchronous learning environment: an educational design research," *International Review of Research in Open and Distributed Learning*, vol. 18, no. 3, pp. 99–118, May 2017, doi: 10.19173/irrodl.v18i3.3034.
[31] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly: Management Information Systems*, vol. 13, no. 3, pp. 319–339, Sep. 1989, doi: 10.2307/249008.

BIOGRAPHIES OF AUTHORS



Quang X. Tran is Vice-Dean of the Faculty of Teacher Education at VNU University of Education, Vietnam National University, Hanoi, Vietnam. He received his Ph.D. in Applied Mathematics and Computational Science from the University of Bordeaux I, University of Bordeaux, France, in 2015. He has more than 18 years of professional experience in higher education, including roles as a lecturer, researcher, academic manager, training manager, teacher trainer, and educational program developer. His research focuses on the following areas: applications of statistics in education, survival analysis, and analysis of stopping systems. He has published several research articles in national and international journals, including those indexed in Scopus. He has successfully supervised two Ph.D. in measurement and assessment in education; 20 master's students in mathematics education (teaching methodology), five master's students in measurement and assessment in education, and published 1 textbook for undergraduate and postgraduate teaching. He can be contacted at email: huong.tran@vnu.edu.vn; quangtx@vnu.edu.vn.



Huong Tran Thi Thu is Vice-Dean of the Faculty of Quality Management at VNU University of Education, Vietnam National University, Hanoi, Vietnam. She received her Ph.D. in International Studies from Vietnam National University, Hanoi, in 2018. She has more than 15 years of professional experience in higher education, including roles as a lecturer, researcher, academic manager, teacher trainer, and educational program developer. Dr. Huong is an officially certified higher education accreditation assessor in Vietnam and serves as an external reviewer for higher education institutions under national quality assurance frameworks. She is actively involved in higher education quality assurance and accreditation activities. Her research interests include measurement and assessment in education, quality management, accreditation, and educational technology. She has published research articles in national and international journals, including Scopus-indexed publications, and has received several recognitions for her academic contributions, including the Innovative Educator Award of Vietnam National University, Hanoi in 2020 and the best teacher of the year award in 2024. She can be contacted at email: huong.tran@vnu.edu.vn.



An T. Thuy Nguyen is a teacher at Pham Van Dong High School, part of the EQuest Education System, Vietnam. She holds a bachelor's degree in Chemical Education. She has over six years of teaching experience, with a strong focus on the application of digital transformation and blended learning in teaching, as well as curriculum development. Her research interests include educational quality management, measurement and assessment in education, and blended learning. She has been actively involved in implementing innovative teaching approaches to enhance student learning outcomes. She was recognized as an excellent teacher in 2024 and 2025. She can be contacted at email: an.nguyen@alphaschool.edu.vn.



Lam Le is the Rector of The College of Dai Viet, Sai Gon, Vietnam and also works as a researcher in higher education. He holds the doctoral degree in Educational Measurement and Assessment and have more than 20 years working in education management. His research interests are quality management, measurement and assessment in education. He can be contacted at email: lelam@daivietsaigon.edu.vn.