International Journal of Education and Practice
2026 Vol. 14, No. 4, pp. 12-26
ISSN(e): 2310-3868
ISSN(p): 2311-6897
DOI: 10.18488/61.v14i4.5095

Validating design indicators for an AI-assisted teaching strategies training module for university EFL lecturers using the Fuzzy Delphi Method

Shiyu Yuan[1+]
Nurfaradilla Mohamad Nasri[2]
Khairul Azhar Jamaludin[3]

[1,2,3]*Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia.*
[1]*Email: p115672@siswa.ukm.edu.my*
[2]*Email: nurfaradilla@ukm.edu.my*
[3]*Email: khairuljamaludin@ukm.edu.my*

(+ Corresponding author)

Article History
Received: 30 January 2026
Revised: 10 April 2026
Accepted: 20 July 2026
Published: 17 August 2026

Keywords
AI-assisted teaching strategies
Design indicators validation
English as a foreign language
Fuzzy Delphi method
Training module design.

ABSTRACT

Artificial intelligence is increasingly used in language education, yet classroom adoption in higher education often remains tool-focused. Lecturers, therefore, receive limited support on how to translate artificial intelligence into coherent and transferable teaching strategies. This study aimed to propose an artificial intelligence–assisted teaching strategies training module for university English as a foreign language lecturers and to validate design indicators to guide the next development stage. Forty-eight draft indicators were organised into five dimensions: objectives, content, unit structure, materials and resources, and assessment methods. Indicator validation used the fuzzy Delphi method with a panel of 15 experts in language pedagogy, teacher education, educational technology, and assessment. Experts rated each indicator, and consensus was evaluated using a predefined distance threshold, percentage agreement, and a fuzzy score. Overall agreement was strong. All indicators for objectives, unit structure, and materials and resources were retained (17/17). Most content indicators were retained (23/24). Assessment indicators were retained more selectively (4/7), reflecting higher expectations for feasible, classroom-ready evidence of strategy use. These findings specify essential design requirements for a strategy-oriented training module and support evidence-informed decisions in subsequent development and later pilot evaluation in authentic teaching settings. The indicator set also provides clear specifications for developing unit activities, resources, and assessment artifacts that can be reviewed and refined through iterative trials.

Contribution/Originality: This study offers a validated indicator framework that supports institutions in planning and standardising professional learning for strategy-oriented artificial intelligence integration. By defining necessary objectives, unit structure, resources, and assessment results based on evidence, it supplies an executable module development blueprint and a clear foundation for later field tests, which is an original contribution of this research work.

1. INTRODUCTION

Artificial intelligence (AI) is increasingly reshaping the education domain. It influences how learning is designed, delivered, and assessed. The Guidance for Generative AI in Education and Research from the United Nations Educational, Scientific, and Cultural Organization (UNESCO) stresses the potential of generative AI to provide personalized learning and formative feedback. It also emphasizes responsible and human-centred adoption (UNESCO, 2023). At the system level, the Organisation for Economic Co-operation and Development (OECD) points out that emerging AI technologies are reshaping teaching and learning roles. They also promote more adaptive and data-based teaching methods (OECD, 2025). These developments indicate that AI is becoming an important driver of educational transformation.

In the field of language education, artificial intelligence has been applied to learning assessment, self-test feedback, voice recognition, and adaptive tutoring services. AI helps the learning process and plays a positive role in improving academic performance. The latest review study found that the scope of application of artificial intelligence in language education is expanding, especially in intelligent tutoring, automatic evaluation feedback, and language processing (Huang, Zou, Cheng, Chen, & Xie, 2023). At the same time, empirical research also shows that adaptive feedback based on artificial intelligence helps English as a foreign language (EFL) learners to improve their learning investment time and writing ability (Luo & Yusuf, 2026). However, the implementation effect of AI in classroom teaching still shows an uneven pattern.

A key issue is that the application of artificial intelligence (AI) in teaching practice and teacher professional development often remains tool-oriented, with insufficient attention to pedagogical coherence and classroom implementation. Systematic review studies show that many educational AI studies focus more on the application of digital tools and less on the alignment of teaching objectives and the actual impact on teaching effectiveness (Qian, 2025; Tan, Cheng, & Ling, 2025). Meanwhile, teacher professional development programs often emphasize the mastery of technical skills but provide limited support in pedagogical reasoning and practice-based integration (Taheri, Nazemi, Pennington, Clark, & Dadgostari, 2025; Tammets & Ley, 2023). In EFL teaching contexts, teachers are using generative AI more often. However, there are still problems with curriculum alignment, teaching objectives, and teaching strategy design (Guan, Lee, Zhang, & Gu, 2025; Kyaw & Deng, 2025). This shows a clear gap. University EFL lecturers need structured support to develop AI-assisted teaching strategies based on pedagogy. They should focus on teaching strategies, not only on using single tools.

This gap is particularly obvious in China. In recent years, China's national-level digital education reform is accelerating (Ministry of Education of the People's Republic of China, 2024). The Ministry of Education has launched the "Education Digitalisation Strategic Action" strategy, proposed to build an "integrated, intelligent and international" digital education system, and further put forward the 2025 National Digital Education Strategy, emphasising the deepening of the application of artificial intelligence in curricula and teaching activities (Ministry of Education of the People's Republic of China, 2025). However, the latest research shows that the implementation at the classroom level still faces many challenges, including problems such as a digital literacy gap, insufficient support resources, and fragmentation of teaching practises (Chen, 2025; Yang, Chen, & Hou, 2025; Zhang & Dong, 2024). Therefore, the application of artificial intelligence in the classroom may still be in the stage of individualisation and exploration, and it is difficult to form a practical model that is driven by teaching strategies and can migrate applications in different teaching situations.

To address this need, the present study focuses on designing an AI-assisted teaching strategy (AITS) training module for university EFL lecturers in China. In this study, "content" is conceptualized as AI-assisted teaching strategies, referring to lecturers' instructional decisions on organizing input, tasks, scaffolding, feedback, practice pathways, and AI output verification. Based on earlier needs analysis work, this research aims to build an expert-validated collection of module parts across five design aspects: module goals, module content, unit framework, materials/sources, and assessment measures, through expert talks and the Fuzzy Delphi Method. Therefore, the research puts forward the following research question.

What components should be included in an AI-assisted teaching strategies (AITS) training module for university EFL lecturers in China?

2. LITERATURE REVIEW

2.1. AI-Assisted Teaching Strategies in University EFL Education

Artificial intelligence (AI) is exerting more and more influence on instructional design, feedback implementation, and learner assistance in various educational environments. In this research, teaching strategies are defined as instructional decisions with a clear purpose that arrange learning activities and steps to reach expected results, and make pedagogy match with assessment (Biggs, Tang, & Kennedy, 2022; Joyce, Weil, & Calhoun, 2009). Therefore, AI-assisted teaching strategies are regarded as strategy-level designs that combine AI's available functions into lesson plan creation, classroom practice, and assessment in a principle-based, goal-led way. It focuses on pedagogical coherence and instructional decision-making, not on treating AI tools as separate add-ons (Holmes & Tuomi, 2022).

In university EFL education, AI-supported teaching practices have been widely discussed in areas such as automatic writing feedback, oral English support, and personalised learning support. Research shows that AI-based automatic written feedback can promote writing progress and revision work when it is inserted into daily teaching activities (Shi & Aryadoust, 2024). AI-driven speech technologies and dialogue systems can increase chances for oral practice, with possible good effects on pronunciation and lower speaking worry during communication tasks (Dennis, 2024). These studies indicate that AI can support key aspects of EFL teaching, but its effectiveness depends on how well it is integrated into coherent instructional planning.

However, a large part of existing research still takes tools as the core, with scarce attention paid to strategy design and classroom implementation. Bibliometric and review research results show that, as Huang et al. (2023) indicate, AI studies usually place tool functions and learner views in prominent positions, but provide little guidance on task arrangement, learning support provision, and assessment design as integrated teaching strategies. In actual teaching situations, teacher professional development activities are also often carried out as one-time tool-show events, instead of long-term help for teachers to design, put into practice, and improve AI-supported teaching strategies in real school environments (Tammets & Ley, 2023). Therefore, lecturers may conduct AI-related teaching attempts without clear instructions on how to match AI support with course goals, handle classroom interaction processes, or assess learning results within a unified and complete teaching strategy framework.

Overall, the literature indicates one gap: while AI's development potential in EFL education is increasing, university EFL lecturers' pedagogical preparedness for systematic AI integration remains insufficient. This gap provides a basis for a targeted training module to support EFL lecturers in developing AI-assisted teaching strategies through step-by-step, classroom-centered design guidance. Based on this gap, the study employs the Fuzzy Delphi Method to identify and validate the key components of the module, aiming to promote strategy-based and sustainable AI integration in university EFL teaching.

2.2. Training Modules for University EFL Lecturers

In Chinese higher education, recent reforms have increased expectations for university English lecturers. Lecturers are expected to help students develop communicative and intercultural competence. Under the Belt and Road Initiative, policy attention has moved beyond exam results. It now focuses more on communication and intercultural ability, so lecturers need to update their teaching approaches (Hu, Li, & Luo, 2024). At the curriculum level, competence-based frameworks such as China's Standards of English Language Ability (CSE) promote application-focused teaching. They also call for better classroom teaching and assessment (Liu, 2024). These changes increase the need for structured professional learning. This support can help lecturers redesign their classroom teaching.

Teacher professional development is commonly conceptualized as a sustained process via which educators enhance content knowledge and polish pedagogy in reaction to instructional needs and reform priorities (Darling-Hammond, Hyler, & Gardner, 2017; Desimone, 2009). In this field, training modules provide a practical form to arrange professional learning into ordered units with clear learning results, learning activities, support materials, and assessment assignments. Module development research often follows systematic design logics such as ADDIE, and it may also adopt a design and development research (DDR) process that combines evidence, expert suggestions, and repeated revision (Li & Abidin, 2024; Low, Balakrishnan, & Yaacob, 2024; Said, Meng, Musa, & Yaakob, 2022).

Across various subject fields, effective teaching modules are generally focused on concrete problems in practical teaching and are supported by practice-centred learning loops (Said et al., 2022). DDR-guided research usually starts from requirement analysis, confirms design components through expert assessment, and then moves on to implementation and evaluation of effects. In the field of language teacher education, modules that combine structured elements with context-based teaching tasks have been shown to increase teachers' sense of relevance and usability (Xiao, Surasin, & Prabjandee, 2020). These patterns are in line with broader views that effective teacher learning must remain closely connected to classroom practice and support cycles of practice and reflection, rather than one-time workshops (Borko, 2004; Clarke & Hollingsworth, 2002). Based on this, the current research uses FDM to examine and validate the dimensions and indicators of an AI-supported teaching strategies training module for university EFL teachers.

2.3. Fuzzy Delphi Method in Educational Research

The Delphi technique was originally developed to elicit and synthesise expert judgement through structured questionnaires and statistical aggregation (Dalkey & Helmer, 1963). In the educational research domain, it has been widely applied for clarifying complex constructs, defining competencies, and supporting planning decisions when empirical evidence still remains in a limited state (Hsu & Sandford, 2007; Johnson & Chan, 2023). Anonymity and systematic consolidation of group responses are central characteristic features; therefore, they help to reduce dominance effects and facilitate the building of reasoned consensus.

Even though it has certain use value, the traditional Delphi method usually needs several rounds of circulation, which raises the level of time consumption and may cause participant tiredness or dropout, especially among occupied professional workers (Hasson, Keeney, & McKenna, 2025; Ishikawa et al., 1993). Moreover, expert judgments are frequently expressed by language words that are hard to represent precisely via clear numerical figures, thus bringing possible vagueness during quantitative analysis (Zadeh, 1965). In order to solve these restricting problems, fuzzy set theory has been added to Delphi procedures; this can better carry out uncertainty modeling of human judgment and lower dependence on repeated circulation rounds (Murray, Pipino, & van Gigch, 1985).

The Fuzzy Delphi Method (FDM) makes this integration operational by changing expert-given ratings into triangular fuzzy numbers and gathering them together so as to judge if every single item satisfies an acceptable consensus threshold (Ishikawa et al., 1993). This offers a transparent foundation to keep, revise, or delete indicators, thus making FDM suitable for instrument refinement and research with design orientation. Inside Design and Development Research (DDR), expert validation further strengthens content validity because it makes sure that design specifications reflect shared expectations before the process of implementation starts (Richey & Klein, 2014). Hence, the present study used FDM to confirm the necessary parts of an AI-assisted teaching strategies training module that is provided for university EFL lecturers.

2.4. Conceptual and Theoretical Underpinnings of the Design Phase

To keep concept consistency and guarantee practice defensibility, the design stage made use of mutually complementary frameworks. These frameworks make clear both which content should enter the module and how the content must be arranged. From a method angle, Design and Development Research (DDR) regards module design as a process with evidence support and repeated operation. To raise usability and context-relatedness, design specifications are polished via systematic validation (Richey & Klein, 2014). In the current research, therefore, DDR provides the top-level logic to move from an initial index group, formed from literature, need analysis, and expert offer, to a verified module component set that is built via expert agreement.

At the instructional design level, an ADDIE-oriented logic was used to organise the questionnaire indicators. In the design stage, ADDIE typically requires a clear definition of learning objectives and content boundaries; it also requires planning of learning activities and delivery structure, together with resources and evaluation requirements that are necessary for coherent development and implementation (Branch, 2009; Molenda, 2003). In terms of conceptual support, TPACK makes clear what kind of knowledge integration is needed for AI-supported instruction in university EFL environments (Koehler & Mishra, 2009; Mishra & Koehler, 2006); it stresses alignment among technology, pedagogy, and language-content goals, rather than a narrow focus on tool operation. Adult learning theory furthermore supports the inclusion of practice-oriented and workplace-related learning experiences, because lecturers tend to prefer autonomy, problem-centred learning, and immediate applicability to teaching tasks (Knowles, 1984; Knowles, Holton, & Swanson, 2014). Finally, task-based principles informed the content indicators by emphasising meaningful classroom tasks and observable communicative outcomes (Nunan, 2004).

Several domain-specialized viewpoints were applied to reinforce the design dimensions. CALL evaluation principles maintained the content indicators as strategy-directed and transferable by framing steady AI capabilities as classroom decision loops instead of tool-specified procedures (Chapelle, 2001). As a balance inspection across the strategy collection, the Nation's Four Strands guaranteed coverage of meaning-focused input and output, language-centered learning, and fluency cultivation (Nation, 2007). The suggested unit sequence was extracted from teacher professional learning research that emphasizes theory input, modeling, guided practice, feedback, and support for transfer into classroom implementation (Joyce & Showers, 2002). Moreover, international guidance on responsible AI application also provided information on minimum governance demands, containing verification, transparency, and human supervision, to back feasible execution and evaluable results (UNESCO, 2024).

2.5. Summary of the Literature Review

This literature review lists key conceptual and methodological points for developing an AI-supported teaching strategies training module for university EFL lecturers. Earlier studies show that AI can support feedback, practice, and various learning supports in EFL teaching. However, many lecturers still lack structured support to integrate AI into pedagogically consistent and daily classroom practices. Therefore, training modules serve as a practical form for long-term professional development and can align with curricular demands. For the design stage, the Fuzzy Delphi Method offers a clear step to validate and improve module indicators through expert consensus within a Design and Development Research framework. The next chapter then details the design steps and related analysis.

3. METHODOLOGY

3.1. Research Design

This study employed the Fuzzy Delphi Method (FDM) to carry out validation of the put-forward indicators for an AI-assisted teaching strategies (AITS) training module, and furthermore to build expert consensus for university EFL lecturers. Because indicator validation depends heavily on expert judgement, and fuzzy set procedures can make consensus quantifiable at the same time as they lower ambiguity in ratings expressed via language (Ishikawa et al., 1993; Okoli & Pawlowski, 2004). FDM was selected for this study. The study was under the orientation of Design and Development Research (DDR), where validated indicators are regarded as design specifications for the development of an educational product with usability (Richey & Klein, 2014).

An original indicator pool was first built from three sources: literature proof, Phase 1 needs analysis discoveries, and first-step semi-structured talks with five field experts. After that, a 15-person expert group assessed the suggested indicators via an FDM questionnaire. Each score was changed into a triangular fuzzy number and gathered to check whether the indicator satisfied the needed consensus level (Ishikawa et al., 1993). Therefore, decisions about keeping or improving indicators were made by applying set FDM standards, which contain threshold value (d), agreement percentage, and defuzzification score (Hsu & Sandford, 2007; Okoli & Pawlowski, 2004). Hence, the following sections will describe the expert selection procedure and the questionnaire's development process.

3.2. Selection of Experts

A group of 15 experts was collected to carry out an assessment of the suggested indicators for the AITS training module. Delphi research commonly indicates that a panel with about 10–18 experts can offer a stable foundation for consensus and, at the same time, stay practical for systematic evaluation (Okoli & Pawlowski, 2004). Methodological discussions also suggest that a medium-sized panel helps balance rigor and practicality in professional settings (Gray, 2016). Therefore, the panel size used in this research was considered suitable.

A purposive sampling method was used to recruit experts with relevant knowledge and practical experience. Based on the Delphi guidelines (Hsu & Sandford, 2007), the selection focused on participants who could give informed and reliable judgments on the research topic. In this study, "experts" referred to university lecturers or academics who met two main criteria. One criterion was having rich experience in university English teaching and/or teacher education. Another was taking an active role in curriculum or module development, educational technology integration, or AI-related work in higher education. This definition also follows the view that Delphi panels should have both subject knowledge and familiarity with the real application context (Ab Latif, Dahlan, Ab Mulud, & Mat Nor, 2017). The final expert panel included senior EFL lecturers, teacher educators, programme coordinators, and educational technology specialists.

Experts were identified through professional networks and institutional recommendation letters, then checked against pre-set participation criteria. Those who qualified were invited to join the study and received a Letter of Appointment explaining the study's aim and required input. Messages about secrecy, name hiding, and informed consent were provided in the questionnaire's information section before the work was completed.

To raise transparency, we describe the expert panel in institutional and regional aspects. The participating experts in China were affiliated with Zhengzhou University and Henan Polytechnic University, as well as higher education institutions in Jiangxi, Inner Mongolia, Jiangsu, Fujian, and Shanghai. Additionally, a small number of experts worked in overseas EFL environments, such as the United States and Malaysia.

Participant characteristics are shown in Table 1. On the whole, the expert group was of high seniority and rich experience: 66.7% of them hold doctoral degrees, 86.6% have more than 11 years of work experience, and 86.7% are associate professors or professors. In the aspect of professional expertise, 86.7% of the experts have backgrounds in university EFL education, 100% report having expertise in curriculum and pedagogy, and 53.3%, which is over half, have expertise in educational technology.

Table 1. Demographic information expert.

Variables	Category	n (N = 15)	%
Level of education	Master's degree	5	33.3
	Doctoral degree	10	66.7
Work experience	5-10 years	2	13.3
	11-15 years	8	53.3
	More than 15 years	5	33.3
Position	Lecturer	2	13.3
	Associate Professor / Professor	13	86.7
	University EFL education	13	86.7
Domain of expertise	Curriculum and pedagogy	15	100.0
	Educational technology	8	53.3

Note: Percentages are calculated based on the total number of experts (N = 15). Domain of expertise is a multiple-response item; experts could select more than one area.

3.3. FDM Questionnaire Development

The FDM questionnaire was created to build expert agreement on the design dimensions and indicators needed for the AITS training module. To let expert judgments feed directly into the module blueprint, the questionnaire used an ADDIE-based structure and divided indicators into five dimensions: Objectives, Content, Structure, Materials/Resources, and Assessment (Branch, 2009; Molenda, 2003). Each item was presented in checklist style to keep statements short and observable, making expert rating easier. Following a DDR orientation, the indicators confirmed via FDM were regarded as design standards for guiding later module development and refinement (Richey & Klein, 2014).

Regarding the Content dimension, "content" was defined as AI-supported teaching tactics instead of training on specific AI tools. Due to the fast speed of tool updates, the indicators were compiled in ability-related words and concentrated on pedagogical choices that can be carried out in the classroom. This method is in line with CALL evaluation logic, which holds that technology application should be validated by learning worth and task suitability, not by the tool itself (Chapelle, 2001). Moreover, Nation's Four Strands were applied as a balance inspection to make sure that the content indicators included meaning-centered input and output, language-centered study, and fluency growth (Nation, 2007).

The remaining dimensions used already built frameworks for teacher learning and instructional integration. TPACK gave information to the Objectives dimension and maintained focus on how technology, pedagogy, and EFL content knowledge combine in university teaching (Koehler & Mishra, 2009; Mishra & Koehler, 2006). For Structure, the Joyce and Showers model formed the unit sequence. It puts strong stress on theory input, demonstration, guided practice, feedback, and follow-up support to make classroom transfer easier (Joyce & Showers, 2002). Materials and resources were regarded as implementation supports, which include guidance notes, templates, and exemplars. These supports were added to raise feasibility and push for more consistent carrying out across teaching contexts (Van Merriënboer, Kirschner, & Frèrejean, 2024). Assessment indicators were defined as evidence-based outputs. They capture lecturers' acceptance of the strategies through formative checks, feedback design, and verification practices (Black & Wiliam, 1998; Hattie & Timperley, 2007).

3.4. Data Collection Procedure

Data collection began after the expert panel was confirmed and the FDM questionnaire finalized. Each participant received a Letter of Appointment as a Fuzzy Delphi Expert, outlining their role, scope of contribution, and participation expectations. Informed consent was obtained before completing the questionnaire.

Wenjuanxing (Questionnaire Star) was used to carry out the questionnaire online, allowing people from different locations to participate and ensuring consistent data capture. The study's title and purpose were displayed at the beginning of the survey. Brief definitions of key terms, such as AI-assisted teaching strategies and module indicators, were provided to help respondents develop a shared understanding. The questionnaire was completed anonymously by experts, with most respondents spending about 5–15 minutes on it. The survey remained open for one week; reminder messages were sent during this period to increase the response rate. After the survey closed, all responses were exported for initial screening and preparation work.

Ethical protection measures were maintained throughout the entire course. Therefore, participation was based on free will, and no personally identifiable information was collected. All data were saved in password-protected folders, and only the research group could access them. After responses were exported, they were examined for completeness and classified according to the five dimensions of Objectives, Content, Structure, Materials/Resources, and Assessment to support the following FDM analysis.

3.5. Data Analysis

Data analysis followed an established FDM procedure to decide whether each proposed indicator met the criteria for retention as a module design specification (Ishikawa et al., 1993). Expert ratings were converted from the 5-point Likert scale into triangular fuzzy numbers (TFNs) using the predefined mapping scheme (Table 2), with TFN representation illustrated in Figure 1. TFNs were then aggregated across experts to generate a collective fuzzy estimate for each indicator.

Table 2. Mapping of Likert ratings to triangular fuzzy numbers.

Agreement level	Likert score	Triangular fuzzy number (m1, m2, m3)
Strongly disagree	1	(0.0, 0.1, 0.2)
Disagree	2	(0.0, 0.2, 0.4)
Partially agree	3	(0.2, 0.4, 0.6)
Agree	4	(0.4, 0.6, 0.8)
Strongly agree	5	(0.6, 0.8, 1.0)

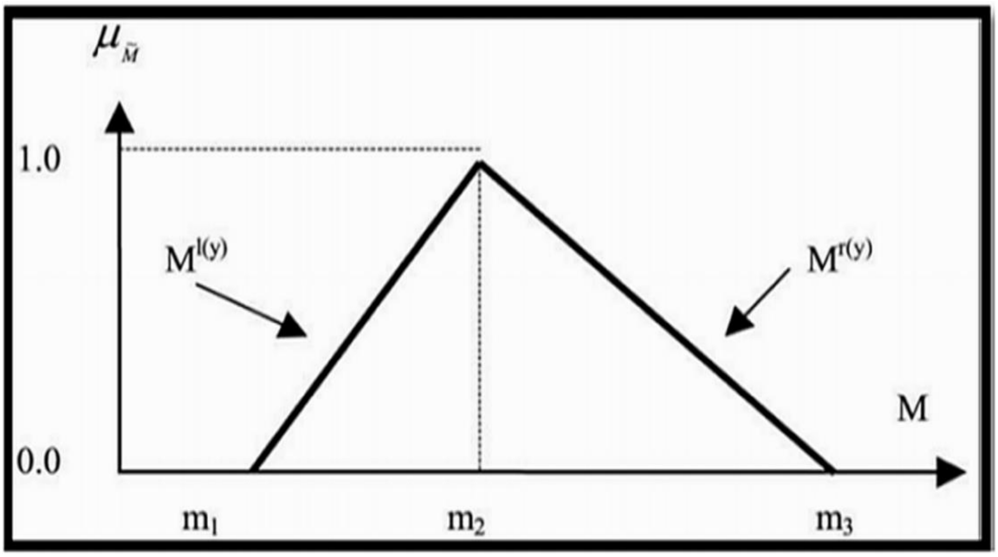

Figure 1. Triangular fuzzy number representation.

Indicator retention was determined using three commonly reported criteria in Delphi and FDM research (Hsu & Sandford, 2007; Okoli & Pawlowski, 2004). First, the threshold distance (d) was calculated to assess dispersion in fuzzy judgements, and indicators were retained when d ≤ 0.20 (Equation 1). Second, consensus proportion was computed, with indicators retained when agreement reached 75% or above. Third, defuzzification converted the aggregated TFNs into a crisp score using A = (m1 + m2 + m3) / 3, and indicators with A ≥ 0.50 were accepted (Equation 2). Defuzzified scores were also used to describe the relative emphasis of indicators within each design dimension. All computations were performed in Microsoft Excel to ensure transparency and allow step-by-step verification of results.

Equation 1. Threshold distance (d).

$$d(\tilde{m},\tilde{n})=\sqrt{\frac{1}{3}[(m_1-n_1)^2+(m_2-n_2)^2+(m_3-n_3)^2]} \quad (1)$$

Equation 2. Defuzzification (A value).

$$A=\frac{m_1+m_2+m_3}{3} \quad (2)$$

3.6. Rigour and Quality Assurance

Several measures were adopted to enhance research strictness. The original indicator pool obtained content from three resources: literature-based proof, Phase 1 results of needs analysis, and first-stage expert talks. This three-source combination helped to consolidate both the research's foundation and its adaptation to local conditions. This approach is consistent with DDR, which emphasises evidence-based design principles before product development (Richey & Klein, 2014).

We solved the instrument clarity problem before the test was carried out. Three experts examined the draft questionnaire, with their focus on wording, consistent terminology, and ambiguous statements. We made revisions according to their feedback. To restrict deviation, experts completed the questionnaire independently and anonymously, which could help decrease peer influence and social desirability effects (Hsu & Sandford, 2007). Explicit decision rules and a traceable workflow from raw ratings to acceptance decisions were also applied in the analysis (Ishikawa et al., 1993; Okoli & Pawlowski, 2004). All data were kept in a safe condition and used only for research-related aims.

4. RESULTS

4.1. Expert Consensus on Module Objectives Design

Six objective indicators were evaluated (Table 3). All six satisfied the cut-off standards of FDM ($d \leq 0.20$, consensus > 75%, and $A \geq 0.50$) and were retained, indicating strong agreement on the module's expected results. The retained objectives covered both strategy implementation (such as designing AI-supported listening, speaking, reading, writing, and translation activities) and responsible classroom management (such as tool–task selection and risk-response procedures). The objective with the highest score stressed establishing AI-supported speaking situations and arranging interactive oral practice with usable feedback ($A = 0.680$, Rank 1). Responsible AI integration principles and differentiated reading adaptation were close behind (both $A = 0.667$, Rank 2), suggesting that experts placed the greatest importance on objectives that can be directly applied in classroom planning and delivery.

Table 3. Expert consensus on module objectives.

Code	Objective indicator (Short label)	d	Consensus (%)	A	Rank
S3-O	Speaking contexts & interactive practice + feedback	0.112	93	0.680	1
U1-O	Responsible AI integration & classroom governance	0.142	80	0.667	2
R4-O	Reading adaptation & differentiated scaffolding	0.142	87	0.667	2
W5-O	Process writing support (Feedback + Revision)	0.117	87	0.653	4
L2-O	Listening task design & difficulty diagnosis	0.128	80	0.640	5
T6-O	Translation/Mediation tasks + Post-editing QC	0.100	80	0.613	6

4.2. Expert Consensus on Module Content Design

Twenty-four content indicators were evaluated across six-unit domains (Table 4): Unit 1 (AI literacy and responsible integration) and five skill-centered units (listening, speaking, reading, writing, translation). Of these, 23 indicators satisfied the FDM criteria ($d \leq 0.20$; consensus > 75%; $A \geq 0.50$) and were kept in reserve. The kept indicators constructed AI use as classroom-executable strategy cycles, which include input adjustment, scaffolding support, diagnosis of learner difficulties, personalized practice, feedback production, and post-editing/quality control.

One item (U1-1) was rejected because consensus did not reach the cut-off value (73% < 75%), even though its d and A values were acceptable ($d = 0.160$; $A = 0.667$). This is because the core content it intended to express was already covered by two retained items: verification and correction (U1-2; $d = 0.107$; consensus = 93%; $A = 0.720$), and academic integrity and transparent disclosure (U1-3; $d = 0.114$; consensus = 93%; $A = 0.693$). This helps reduce content overlap and, at the same time, keep all key governance elements.

Table 4. Top-ranked and rejected content indicators.

Code	Content indicator (short label)	d	Consensus (%)	A	Decision	Rank
U1-2	AI limitation awareness & verification/Correction	0.107	93	0.720	Accepted	1
S3-1	Speaking prompts & context generation	0.096	100	0.720	Accepted	2
R4-4	Differentiated reading practice support	0.096	100	0.720	Accepted	2
U1-4	Tool–task alignment by instructional purpose	0.124	87	0.707	Accepted	3
U1-1	Ethics & privacy mechanisms	0.160	73	0.667	Rejected	—

Note: Overall, 23 out of 24 content indicators were retained; one item (U1-1) was removed due to marginally insufficient consensus.

4.3. Expert Consensus on Unit Structure Design

Five unit-structure indicators were assessed (Table 5): Theory explanation, Demonstration, Guided practice, Feedback and revision, and Transfer and follow-up. All met the FDM criteria (d ≤ 0.20; consensus > 75%; A ≥ 0.50) and were retained. Dispersion was low (d = 0.085–0.128), and consensus remained high (87%–93%) across items.

Demonstration (B3-2) and Guided practice (B3-3) received the strongest support (consensus = 87%–93%, A = 0.693). Theory explanation (B3-1) and Feedback and revision (B3-4) were also well supported (d = 0.107–0.124; consensus = 87%–93%; A = 0.667). Transfer and follow-up (B3-5) showed slightly lower A (A = 0.640) but remained within the acceptance range (d = 0.085; consensus = 93%).

Table 5. Result of unit structure design consensus.

No.	Structure element	d	Consensus (%)	A	Rank
1	Demonstration	0.128	87	0.693	1
2	Guided practice	0.114	93	0.693	1
3	Theory explanation	0.124	87	0.667	2
4	Feedback and revision	0.107	93	0.667	2
5	Transfer and follow-up	0.085	93	0.640	3

4.4. Expert Consensus on Materials and Resources Design

All six materials/resources indicators satisfied the FDM criteria (d ≤ 0.20; consensus > 75%; A ≥ 0.50) and were retained. For space efficiency, Table 6 reports category-level results, grouped into core operational supports (R1, R3–R5; A = 0.680) and governance & evaluation supports (R2, R6; A = 0.640).

Operational supports included a module guide, prompt/workflow toolkit, worked exemplars, and templates (e.g., micro-lesson plans/task chains, rubrics, verification records). Governance-oriented supports included a responsible use guide and unified feedback tools (e.g., rubrics/checklists and peer-review forms). Item-level results showed acceptable convergence (d ≈ 0.11–0.13; consensus ≥ 80%).

Table 6. Result of materials and resources design consensus.

Resource type	Items included	A (Range)	Rank pattern
Core operational supports	R1, R3, R4, R5	0.680	Rank = 1
Governance & evaluation supports	R2, R6	0.640	Rank = 2

4.5. Expert Consensus on Assessment Methods Design

Seven assessment indicators were evaluated using the FDM criteria (d ≤ 0.20; consensus > 75%; A ≥ 0.50). Four indicators were therefore retained in the study (Table 7), while three indicators were eliminated because their consensus levels were below the cut-off value, even though their d and A values met the acceptable standard.

The retained deliverables focused on evidence-based teaching artifacts: a micro-lesson plan/task chain, a classroom materials package, lesson-level assessment and feedback design, and a verification and revision record documenting human checks of AI outputs. The classroom materials package (B5-2) received the strongest support (A

= 0.680, Rank 1). The verification and revision record (B5-5) was also retained (A = 0.613). For rejected indicators, their functions (e.g., transparency of AI use and transfer evidence) were incorporated into the retained artifacts through template fields and rubric criteria to avoid redundancy.

Table 7. Retained assessment deliverables.

Code	Assessment deliverable (short label)	d	Consensus (%)	A	Rank
B5-2	Classroom materials package	0.112	93	0.680	1
B5-1	Micro-lesson plan / Task chain	0.116	80	0.627	2
B5-4	Lesson-specific assessment & feedback design	0.116	80	0.627	2
B5-5	Verification & revision record (Human checking)	0.100	87	0.613	3

4.6. Summary of the Results

Across the five dimensions, expert consensus was strongest in Objectives, Unit Structure, and Materials/Resources; therefore, all indicators in these areas were retained. Content reached near-full indicator retention; however, one indicator was rejected because its consensus percentage was slightly below the required level. Assessment was relatively more selective, with four indicators retained and three removed as independent deliverables. The retained indicators provide an expert-validated blueprint to guide subsequent module development.

5. DISCUSSION

This FDM study established a consensus-based indicator set for designing an AI-assisted teaching strategies (AITS) training module for university EFL lecturers. Across dimensions, experts consistently favored indicators that can be enacted in real lessons and produce observable teaching artifacts. This supports a design focus on transferable classroom decision-making rather than tool familiarization alone.

One contribution of this study lies in the way the module content is conceptualized. Experts supported indicators that regard AI integration as one set of teachable strategy cycles. These cycles contain input adaptation, scaffolding design, difficulty diagnosis, personalized practice organization, and feedback generation. Therefore, the emphasis lies on capability, not specific tools. This view conforms to CALL evaluation principles, since technology is judged by learning value and task fit, not by novelty (Chapelle, 2001). The consensus also points to the need for balance in strategy coverage. The validated set surpasses material generation and includes meaning-focused input and output, language-focused learning, and fluency development (Nation, 2007). Moreover, experts also strongly approved of verification and correction. This places quality control at the core of AI-assisted teaching, which is in line with guidance on trustworthy AI that highlights human oversight and transparent handling of AI outputs (National Institute of Standards and Technology, 2023; OECD, 2019; UNESCO, 2021).

The unit-structure research outcomes cast illumination on delivery priority arrangements. Demonstration and guided practice obtained the most powerful degree of support. Thus, it indicates that lecturers can gain benefits from worked examples and protected time to carry out rehearsal of new practices. This pattern is in alignment with Joyce and Showers' staff development model, which connects classroom transfer to modelling, practice, feedback, and follow-up, instead of only to theory input (Joyce & Showers, 2002). Transfer and follow-up were kept but positioned at lower ranks. Therefore, this may point out a preference for embedding transfer via unit outputs and daily routines, without adding a large workload that exceeds the core learning activities.

The results concerning materials and resources strengthen the module's implementation emphasis. Experts gave high priority to practical supports, including a module guide, toolkits, exemplars, templates, and unified feedback tools. Such resources can reduce preparation time and make enactment more consistent, particularly when time is limited, and learner needs are diverse. Governance-related supports were also retained, as they serve as safeguards for responsible use and support safer, more sustainable implementation rather than forming the core learning content.

Assessment indicators attracted more cautious endorsement. Experts preferred evidence-based deliverables that demonstrate teaching competence through classroom-ready artefacts (e.g., lesson/task-chain plans, materials packages, assessment designs, and verification records). Items based on reflective or declarative reporting received less support, likely because similar evidence can be captured within templates, rubrics, and documentation fields without increasing workload. This suggests that assessment in the AITS module should remain closely aligned with unit outputs, primarily formative in function, and explicit in documenting responsible AI use through verification practices.

Overall, the findings support a module blueprint centered on strategy capability, practice-based learning sequences, and implementation-ready supports. The validated indicators provide a defensible basis for subsequent module development and pilot evaluation of usability, user acceptance, and instructional impact.

6. CONCLUSIONS

This study applied the Fuzzy Delphi Method to confirm a consensus-based indicator set for designing an AI-assisted teaching strategies (AITS) training module for university EFL lecturers. The validated indicators position AI integration as a strategy-focused classroom decision-making tool rather than just an operation. The resulting blueprint outlines requirements for module objectives, content, unit structure, materials/resources, and assessment, serving as a workable starting point for further module development.

Two implications are clearly pointed out by the research findings. Firstly, experts consistently gave higher preference to indicators that can be changed into classroom action; especially, guided practice, worked examples, and practical support matters such as templates, exemplars, and toolkits are included. Secondly, responsible AI usage is also regarded as part of teaching ability. Therefore, this is shown through the inclusion of verification and correction practices, together with demands related to transparency and academic integrity. Hence, assessment preferences further emphasized feasibility because experts tended to prioritize evidence-based artifacts that make the lecturers' strategy use clearly visible in teaching products.

The study also has limitations that need to be acknowledged. Although suitable for an FDM study, the expert panel was limited in size and contextual coverage, which may reduce the generalizability of the findings. Additionally, the study validates design indicators rather than demonstrating instructional effectiveness. Future work should develop the full module from the validated blueprint and conduct pilot implementation and evaluation to examine usability, acceptance, and impact on lecturers' teaching practices.

Funding: This research received no external funding.
Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Henan Polytechnic University (protocol code 20250127003, date 27, 01, 2025).
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Disclosure of AI Use: OpenAI's ChatGPT was used to assist with English language editing (clarity and grammar). The authors reviewed, revised, and take full responsibility for the final content.

REFERENCES

Ab Latif, R., Dahlan, A., Ab Mulud, Z., & Mat Nor, M. Z. (2017). The Delphi technique as a method to obtain consensus in health care education research. *Education in Medicine Journal, 9*(3), 89–102. https://doi.org/10.21315/eimj2017.9.3.10

Biggs, J., Tang, C., & Kennedy, G. (2022). *Teaching for quality learning at university* (5th ed.). UK: McGraw-Hill Education.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15. https://doi.org/10.3102/0013189X033008003

Branch, R. M. (2009). *Instructional design: The ADDIE approach.* New York, NY: Springer. https://doi.org/10.1007/978-0-387-09506-6

Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research.* Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781139524681

Chen, P. (2025). Teaching adaptation strategies for Chinese university English teachers in the AI era. *Advances in Social Science and Culture, 7*(5), 49–56. https://doi.org/10.22158/assc.v7n5p49

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947-967. https://doi.org/10.1016/S0742-051X(02)00053-7

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458-467. https://doi.org/10.1287/mnsc.9.3.458

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development.* Palo Alto, CA: Learning Policy Institute. https://doi.org/10.54300/122.311

Dennis, N. K. (2024). Using AI-powered speech recognition technology to improve English pronunciation and speaking skills. *IAFOR Journal of Education, 12*(2), 107-126. https://doi.org/10.22492/ije.12.2.05

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199. https://doi.org/10.3102/0013189X08331140

Gray, C. J. (2016). The Delphi technique: Lessons learned from a first time researcher. *Issues in Information Systems, 17*(4), 91–97. https://doi.org/10.48009/4_iis_2016_91-97

Guan, L., Lee, J. C.-K., Zhang, Y., & Gu, M. M. (2025). Investigating the tripartite interaction among teachers, students, and generative AI in EFL education: A mixed-methods study. *Computers and Education: Artificial Intelligence, 8*, 100384. https://doi.org/10.1016/j.caeai.2025.100384

Hasson, F., Keeney, S., & McKenna, H. (2025). Revisiting the Delphi technique-research thinking and practice: A discussion paper. *International Journal of Nursing Studies, 168*, 105119. https://doi.org/10.1016/j.ijnurstu.2025.105119

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*(4), 542-570. https://doi.org/10.1111/ejed.12533

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation, 12*(1), 10. https://doi.org/10.7275/pdz9-th90

Hu, H., Li, F., & Luo, Z. (2024). The evolution of China's English education policy and challenges in higher education: Analysis based on LDA and Word2Vec. *Frontiers in Education, 9*, 1385602. https://doi.org/10.3389/feduc.2024.1385602

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society, 26*(1), 112-131.

Ishikawa, A., Amagasa, M., Shiga, T., Tomizawa, G., Tatsuta, R., & Mieno, H. (1993). The max-min Delphi method and fuzzy Delphi method via fuzzy integration. *Fuzzy Sets and Systems, 55*(3), 241-253. https://doi.org/10.1016/0165-0114(93)90251-C

Johnson, A., & Chan, T. C. (2023). Effectiveness of the Delphi technique as an educational planning tool. *Educational Planning, 30*(1), 43-57.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Joyce, B. R., Weil, M., & Calhoun, E. (2009). *Models of teaching* (8th ed.). Boston, MA: Pearson/Allyn and Bacon.

Knowles, M. S. (1984). *The adult learner: A neglected species* (3rd ed.). Houston, TX: Gulf Publishing.

Knowles, M. S., Holton, E. F., & Swanson, R. A. (2014). *The adult learner: The definitive classic in adult education and human resource development* (8th ed.). United Kingdom: Routledge.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60-70.

Kyaw, E. M., & Deng, J. (2025). Systematic review of artificial intelligence integration in English language teaching: Trends, applications, and pedagogical implications. *Educalitra: English Education, Linguistics, and Literature Journal*, *4*(2), 153-181.

Li, C. L., & Abidin, M. J. B. Z. (2024). Instructional design of classroom instructional skills based on the ADDIE model. *Technium Social Sciences Journal*, *55*(1), 167–178.

Liu, X. (2024). *Research on college English teaching reform based on China's Standards of English language ability*. Paper presented at the Proceedings of the 3rd International Conference on Education, Language and Art (ICELA 2023) (pp. 194–202). Paris, France: Atlantis Press.

Low, J. Y., Balakrishnan, B., & Yaacob, M. I. H. (2024). The design and development of a teacher training module on the game-based learning approach in teaching physics: A conceptual framework based on design and development research (DDR) approach. *Journal of Education and Social Sciences*, *27*(1), 73–82.

Luo, L., & Yusuf, A. (2026). Bridging AI and pedagogy: How AI-adaptive feedback shapes Chinese EFL students' writing engagement, metacognitive writing strategies, and writing performance. *Assessment & Evaluation in Higher Education*, *51*(3), 508–531. https://doi.org/10.1080/02602938.2025.2548919

Ministry of Education of the People's Republic of China. (2024). *MOE launches special action plan for integrated, intelligent, and internationalized digital education*. China: Ministry of Education of the People's Republic of China.

Ministry of Education of the People's Republic of China. (2025). *MOE outlines 2025 national strategy for digital education*. China: Ministry of Education of the People's Republic of China.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education*, *108*(6), 1017-1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Molenda, M. (2003). In search of the elusive ADDIE model. *Performance Improvement*, *42*(5), 34-36. https://doi.org/10.1002/pfi.4930420508

Murray, T. J., Pipino, L. L., & van Gigch, J. P. (1985). A pilot study of fuzzy set modification of Delphi. *Human Systems Management*, *5*(1), 76-80. https://doi.org/10.3233/HSM-1985-5111

Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, *1*(1), 2-13. https://doi.org/10.2167/illt039.0

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. United States: National Institute of Standards and Technology.

Nunan, D. (2004). *Task-based language teaching*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511667336

OECD. (2019). *Artificial intelligence in society*. Paris, France: OECD Publishing. https://doi.org/10.1787/eedfee77-en

OECD. (2025). *Trends shaping education 2025*. Paris, France: OECD Publishing. https://doi.org/10.1787/ee6587fd-en

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management*, *42*(1), 15-29. https://doi.org/10.1016/j.im.2003.11.002

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, *69*, 1105–1120. https://doi.org/10.1007/s11528-025-01100-1

Richey, R. C., & Klein, J. D. (2014). *Design and development research: Methods, strategies, and issues*. United Kingdom: Routledge.

Said, N., Meng, C. C., Musa, M., & Yaakob, M. N. (2022). Development of financial education training module for secondary school mathematics teachers: Design and development research. *International Journal of Academic Research in Progressive Education and Development*, *11*(1), 827-836. https://doi.org/10.6007/IJARPED/v11-i1/12342

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, *36*(2), 187-209. https://doi.org/10.1017/S0958344023000265

Taheri, R., Nazemi, N., Pennington, S. E., Clark, J. A., & Dadgostari, F. (2025). Factors influencing educators' AI adoption: A grounded meta-analysis review. *Computers and Education: Artificial Intelligence*, *9*, 100464. https://doi.org/10.1016/j.caeai.2025.100464

Tammets, K., & Ley, T. (2023). Integrating AI tools in teacher professional learning: A conceptual model and illustrative case. *Frontiers in Artificial Intelligence*, *6*, 1255089. https://doi.org/10.3389/frai.2023.1255089

Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, *8*, 100355. https://doi.org/10.1016/j.caeai.2024.100355

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.* Paris, France: UNESCO.

UNESCO. (2023). *Guidance for generative AI in education and research.* Paris, France: UNESCO Publishing.

UNESCO. (2024). *AI competency framework for teachers.* Paris, France: UNESCO.

Van Merriënboer, J. J. G., Kirschner, P. A., & Frèrejean, J. (2024). *Ten steps to complex learning: A systematic approach to four-component instructional design* (4th ed.). UK: Routledge.

Xiao, Y., Surasin, J., & Prabjandee, D. (2020). Development of a training module to improve initial ELT proficiency among student-teachers in multi-ethnic community schools. *Journal of Language and Linguistic Studies*, *16*(1), 366-389. https://doi.org/10.17263/jlls.712849

Yang, C., Chen, J., & Hou, S. (2025). Educational technology in EFL teaching and learning in China: Types, effectiveness, determinants, and challenges. *Asian Journal of Distance Education*, *20*(2), 1-24.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhang, Y., & Dong, C. (2024). Exploring the digital transformation of generative AI-assisted foreign language education: A socio-technical systems perspective based on mixed-methods. *Systems*, *12*(11), 462. https://doi.org/10.3390/systems12110462