International Journal of Education and Practice
2027 Vol. 15, No. 1, pp. 1-18
ISSN(e): 2310-3868
ISSN(p): 2311-6897
DOI: 10.18488/61.v15i1.5359
© 2027 Conscientia Beam. All Rights Reserved.

Comparing Arabic and English instruction in generative AI-supported teacher education: A quasi-experimental study

Mohammad Salman Alkhazaleh[1+]
Robin Kabha[2]
Abhijit Kelkar[3]
Maram Salah Jaradat[4]
Mohammad Awad Alkhawaldeh[5]

[1,4]*College of Education, Humanities and Social Sciences, Al Ain University, Al Ain, United Arab Emirates.*
[1]*Email: mohammad.alkhazaleh@aau.ac.ae*
[4]*Email: maram.jaradat@aau.ac.ae*
[2]*Fujairah University, Fujairah, United Arab Emirates.*
[2]*Email: robinkab@yahoo.co.uk*
[3]*College of Business Administration, Fujairah University, United Arab Emirates.*
[3]*Email: abhijeetsharadkelkar@gmail.com*
[5]*Exegesis and Quranic Sciences, Alqasimia University, Sharjah, United Arab Emirates.*
[5]*Email: malkhawaldeh@alqasimia.ac.ae*



(+ Corresponding author)

Article History
Received: 27 July 2026
Revised: 16 September 2026
Accepted: 12 October 2026
Published: 13 November 2026

Keywords
Arabic-medium instruction
English-medium instruction
Generative artificial intelligence
Language of instruction
Pedagogical competencies
Professional performance
Quasi-experimental design
Teacher preparation.

ABSTRACT

This study examined the effect of the language of instruction in generative artificial intelligence (AI)-supported learning environments on pre-service teachers' pedagogical competencies and professional performance. A quasi-experimental design with two equivalent groups and pre- and post-measurements was used. The sample consisted of 64 professional teaching diploma students divided equally into an Arabic-medium group (n = 32) and an English-medium group (n = 32). Over eight weeks, both groups studied the same pedagogical content and completed comparable generative AI-supported activities; the language of instruction represented the experimental variable. Data were collected using a pedagogical competencies scale and a professional performance observation checklist. Pre-measurement results confirmed equivalence between the groups. Post-measurement findings showed statistically significant improvement in both groups, indicating the usefulness of structured generative AI-supported instruction in teacher preparation. However, the Arabic-medium group achieved higher post-test scores in pedagogical competencies and professional performance, whereas the English-medium group performed better in pedagogical terminology and AI prompt interaction. The findings suggest that Arabic may support deeper contextualized pedagogical understanding and professional application, while English may strengthen access to technical terminology and prompt design. The study recommends a moderated bilingual model that uses Arabic for conceptual and professional application and English for terminology, international resources, and generative AI interaction.

Contribution/Originality: This study is one of the very few studies that have investigated the effect of language of instruction within generative AI-supported teacher education. It provides a quasi-experimental comparison of Arabic- and English-medium instruction and documents their complementary effects on pedagogical competencies, professional performance, terminology development, and AI prompt interaction.

1. INTRODUCTION

Generative artificial intelligence has brought about a visible transformation in teaching and learning environments, particularly within higher education and teacher preparation programs. AI tools are no longer confined to generating text or answering questions; they now support student-teachers in ideation, activity construction,

© 2027 Conscientia Beam. All Rights Reserved.

lesson plan design, classroom situation analysis, feedback provision, and the nurturing of professional reflection. UNESCO has affirmed that the integration of generative AI in education should occur within a clearly articulated pedagogical and humanistic vision, one that safeguards equity, privacy, the role of the teacher, and the quality of learning (Miao & Holmes, 2023).

The significance of generative AI grows considerably within teacher preparation contexts because it offers a flexible training environment in which student-teachers can simulate authentic instructional situations, examine classroom challenges, and develop professional responses grounded in reflective thinking. Contemporary literature indicates that generative AI can support the construction of technological, pedagogical, and content knowledge among student-teachers, yet it requires deliberate and structured use to prevent its reduction to superficial or mechanical application (Mishra, Warr, & Islam, 2023; Wen & Wen, 2024).

Language of instruction constitutes an important variable that may influence the effectiveness of generative AI-supported learning environments. Language is not merely a medium for conveying content; it functions as an instrument for constructing understanding, analyzing meaning, articulating thought, and converting knowledge into practice. Educational literature suggests that learning in a language closer to the learner's experience and cultural context can foster deep comprehension and the ability to link concepts to real-world practice, whereas learning in a global language such as English may support access to terminology, contemporary scholarship, and interaction with global digital tools (Cummins, 2000; UNESCO, 2023).

Within teacher preparation programs specifically, Arabic-medium instruction may help students develop a deep understanding of pedagogical concepts and connect them to local and professional contexts, while English-medium instruction may support professional terminology, engagement with international literature, and interaction with AI tools that frequently operate with greater efficiency in English. The question, therefore, is not which language to select absolutely, but rather which pedagogical function each language can serve within a given learning environment.

Despite the growing body of research on generative AI in education, most studies have focused on general patterns of tool use or on student and teacher attitudes, while quasi-experimental investigations that examine the effect of instruction language inside generative AI environments remain limited, particularly in teacher preparation. This study was designed to address that gap by testing the differential effects of Arabic and English as languages of instruction within a generative AI-supported educational environment.

Accordingly, this study does not treat language as a formal variable, but as an independent experimental variable capable of influencing teacher preparation outcomes. For this reason, a quasi-experimental methodology was adopted, as it permits the examination of this variable's effect in a real instructional setting where full random assignment of students is impractical, while allowing control over factors such as content, duration, AI tools, measurement instruments, and assessment procedures (Creswell & Creswell, 2023; Shadish, Cook, & Campbell, 2002).

1.1. Statement of the Problem

Teacher preparation programs face a growing challenge: how to integrate generative AI in pedagogically meaningful ways that genuinely develop student-teachers' competencies and professional performance. Simultaneously, the language of instruction emerges as a potentially influential factor in students' capacity to comprehend concepts, internalize terminology, engage with AI tools, and translate knowledge into educational practice.

Despite the recognized value of English for accessing global literature and professional vocabulary, Arabic-medium instruction may be more conducive to deepening pedagogical understanding, particularly when professional practice is embedded in an Arabic-speaking local context. Moreover, using generative AI tools within teacher preparation is not simply a matter of tool availability; it is shaped by the language through which the tool is operated, the way prompts are constructed, and the student-teacher's ability to apply the tool's outputs in authentic instructional situations.

© 2027 Conscientia Beam. All Rights Reserved.

This situation reveals the need to test the effect of instruction language within a generative AI-supported learning environment through a quasi-experimental design that permits comparison between two groups that are equivalent at baseline, while controlling for content, duration, tools, and procedures as much as possible. The central research question is as follows: What is the effect of instruction language in generative AI-supported learning environments on the development of pedagogical competencies and professional performance among professional teaching diploma students?

1.2. Research Questions

The study sought to answer the following questions:

1. What is the effect of Arabic-medium instruction within a generative AI-supported learning environment on the development of pedagogical competencies among professional teaching diploma students?
2. What is the effect of English-medium instruction within a generative AI-supported learning environment on the development of pedagogical competencies among professional teaching diploma students?
3. What is the effect of Arabic-medium instruction within a generative AI-supported learning environment on the development of professional performance among professional teaching diploma students?
4. What is the effect of English-medium instruction within a generative AI-supported learning environment on the development of professional performance among professional teaching diploma students?
5. Are there statistically significant differences between the Arabic group and the English group in pedagogical competencies following the intervention?
6. Are there statistically significant differences between the Arabic group and the English group in professional performance following the intervention?
7. Are there differences between the two groups in the dimensions of pedagogical terminology and interaction with generative AI prompts?
8. Which language model appears most suitable for generative AI-supported teacher preparation programs: Arabic-only, English-only, or a bilingual model?

1.3. Research Hypotheses

The study tested the following hypotheses:

1. *There are statistically significant differences at $\alpha \leq 0.05$ between pre- and post-measurement mean scores of the Arabic group in pedagogical competencies, in favor of the post-measurement.*
2. *There are statistically significant differences at $\alpha \leq 0.05$ between the pre- and post-measurement mean scores of the English group in pedagogical competencies, in favor of the post-measurement.*
3. *There are statistically significant differences at $\alpha \leq 0.05$ between the pre- and post-measurement mean scores of the Arabic group in professional performance, in favor of the post-measurement.*
4. *There are statistically significant differences at $\alpha \leq 0.05$ between the pre- and post-measurement mean scores of the English group in professional performance, in favor of the post-measurement scores.*
5. *There are statistically significant differences at $\alpha \leq 0.05$ between the Arabic and English groups in post-measurement pedagogical competencies in favor of the Arabic group.*
6. *There are statistically significant differences at $\alpha \leq 0.05$ between the Arabic and English groups in post-measurement professional performance, favoring the Arabic group.*
7. *There are statistically significant differences between the Arabic and English groups at $\alpha \leq 0.05$ in the dimensions of pedagogical terminology and interaction with generative AI prompts, favoring the English group.*

1.4. Study Objectives

The study aimed to:

© 2027 Conscientia Beam. All Rights Reserved.

1. Measure the effect of Arabic-medium instruction within a generative AI-supported environment on the development of pedagogical competencies.
2. Measure the effect of English-medium instruction within a generative AI-supported environment on the development of pedagogical competencies.
3. Measure the effect of Arabic-medium instruction on the development of professional performance.
4. Measure the effect of English-medium instruction on the development of professional performance.
5. Compare the two groups' post-measurement pedagogical competencies and professional performance.
6. Identify the dimensions in which each language demonstrates an advantage.
7. Propose an applied framework for a bilingual model in generative AI-supported teacher preparation programs.

1.5. Significance of the Study

This study derives its significance from addressing a contemporary intersection of generative artificial intelligence, language of instruction, and teacher preparation outcomes. It goes beyond describing the relationship among these variables by employing a quasi-experimental design that quantifies the effect of instruction language in a real-world learning setting.

Theoretically, the study enriches the literature on generative AI in teacher education by demonstrating that the effectiveness of such tools cannot be separated from the language through which the educational experience is delivered. It also contributes to clarifying the relationship between language of instruction and the construction of professional competencies in student teachers.

Practically, the study assists professional teaching diploma programs in making more precise pedagogical decisions regarding language of instruction, AI integration strategies, professional activity design, and assessment procedures. Its findings can also support decision-makers in teacher preparation institutions in adopting bilingual instructional models that combine the depth of local understanding with openness to global knowledge.

1.6. Theoretical Framework

The study is grounded in the view that learning with generative AI is shaped by the interaction among technological affordances, pedagogical design, and the language through which learners interpret professional knowledge. In this sense, the TPACK perspective explains how AI tools become educationally meaningful only when they are integrated with content and pedagogy, rather than used as isolated technical applications.

The study also draws on bilingual and translanguaging perspectives, which regard language as a cognitive and sociocultural mediator of learning. Arabic-medium instruction may support contextualized meaning-making and professional reflection in local school settings, whereas English-medium instruction may facilitate access to international terminology, scholarly sources, and prompt construction practices used by many AI systems. This framework supports examining language of instruction as an active experimental variable within AI-supported teacher preparation.

1.7. Review of Related Literature

Recent scholarship on generative AI in education presents a consistent but qualified picture. Reviews show that tools such as ChatGPT can support explanation, feedback, content generation, planning, and learner engagement, while also raising concerns about accuracy, over-reliance, assessment validity, and academic integrity (Albadarin, Saqr, Pope, & Tukiainen, 2024; Cotton, Cotton, & Shipway, 2024; Farrokhnia, Banihashem, Noroozi, & Wals, 2024; Lo, 2023; Tlili et al., 2023). Evidence from higher education likewise suggests that the educational value of generative AI depends less on access to the tool itself than on purposeful task design, teacher guidance, and verification of outputs (Baig & Yadegaridehkordi, 2024; Bhullar, Joshi, & Chugh, 2024; McGrath, Farazouli, & Cerratto-Pargman, 2025).

© 2027 Conscientia Beam. All Rights Reserved.

This pattern supports the present study's treatment of AI as a structured pedagogical environment rather than as an isolated technological intervention. Additional reviews and conceptual analyses reach similar conclusions, emphasizing both the instructional opportunities of generative AI and the need to address reliability, academic integrity, accessibility, and responsible implementation (Alasadi & Baiz, 2023; Ali et al., 2024; Montenegro-Rueda, Fernández-Cerero, Fernández-Batanero, & López-Meneses, 2023).

Teacher-education research narrows this issue further. Generative AI has been used to make pedagogical reasoning visible, support lesson and unit design, and stimulate reflection on pedagogical content knowledge among pre-service teachers (Blonder, Feldman-Maggor, & Rap, 2025; Li, Ji, & Zhan, 2024; Wen & Wen, 2024). At the same time, these studies indicate that benefits are contingent on the learner's ability to evaluate AI output and connect it to disciplinary and classroom knowledge. Thus, the literature supports examining observable pedagogical competencies and professional performance, rather than relying only on attitudes toward AI or self-reported frequency of use.

Language of instruction introduces a second body of evidence that has rarely been connected experimentally with generative AI. Research on English-medium instruction in higher education shows mixed effects on content learning but more consistent advantages for English-language development and disciplinary vocabulary; outcomes also depend on prior proficiency, instructional support, and context (Alqarni, Mahdi, Ali, & Curle, 2024; Feng, He, & Yiganmu, 2023; Macaro, Curle, Pun, An, & Dearden, 2018; Rose et al., 2026). Translanguaging research, in contrast, emphasizes the pedagogical value of drawing strategically on learners' full linguistic repertoires rather than treating languages as mutually exclusive instructional systems (Tang, Rousse-Malpat, & Duarte, 2026; Wang, Zaid, Ibrahim, & Haladin, 2024). Together, these findings suggest that Arabic and English may contribute differently to conceptual understanding, professional application, terminology, and access to international resources.

A related emerging literature concerns prompt competence. Studies of generative-AI use increasingly treat prompt construction as a learnable academic skill rather than a purely technical act. Systematic and experimental evidence indicates that explicit attention to prompting can improve the quality of human-AI interaction, although outcomes vary by task and by learners' evaluative skills (Jin & Sercu, 2025; Lee & Palmer, 2025). Research also shows that ChatGPT can complement selected forms of assessment and feedback, but should not replace professional judgment (Lu et al., 2024; Meyer et al., 2023). The unresolved issue, therefore, is not simply whether generative AI is useful, but whether the language through which learners study and interact with AI changes what they learn and how effectively they translate that learning into professional practice. This quasi-experimental comparison addresses this gap.

1.8. Contribution and Originality of the Study

The originality of this study lies in its quasi-experimental examination of the language of instruction within generative AI-supported teacher preparation, an area that has received limited empirical attention. Rather than treating AI use as a general technological intervention, the study isolates language of instruction as the main experimental variable and compares its effect on both pedagogical competencies and observable professional performance.

The study contributes to the teacher education literature by showing that Arabic and English may serve complementary functions in AI-supported learning environments: Arabic may strengthen conceptual understanding and professional application, while English may support terminology development and prompt interaction. These findings provide a practical basis for a moderated bilingual model in professional teaching diploma programs.

More specifically, the study connects three strands that are usually examined separately, generative AI integration, language of instruction, and performance-based teacher preparation, and tests their intersection under controlled instructional conditions. This combination constitutes the study's principal empirical contribution.

© 2027 Conscientia Beam. All Rights Reserved.

1.9. Operational Definitions

Generative Artificial Intelligence: A set of digital models and applications capable of producing new content in the form of texts, ideas, activities, images, or feedback, based on data and patterns from training. UNESCO defines generative AI as technology that can produce new content usable in education and research, provided it is governed ethically and pedagogically in ways that account for privacy, equity, and learning quality Miao and Holmes (2023). Mishra et al. (2023) note that generative AI calls for a reconceptualization of the relationship between technological, pedagogical, and content knowledge in instructional practice. Operationally, in this study, generative AI refers to the use of AI tools in lesson planning, activity generation, classroom situation analysis, educational feedback, and professional reflection support among teaching diploma students.

Language of Instruction: The language used to deliver educational content, manage classroom interaction, explain concepts, build activities, and guide learners toward achieving learning outcomes. Educational literature underscores that the language of instruction is not a neutral conduit for knowledge, but actively shapes meaning construction, understanding, and the learner's capacity for expression, analysis, and application (Cummins, 2000; García & Li, 2014). Operationally, it refers to the language used in teaching pedagogical content and interacting with generative AI tools during program implementation, encompassing two conditions: Arabic-medium instruction and English-medium instruction.

Pedagogical Competencies: The cluster of knowledge, skills, and dispositions a teacher requires to perform the instructional role effectively, encompassing planning, teaching strategies, classroom management, assessment, differentiation, professional reflection, and technology integration. Pedagogical competencies rest on the understanding that teacher preparation extends beyond theoretical knowledge to the ability to translate that knowledge into effective instructional practice (Darling-Hammond, 2017; Mishra & Koehler, 2006). Operationally, this denotes the score obtained by the student-teacher on the pedagogical competencies scale developed for this study, covering dimensions of planning, teaching, classroom management, assessment, differentiation, professional reflection, and generative AI application.

Professional Performance: The student-teacher's capacity to translate pedagogical knowledge into observable instructional behavior in educational situations, including objective clarity, content organization, activity implementation, interaction management, feedback delivery, and classroom situation handling. Professional performance is a key indicator for judging teacher preparation quality because it reflects the student-teacher's ability to transition from theoretical understanding to actual classroom practice (Danielson, 2013; Darling-Hammond, 2017). Operationally, this refers to the score on the professional performance observation checklist developed for this study, assessing objective clarity, content organization, activity implementation, classroom interaction management, feedback use, classroom situation handling, and AI integration for performance improvement.

Quasi-Experimental Design: A research approach that tests the effect of an independent variable on dependent variables within a real-world setting without full random assignment of participants, as in true experiments, while employing control procedures such as pre-measurement, treatment standardization, measurement instrument uniformity, and group comparison. This design is well suited to educational studies where complete experimental control or random assignment is not feasible (Creswell & Creswell, 2023; Shadish et al., 2002). Operationally, it denotes the design used to compare two existing groups—one studying content in Arabic, the other studying the same content in English, with a competency scale and a performance checklist administered pre- and post-intervention to determine the effect of the language of instruction within a generative AI-supported learning environment.

1.10. Study Delimitations

The study's findings are bounded by the following delimitations:

- Population: The study was limited to professional teaching diploma students.

© 2027 Conscientia Beam. All Rights Reserved.

- Content scope: The study addressed the effect of the language of instruction within generative AI-supported environments on pedagogical competencies and professional performance.
- Methodological scope: A quasi-experimental design with two equivalent groups and pre- and post-measurements was used.
- Temporal scope: The program was implemented over eight academic weeks with two sessions per week, totaling sixteen training and instructional sessions.
- Instrumentation: The study relied on a pedagogical competencies scale and a professional performance observation checklist.
- Technological scope: Generative AI tools were employed in lesson planning, activity generation, scenario construction, and feedback provision.

2. RESEARCH METHODOLOGY

The study employed a quasi-experimental design, one of the most appropriate approaches for educational research aiming to measure the effect of an independent variable on dependent variables within an authentic instructional context. This methodology was applied to examine the effect of the language of instruction in generative AI-supported learning environments on the development of pedagogical competencies and professional performance among professional teaching diploma students.

Unlike fully experimental designs, this approach does not require complete random assignment; instead, it works with intact or near-intact instructional groups, applying systematic control procedures such as pre-measurement, content standardization, duration uniformity, instrument standardization, and task equivalence (Shadish et al., 2002).

Rationale for the quasi-experimental approach: This design was selected for several methodological reasons: the study aimed to measure the effect of an independent variable, language of instruction within a generative AI-supported environment, on two dependent variables; the study was conducted in an authentic educational setting where full random assignment was impractical; the design allowed comparison between two existing groups; it permitted baseline equivalence verification via pre-measurement; post-measurement enabled quantification of change following the intervention; and controlling for content, duration, AI tools, and measurement instruments reduced the influence of extraneous variables and linked post-intervention differences more closely to the instruction language variable.

2.1. Quasi-Experimental Design

The study adopted a non-randomized equivalent groups design with pre- and post-measurements, structured as follows: Table 1 presents the quasi-experimental design and measurement schedule used for the two instructional groups.

Table 1. Quasi-Experimental Design and Measurement Schedule.

Group	Pre-Measurement	Experimental Treatment	Post-Measurement
Arabic Group	Competencies Scale + Performance Checklist	Content instruction in Arabic within a generative AI-supported environment	Competencies Scale + Performance Checklist
English Group	Competencies Scale + Performance Checklist	Same content instruction in English within a generative AI-supported environment	Competencies Scale + Performance Checklist

Both groups were measured prior to the program, then received their respective treatments, and then underwent post-measurement. The primary distinction between the groups was the language of instruction; all other elements of the experiment were standardized to the greatest extent possible.

© 2027 Conscientia Beam. All Rights Reserved.

2.2. Population and Sample

The study population consisted of professional teaching diploma students enrolled in teacher preparation programs. The sample comprised 64 students distributed equally into two groups. Table 2 displays the distribution of the sample according to the language of instruction and treatment.

Table 2. Sample Distribution by Instruction Language and Treatment.

Group	Language	N	Treatment
Group One	Arabic	32	Generative AI-supported learning in Arabic
Group Two	English	32	Generative AI-supported learning in English
Total	—	64	—

The sample was selected purposively, given the study's requirement for two groups studying identical content in two different languages. Baseline equivalence was verified for both pedagogical competencies and professional performance prior to program implementation.

The data were collected from professional teaching diploma students enrolled in the College of Education, Humanities and Social Sciences at Al Ain University, United Arab Emirates. Participants were enrolled in comparable course sections and studied the same pedagogical content during the implementation period.

Because the study was conducted in a bilingual teacher preparation context, participants had prior exposure to both Arabic and English through their academic program. Group allocation followed the available instructional sections: one section received the Arabic-medium version of the intervention, and the other received the English-medium version. Full random assignment was therefore not possible; however, pre-measurement equivalence was statistically verified before the intervention.

Participants' language readiness was checked through program records and pre-intervention orientation activities to ensure that students could participate in the assigned instructional language. Any differences in language proficiency were treated as an important contextual factor when interpreting the findings.

2.3. Study Variables

The study variables were organized into independent, dependent, and control variables. Language of instruction within the generative AI-supported learning environment served as the independent variable, while pedagogical competencies and professional performance represented the dependent variables. Educational content, program duration, AI tools, and measurement instruments were controlled across both groups. Table 3 summarizes the study variables and their respective levels and dimensions.

Table 3. Study Variables and Dimensions.

Variable Type	Variables	Levels / Dimensions
Independent Variable	Language of instruction within a generative AI-supported learning environment	Arabic / English
Dependent Variable	Pedagogical competencies	Planning, teaching, classroom management, assessment, differentiation, reflection, AI application
Dependent Variable	Professional performance	Objective clarity, content organization, activity implementation, interaction management, feedback, classroom situations
Control Variable	Educational content	Identical content for both groups
Control Variable	Program duration	Identical duration for both groups
Control Variable	AI tools	Identical tools for both groups
Control Variable	Measurement instruments	Same scale and checklist for both groups

2.4. Quasi-Experimental Treatment

The treatment consisted of teaching a unified pedagogical curriculum to both groups within a generative AI-supported learning environment, with language of instruction as the sole differentiating factor. The program ran for

© 2027 Conscientia Beam. All Rights Reserved.

eight academic weeks with two sessions per week, totaling sixteen sessions. It covered pedagogical and professional topics related to lesson planning, teaching strategies, classroom management, assessment, differentiation, professional reflection, and the integration of generative AI into teaching practice.

Arabic Group: The group studied content in Arabic and used generative AI tools in Arabic for simplifying pedagogical concepts, generating classroom examples from the local school context, building lesson plans in Arabic, analyzing authentic classroom situations, receiving feedback on performance and planning, and writing professional reflections in Arabic.

The English Group studied the same content in English and used generative AI tools in English for understanding English pedagogical terminology, constructing directed prompts for AI tools, generating instructional activities in English, analyzing classroom scenarios in English, receiving professional feedback in English, and accessing examples and concepts from international literature.

2.5. Experimental Control Procedures

To enhance the rigor of the quasi-experimental design and minimize external influences, the following control procedures were implemented: standardization of educational content across both groups; uniform program duration of eight weeks; an equal number of sessions (sixteen) for each group; identical generative AI tools for both groups; identical measurement instruments applied to both groups pre- and post-intervention; pre-measurement to verify baseline equivalence; standardized instructional tasks in planning, situation analysis, activity design, and professional reflection; designation of the language of instruction as the sole differing variable; unified performance assessment criteria for both groups; and training of observers on the checklist prior to implementation to reduce rating bias.

2.6. Research Instruments

Pedagogical Competencies Scale: The scale was developed through a review of the literature on teacher preparation, professional competencies, and technology and AI integration in education, drawing on works by Darling-Hammond (2017); Mishra and Koehler (2006); and Mishra et al. (2023). The final version comprised 35 items distributed across seven dimensions. Table 4 presents the dimensions and number of items included in the Pedagogical Competencies Scale.

Table 4. Dimensions of the Pedagogical Competencies Scale.

Dimension	Number of Items
Lesson planning	5
Teaching strategies	5
Classroom management	5
Educational assessment	5
Differentiation	5
Professional reflection	5
Generative AI application	5
Total	35

Items were rated on a five-point Likert scale from 1 to 5, with higher scores indicating stronger pedagogical competencies.

Professional Performance Observation Checklist: The checklist was developed in light of the literature on teacher performance evaluation and effective instructional practices, including the (Danielson, 2013) teaching evaluation framework and contemporary literature on AI applications in planning, assessment, and feedback (Lee & Moore, 2024; Ogunleye, Zakariyyah, Ajao, Olayinka, & Sharma, 2024). It comprised 28 items across seven dimensions. Table 5 outlines the dimensions and number of items included in the Professional Performance Observation Checklist.

© 2027 Conscientia Beam. All Rights Reserved.

Table 5. Dimensions of the Professional Performance Observation Checklist.

Dimension	Number of Items
Clarity of learning objectives	4
Content organization	4
Instructional activity implementation	4
Classroom interaction management	4
Feedback use	4
Handling classroom situations	4
AI application for performance improvement	4
Total	28

The checklist was used to rate student-teachers' professional performance in applied tasks and micro-teaching activities on a five-point rating scale ranging from very low to very high.

2.7. Validity and Reliability of Instruments

Validity: Both instruments were submitted to a panel of nine specialists in curriculum and instruction, educational technology, and measurement and evaluation. Reviewers assessed item clarity, alignment with study objectives, dimensional relevance, appropriateness for the study sample, and linguistic accuracy. Items meeting the 80% agreement threshold were retained. Based on reviewers' feedback, some items were reworded, overlapping items were merged, and items that failed to reach the required agreement threshold were removed. The instruments in their final form were deemed appropriate for administration to the study sample.

Reliability: Internal consistency was calculated using Cronbach's alpha. Results were as follows:

Table 6 reports the reliability coefficients of the research instruments.

Table 6. Reliability Coefficients of the Research Instruments.

Instrument	Items	Cronbach's Alpha
Pedagogical Competencies Scale	35	0.91
Professional Performance Checklist	28	0.89
Generative AI Application dimension	5	0.87

These values indicate a high degree of internal consistency, rendering the instruments suitable for measurement purposes in this study.

Inter-Rater Reliability: Because the professional performance checklist relies on the assessment of observable behavior, two observers were trained on the instrument and its rating criteria and then applied it to a pilot sample of micro-teaching performances. The inter-rater agreement coefficient reached 0.86, an educationally acceptable value indicating a satisfactory level of consistency in professional performance ratings. Discrepancies between raters were discussed prior to final administration to ensure a shared operational understanding of the checklist indicators.

2.8. Study Procedures

The study was implemented through the following steps.

1. Identification of the pedagogical content to be taught to both groups.
2. Preparation of Arabic and English versions of the same content, preserving unified objectives, concepts, and activities.
3. Selection of generative AI tools to be used in the implementation.
4. Training students in both groups on how to use AI tools for lesson planning, activity generation, situation analysis, and feedback acquisition.
5. Administration of the pedagogical competencies scale and professional performance checklist as pre-measurements for both groups.

© 2027 Conscientia Beam. All Rights Reserved.

6. Implementation of the Arabic-medium program for the Arabic group over eight weeks.
7. Implementation of the English-medium program for the English group over eight weeks.
8. Post-administration of the competency scale and performance checklist.
9. Data entry and statistical analysis.
10. Interpretation of results in light of the study's questions and hypotheses.

Before data collection, verbal institutional permission to conduct the study was obtained through the Dean of the College of Education, Humanities and Social Sciences at Al Ain University, in accordance with the College's internal procedure for coordinating and facilitating educational research. Informed verbal consent was subsequently obtained from all participants. The detailed ethical approval statement is provided in the declarations section at the conclusion.

2.9. Statistical Analyses

The following statistical procedures were employed.

- Descriptive statistics (means and standard deviations) were used to characterize each group's performance.
- An independent-samples t-test was used to verify baseline equivalence between the groups.
- A paired-samples t-test was used to examine within-group differences between pre- and post-measurements.
- An independent-samples t-test was used to examine between-group differences at post-measurement.
- Cronbach's alpha was used to assess instrument reliability.
- An inter-rater agreement coefficient was used to assess checklist reliability.
- Cohen's d effect size was interpreted in accordance with Cohen (1988) to determine the magnitude of the experimental treatment's effect.

3. RESULTS AND ANALYSIS

Prior to addressing the research questions, baseline equivalence between the two groups was verified for both pedagogical competencies and professional performance to ensure that any post-intervention differences could be attributed, to a meaningful degree, to the quasi-experimental treatment, namely, language of instruction within a generative AI-supported environment. Table 7 presents the baseline equivalence results for the Arabic and English groups before the intervention.

Table 7. Baseline Equivalence Between Arabic and English Groups Before the Intervention.

Variables	Group	N	Mean	SD	t	p
Pedagogical Competencies	Arabic	32	3.01	0.41	0.39	0.697
Pedagogical Competencies	English	32	2.97	0.43		
Professional Performance	Arabic	32	2.94	0.38	0.45	0.654
Professional Performance	English	32	2.90	0.41		

The significance values of 0.697 for pedagogical competencies and 0.654 for professional performance both exceeded the 0.05 threshold, confirming the absence of statistically significant pre-intervention differences. The groups were therefore equivalent at baseline, strengthening the internal validity of the quasi-experimental design and supporting the interpretation of post-intervention differences as attributable to the treatment variable.

3.1. Research Question One

What is the effect of Arabic-medium instruction within a generative AI-supported learning environment on the development of pedagogical competencies among professional teaching diploma students? Table 8 reports the paired-samples t-test results for the Arabic-medium group's pedagogical competencies.

© 2027 Conscientia Beam. All Rights Reserved.

Table 8. Paired-Samples t-Test Results for Arabic-Medium Instruction and Pedagogical Competencies.

Variables	Measurement	Mean	SD	t	p	Cohen's d
Pedagogical Competencies	Pre	3.01	0.41	11.82	0.000	1.91
Pedagogical Competencies	Post	4.21	0.38			

The results indicate statistically significant differences between pre- and post-measurements in the Arabic group's pedagogical competencies. The mean increased from 3.01 to 4.21, with $t = 11.82$ and $p < 0.001$. Cohen's $d = 1.91$ indicates a large within-group effect.

3.2. Research Question Two

What is the effect of English-medium instruction within a generative AI-supported learning environment on the development of pedagogical competencies among professional teaching diploma students? Table 9 reports the paired-samples t-test results for the English-medium group's pedagogical competencies.

Table 9. Paired-Samples t-Test Results for English-Medium Instruction and Pedagogical Competencies.

Variables	Measurement	Mean	SD	t	p	Cohen's d
Pedagogical Competencies	Pre	2.97	0.43	8.74	0.000	1.46
Pedagogical Competencies	Post	3.91	0.42			

The results show statistically significant differences between pre- and post-measurements in the English group's pedagogical competencies. The mean increased from 2.97 to 3.91, with $t = 8.74$ and $p < 0.001$. Cohen's $d = 1.46$ indicates a large within-group effect. The post-measurement score was therefore significantly higher than the pre-measurement score.

3.3. Research Question Three

What is the effect of Arabic-medium instruction within a generative AI-supported learning environment on the development of professional performance among professional teaching diploma students? Table 10 presents the paired-samples t-test results for Arabic-medium instruction and professional performance.

Table 10. Paired-Samples t-Test Results for Arabic-Medium Instruction and Professional Performance.

Variables	Measurement	Mean	SD	t	p	Cohen's d
Professional Performance	Pre	2.94	0.38	12.36	0.000	2.02
Professional Performance	Post	4.18	0.36			

The results demonstrate statistically significant differences between pre- and post-measurements in the Arabic group's professional performance. The mean increased from 2.94 to 4.18, with $t = 12.36$ and $p < 0.001$. Cohen's $d = 2.02$ indicates a very large within-group effect.

3.4. Research Question Four

What is the effect of English-medium instruction within a generative AI-supported learning environment on the development of professional performance among professional teaching diploma students? Table 11 presents the paired-samples t-test results for English-medium instruction and professional performance.

© 2027 Conscientia Beam. All Rights Reserved.

Table 11. Paired-Samples t-Test Results for English-Medium Instruction and Professional Performance.

Variables	Measurement	Mean	SD	t	p	Cohen's d
Professional Performance	Pre	2.90	0.41	9.16	0.000	1.53
Professional Performance	Post	3.88	0.40			

Results indicate statistically significant differences between pre- and post-measurements in the English group's professional performance. The mean increased from 2.90 to 3.88, with $t = 9.16$ and $p < 0.001$. Cohen's $d = 1.53$ indicates a large within-group effect.

3.5. Research Question Five

Are there statistically significant differences between the Arabic and English groups in pedagogical competencies following the intervention? Table 12 displays the post-intervention differences between the Arabic and English groups in pedagogical competencies.

Table 12. Post-Intervention Differences Between Groups in Pedagogical Competencies.

Variables	Group	N	Mean	SD	t	p	Direction
Pedagogical Competencies	Arabic	32	4.21	0.38	2.99	0.004	Favors Arabic
Pedagogical Competencies	English	32	3.91	0.42			

Statistically significant post-measurement differences were found between the groups in overall pedagogical competencies, with the Arabic group ($M = 4.21$, $SD = 0.38$) scoring higher than the English group ($M = 3.91$, $SD = 0.42$), $t = 2.99$, $p = 0.004$. The direction of the observed difference therefore favored the Arabic group.

3.6. Research Question Six

Are there statistically significant differences between the Arabic and English groups in professional performance following the intervention? Table 13 displays the post-intervention differences between the Arabic and English groups in professional performance.

Table 13. Post-Intervention Differences Between Groups in Professional Performance.

Variables	Group	N	Mean	SD	t	p	Direction
Professional Performance	Arabic	32	4.18	0.36	3.13	0.003	Favors Arabic
Professional Performance	English	32	3.88	0.40			

Statistically significant post-measurement differences were found between the groups in professional performance, with the Arabic group ($M = 4.18$, $SD = 0.36$) scoring higher than the English group ($M = 3.88$, $SD = 0.40$), $t = 3.13$, $p = 0.003$. The direction of the observed difference therefore favored the Arabic group.

3.7. Research Question Seven

Are there differences between the groups in the dimensions of pedagogical terminology and interaction with generative AI prompts? Table 14 exhibits the group differences in pedagogical terminology and generative AI prompt interaction.

Table 14. Group Differences in Pedagogical Terminology and Generative AI Prompt Interaction.

Dimension	Group	N	Mean	SD	t	p	Cohen's d	Direction
AI Prompt Interaction	Arabic	32	3.87	0.44	-2.30	0.025	0.58	Favors English
AI Prompt Interaction	English	32	4.12	0.43				
Pedagogical Terminology	Arabic	32	3.93	0.42	-2.21	0.031	0.55	Favors English
Pedagogical Terminology	English	32	4.16	0.41				

© 2027 Conscientia Beam. All Rights Reserved.

The results reveal statistically significant differences favoring the English group in both dimensions. For AI prompt interaction, the English group (M = 4.12, SD = 0.43) scored higher than the Arabic group (M = 3.87, SD = 0.44). When the statistic was defined as Arabic minus English, t = -2.30, p = 0.025, d = 0.58. For pedagogical terminology, the English group (M = 4.16, SD = 0.41) scored higher than the Arabic group (M = 3.93, SD = 0.42), t = -2.21, p = 0.031, d = 0.55. The negative signs are consistent with the group order shown in Table 14 and with higher English-group means.

3.8. Research Question Eight

Which language model appears most suitable for generative AI-supported teacher preparation programs?

In light of the preceding findings, a bilingual model appears to offer a balanced and context-sensitive framework for generative AI-supported teacher preparation. The Arabic group's higher scores in overall pedagogical competencies and professional performance highlight the potential value of Arabic for deepening educational understanding and professional application, while the English group's advantage in terminology and AI prompt interaction highlights the value of English for engaging with global scholarship and digital tools.

Combining both languages may therefore represent a balanced approach: Arabic can be used to explain foundational pedagogical concepts, analyze classroom situations, and articulate professional reflections, while English is leveraged for terminology, international references, prompt construction, and technical engagement with generative AI tools.

4. DISCUSSION

The findings confirm that the quasi-experimental design was well suited to examining the effect of instruction language in an authentic generative AI-supported environment. Pre-measurement verified group equivalence, and post-measurement quantified the change attributable to the treatment. Confirmed baseline equivalence is a critical indicator of quasi-experimental design quality, as it reduces the likelihood that post-intervention differences stem from pre-existing group differences.

Both groups demonstrated improvement following the intervention, indicating that the structured AI-supported learning activities were associated with gains in the measured outcomes. This pattern is compatible with the wider literature, which reports potential benefits of generative AI for explanation, feedback, planning, and learner support while emphasizing the need for human oversight and purposeful pedagogical design (Albadarin et al., 2024; Baig & Yadegaridehkordi, 2024; Farrokhnia et al., 2024; Lo, 2023; McGrath et al., 2025; Tlili et al., 2023). Because the present study did not include a no-AI control group, however, the improvement should not be interpreted as a causal estimate of AI use alone; the principal experimental comparison concerns the language of instruction within the shared AI-supported environment.

The Arabic group's higher post-test scores in overall pedagogical competencies and professional performance may reflect the linguistic and contextual proximity of Arabic to the participants' professional setting. One plausible interpretation is that working through Arabic reduced the linguistic distance between pedagogical concepts and the classroom situations in which those concepts were applied. This interpretation should be treated cautiously because language proficiency was not manipulated independently and the groups were not randomly assigned. It is nevertheless consistent with bilingual and translanguaging perspectives that view learners' familiar linguistic resources as potential supports for meaning-making and professional reflection (García & Li, 2014; Tang et al., 2026; Wang et al., 2024).

The result therefore does not establish that Arabic is inherently superior for teacher education. Rather, within this sample and instructional context, Arabic was associated with stronger scores on outcomes requiring contextualized pedagogical application. Research on English-medium instruction similarly cautions that content

© 2027 Conscientia Beam. All Rights Reserved.

outcomes are shaped by prior proficiency, disciplinary demands, and instructional support, not by language choice alone (Feng et al., 2023; Macaro et al., 2018; Rose et al., 2026).

Conversely, the English group's higher scores in pedagogical terminology and AI prompt interaction may be related to greater exposure to English-language technical vocabulary and to the English-dominant documentation and examples surrounding many generative-AI tools. This interpretation is consistent with research showing that English-medium instruction can support disciplinary vocabulary and with emerging work that treats prompt construction as a specific form of AI literacy (Alqarni et al., 2024; Lee & Palmer, 2025). Teacher-education studies also suggest that generative AI can support planning and pedagogical reasoning when learners are able to formulate, evaluate, and refine interactions with the system (Blonder et al., 2025; Li et al., 2024; Wen & Wen, 2024).

These findings collectively indicate that the choice between Arabic and English should not be absolute, since each language serves a distinct pedagogical function within an AI-supported learning environment. Arabic was more strongly associated with deep understanding and professional application; English was more closely linked to terminology and technical engagement with AI tools. The study's results therefore suggest considering a moderated bilingual model that combines the strengths of both languages. They further affirm that integrating generative AI into teacher preparation should not be treated merely as the introduction of a technical tool into the classroom, but must be accompanied by a clear pedagogical design, appropriate language selection, applied tasks, and assessment criteria that measure actual competencies and performance. Success in AI integration depends on how the tool is used, not merely whether it is present.

5. INTERNAL AND EXTERNAL VALIDITY

Internal Validity: The study strengthened internal validity through several methodological procedures: pre-measurement to verify group equivalence; standardized educational content across groups; uniform duration; identical AI tools; consistent measurement instruments; designation of language of instruction as the primary independent variable; and appropriate statistical tests for within- and between-group comparisons. These measures reduce the influence of confounding variables and increase confidence that post-intervention differences are substantially attributable to differences in language of instruction. The availability of both pre- and post-measurements provided a methodological basis for understanding the magnitude of change following the experimental treatment.

External Validity: The findings may be cautiously generalized to comparable teacher preparation programs operating in bilingual or multilingual environments, particularly those integrating generative AI into instruction and professional training. Generalizability nonetheless remains conditional on similarity in educational context, student characteristics, program design, and participants' linguistic and digital proficiency levels. Results are likely most transferable to educational settings where students use Arabic for professional comprehension and interaction, and English for terminology, references, and digital applications, and should therefore be interpreted within the contextual boundaries of the current study.

6. CONCLUSION

Through a quasi-experimental design based on two statistically equivalent groups and paired pre- and post-measurements, this study found that Arabic-medium instruction within a generative AI-supported learning environment was associated with greater gains in pedagogical competencies and professional performance among professional teaching diploma students than English-medium instruction.

Nevertheless, the English group showed an advantage in the dimensions of pedagogical terminology and AI prompt interaction. The most significant takeaway, therefore, is not an absolute preference for Arabic or English, but rather the potential usefulness of a moderated bilingual model that employs Arabic to deepen understanding and professional application while employing English to reinforce terminology, provide access to global scholarship, and enable technical engagement with AI tools.

© 2027 Conscientia Beam. All Rights Reserved.

The study further suggests that generative AI can serve as an effective instrument in teacher preparation, provided it is embedded within a structured pedagogical design, and that teacher preparation programs must attend carefully to the relationship among language, understanding, application, and professional competencies.

7. RECOMMENDATIONS

1. Consider adopting a moderated bilingual model in professional teaching diploma programs, using Arabic to deepen pedagogical understanding and professional application, and English to reinforce terminology and access to global literature.
2. Use Arabic for explaining foundational pedagogical concepts and connecting them to local school contexts, particularly for topics requiring classroom situation analysis and professional problem-solving.
3. Use English for professional terminology, international references, and AI prompt construction to enhance students' capacity to engage with contemporary digital tools.
4. Train student-teachers in the deliberate pedagogical use of generative AI tools, extending beyond text production to planning, assessment, situation analysis, and professional reflection.
5. Design quasi-experimental activities within teacher preparation programs that measure the effect of AI on actual performance rather than solely on attitudes or perceptions.
6. Develop pedagogical competency measurement instruments that include a dedicated dimension for generative AI application in teaching.
7. Train faculty members to design instructional activities that account for the relationship among language of instruction, AI, and professional competencies.
8. Strengthen student-teachers' capacity to construct effective prompts for generative AI tools in both Arabic and English.
9. Account for cultural and linguistic context when integrating AI into teacher preparation programs, so that tool use does not become a technical practice disconnected from pedagogical objectives.
10. Conduct follow-up studies using alternative quasi-experimental designs, such as multiple-group designs, repeated-measures designs, or comparisons across varying levels of linguistic and digital proficiency.

8. DIRECTIONS FOR FUTURE RESEARCH

The study proposes the following areas for subsequent investigation.

- Examining the effect of bilingual instructional models on student-teacher competency development within generative AI-supported environments.
- Comparing the effects of Arabic- and English-medium generative AI use across different pedagogical specializations.
- Investigating the relationship between students' language proficiency levels and the effectiveness of generative AI tool use.
- Analyzing the effect of prompt engineering on student-teachers' professional performance development.
- Conducting longitudinal studies to measure the effect of generative AI use on student-teachers' performance during fieldwork placements.
- Exploring the effect of generative AI on the development of professional reflection among student-teachers.
- Comparing the effectiveness of monolingual and bilingual instructional approaches in teacher preparation programs.
- Constructing a proposed training model for integrating generative AI into the development of planning, assessment, and classroom management competencies.

© 2027 Conscientia Beam. All Rights Reserved.

Funding: This study received no specific financial support.
Institutional Review Board Statement: The Ethical Committee of Al Ain University, United Arab Emirates, has granted approval for this study on 2 January 2026 (Ref. No. CEHSS-2026-56).
Informed Consent Statement: Informed verbal consent was obtained from all participants, and all data were anonymized to protect participant confidentiality.
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: The authors used OpenAI's ChatGPT (GPT-5.6 Sol) to support language editing, refinement of academic wording, organization of selected sections, and consistency checking during revision. The tool was not used to generate or alter study data, statistical analyses, or empirical results. All AI-assisted outputs were reviewed, verified, and revised by the authors, who remain fully responsible for the content of the manuscript.

REFERENCES

Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education, 100*(8), 2965-2971. https://doi.org/10.1021/acs.jchemed.3c00323

Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education, 3*(1), 60. https://doi.org/10.1007/s44217-024-00138-2

Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences, 14*(6), 643. https://doi.org/10.3390/educsci14060643

Alqarni, O. M., Mahdi, H. S., Ali, J. K. M., & Curle, S. (2024). English medium instruction in Saudi Arabia: A systematic review. *Language Teaching Research Quarterly, 42*, 21-37. https://doi.org/10.32038/ltrq.2024.42.02

Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research, 127*, 102411. https://doi.org/10.1016/j.ijer.2024.102411

Bhullar, P. S., Joshi, M., & Chugh, R. (2024). ChatGPT in higher education-a synthesis of the literature and a future research agenda. *Education and Information Technologies, 29*(16), 21501-21522. https://doi.org/10.1007/s10639-024-12723-x

Blonder, R., Feldman-Maggor, Y., & Rap, S. (2025). Are they ready to teach? Generative AI as a means to uncover pre-service science teachers' PCK and enhance their preparation program. *Journal of Science Education and Technology, 34*(6), 1301-1310. https://doi.org/10.1007/s10956-024-10180-2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire.* Clevedon, England: Multilingual Matters.

Danielson, C. (2013). *The framework for teaching evaluation instrument: 2013 edition.* Princeton, NJ: The Danielson Group.

Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education, 40*(3), 291-309. https://doi.org/10.1080/02619768.2017.1315399

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(3), 460-474. https://doi.org/10.1080/14703297.2023.2195846

Feng, L., He, L., & Yiganmu, A. (2023). Determinants of students' academic success in English as a medium of instruction (EMI) classes: A systematic review. *Heliyon, 9*(10), e20421. https://doi.org/10.1016/j.heliyon.2023.e20421

García, O., & Li, W. (2014). *Translanguaging: Language, bilingualism and education.* London, England: Palgrave Macmillan.

Jin, Y., & Sercu, L. (2025). ChatGPT interventions in higher education: A systematic review of experimental studies. *Journal of Computer Assisted Learning, 41*(4), e70072. https://doi.org/10.1111/jcal.70072

© 2027 Conscientia Beam. All Rights Reserved.

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education, 22*(1), 7. https://doi.org/10.1186/s41239-025-00503-7

Lee, S. S., & Moore, R. L. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning, 28*(3), 82-106.

Li, T., Ji, Y., & Zhan, Z. (2024). Expert or machine? Comparing the effect of pairing student teacher with in-service teacher and ChatGPT on their critical thinking, learning performance, and cognitive load in an integrated-STEM course. *Asia Pacific Journal of Education, 44*(1), 45-60. https://doi.org/10.1080/02188791.2024.2305163

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assessment & Evaluation in Higher Education, 49*(5), 616-633. https://doi.org/10.1080/02602938.2024.2301722

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching, 51*(1), 36-76. https://doi.org/10.1017/S0261444817000350

McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2025). Generative AI chatbots in higher education: A review of an emerging research area. *Higher Education, 89*(6), 1533-1549. https://doi.org/10.1007/s10734-024-01288-w

Meyer, J. G., Urbanowicz, R. J., Martin, P. C. N., O'Connor, K., Li, R., Peng, P.-C., . . . Moore, J. H. (2023). ChatGPT and large language models in academia: Opportunities and challenges. *BioData Mining, 16*(1), 20. https://doi.org/10.1186/s13040-023-00339-9

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research.* Paris, France: UNESCO.

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

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education, 39*(4), 235-251. https://doi.org/10.1080/21532974.2023.2247480

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers, 12*(8), 153. https://doi.org/10.3390/computers12080153

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences, 14*(6), 636. https://doi.org/10.3390/educsci14060636

Rose, H., Sahan, K., Wei, M., Aizawa, I., Zhou, S., & Shepard, C. (2026). A systematic review of English medium instruction in higher education: An update of Macaro et al.(2018). *System, 136*, 103892. https://doi.org/10.1016/j.system.2025.103892

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Tang, X., Rousse-Malpat, A., & Duarte, J. (2026). Implementing translanguaging strategies in the English writing classroom in higher education: A systematic review. *AILA Review, 39*(1), 45-79. https://doi.org/10.1075/aila.23018.tan

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

UNESCO. (2023). *Global education monitoring report 2023: Technology in education—A tool on whose terms?* Paris, France: UNESCO.

Wang, J., Zaid, Y. H., Ibrahim, N. A., & Haladin, N. B. (2024). Teachers' and students' beliefs about translanguaging in higher education: A systematic review. *Journal of Nusantara Studies, 9*(2), 402-419. https://doi.org/10.24200/jonus.vol9iss2pp402-419

Wen, W., & Wen, H. (2024). Bridging generative AI technology and teacher education: Understanding preservice teachers' processes of unit design with ChatGPT. *Contemporary Issues in Technology and Teacher Education, 24*(4), 582-611.

Views and opinions expressed in this article are the views and opinions of the author(s), International Journal of Education and Practice shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.

© 2027 Conscientia Beam. All Rights Reserved.