frontiers in Psychology

TYPE Original Research
PUBLISHED 17 November 2026
DOI 10.3389/fpsyg.2026.1869502

Check for updates

OPEN ACCESS

EDITED BY
Daniel H. Robinson,
The University of Texas at Arlington College of Education, United States

REVIEWED BY
Jinming Du,
University of Otago, New Zealand
Muhamad Riyan Maulana,
Yogyakarta State University, Indonesia

*CORRESPONDENCE
Bimei Wang
✉ 303509725@qq.com
Yuyin Jiang
✉ jyuyin@163.com

RECEIVED 30 April 2026
REVISED 24 August 2026
ACCEPTED 29 October 2026
PUBLISHED 17 November 2026

CITATION
Wang B, Zhang Q, Zheng W, Jiang Y, Gao J and Li M (2026) Comparing human-AI collaboration and teacher-student collaboration in developing pre-service teachers' instructional design competence: an experimental study.
Front. Psychol. 17:1869502.
doi: 10.3389/fpsyg.2026.1869502

COPYRIGHT
© 2026 Wang, Zhang, Zheng, Jiang, Gao and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comparing human-AI collaboration and teacher-student collaboration in developing pre-service teachers' instructional design competence: an experimental study

Bimei Wang*, Qi Zhang, Wenfeng Zheng, Yuyin Jiang*, Jie Gao and Mingjun Li

Shaanxi Xueqian Normal University, Xi'an, China

Introduction: The application of generative artificial intelligence in learning has become increasingly prominent, and how it can complement and augment teachers' roles has become a key topic of discussion.
Methods: This study adopted an experimental research paradigm to examine the effects of human-AI collaboration and teacher-student collaboration on the improvement of pre-service teachers' instructional design ability from the perspectives of information feedback and theory of mind.
Results: The results showed that both human-AI collaboration and teacher-student collaboration could enhance pre-service teachers' instructional design ability. However, at different time points, pre-service teachers' instructional design ability after eight weeks of training was significantly higher than that before training and after four weeks of training. Under different learning strategies, no statistically significant difference was detected between teacher-student collaboration and human-AI collaboration in the overall instructional design competence score, although the two strategies showed distinct advantages at the dimensional level and in learning experience. However, the subject-specific analysis revealed a significant interaction between learning strategy and subject discipline: the difference in the effects of the two learning strategies was more pronounced in the Chinese language subject than in the mathematics subject. Moreover, both the flow experience measurement and self-reports indicated significant differences in the learning process between teacher-student collaboration and human-AI collaboration.
Discussion: Based on these findings, it is recommended to establish an interactive dual-feedback loop of information, enhance feedback interactivity and emotional experience, and focus on subject-specific knowledge logic for guidance and feedback.

KEYWORDS
generative artificial intelligence, human-AI collaboration, instructional design competence, learning strategies, teacher-student collaboration

1 Introduction

In recent years, the rapid rise of artificial intelligence has driven innovation and development in education and teaching, transforming educational scenarios, learning pathways, and teacher-student relationships (Farrokhnia et al., 2023). As representative applications such as ChatGPT and Sora have emerged, education has "entered the digital era of 'new-quality talent cultivation'" (Zhu et al., 2024), with generative AI driving changes in learning modes among university students and facilitating the transformation of educational philosophies in higher education.

Alongside the development of generative AI, Intelligent Tutoring Systems (ITS) have been continuously upgraded and widely adopted by learners, including platforms such as Wenxin Yiyan, Khmingo, and Learn About. These large model-based educational applications not only assist students in completing, revising, and improving written texts but also support problem-solving in practical applications. Existing research has explored the potential of generative AI in empowering teacher professional development from theoretical models and practical pathways (Huang et al., 2025) and enhancing student learning through the provision of precise digital educational resources (Ibrahim et al., 2023; Dalgiç et al., 2024) and research on empowering teachers' instructional design (Ahn et al., 2024); Hashem et al. (2024) found that ChatGPT-generated lesson plans align with the 5E model, incorporating elements such as students sharing prior knowledge, engaging in group collaboration, classroom debates, and self-assessment activities and the reconstruction of smart learning environments (Wu et al., 2025). While generative AI has demonstrated unique advantages in student learning and teaching practice, leading to educational model reforms and diverse human-AI collaborative learning modes (Gimpel et al., 2023; OpenAI, 2023), current research remains primarily focused on knowledge-level exploration. This study first examines whether generative AI can facilitate the improvement of pre-service teachers' professional skills during learning.

Pre-service teacher' instructional design ability is a key competency they must possess before entering the teaching profession. It is an important component of talent cultivation in higher normal institutions and a core element of teachers' professional development (Zaragoza et al., 2024; Großmann and Krüger, 2024). Existing research has explored strategies for enhancing pre-service teachers' teaching competence, including micro-video production to improve practical teaching and technology application capabilities (Li et al., 2020), peer and expert feedback in micro-teaching to promote reflective teaching (Liu et al., 2023), and intelligent technology support to enhance digital teaching competence (Gou et al., 2024). Contextual self-directed learning has been found to exert a progressive effect on the improvement of pre-service teachers' teaching skills (Li Z. et al., 2023). Analysis of existing enhancement strategies reveals that teacher guidance is a critical factor in the development of pre-service teachers' teaching skills. As the most fundamental teaching skill and a key training objective in pre-service education, this study second examines whether teacher-guided collaborative learning facilitates the improvement of pre-service teachers' instructional design competence.

A growing body of research has begun to explore AI's application in instructional design. Studies have shown that AI can generate high-quality lesson plans aligned with established pedagogical frameworks such as the 5E model (Hashem et al., 2024), and can assist pre-service teachers in producing lesson plans that receive higher quality ratings from external evaluators (Zaugg, 2024). AI provides structured frameworks, generates diverse teaching activities, and reduces the initial difficulty of instructional design for novice teachers (Wickard et al., 2025; Peikos and Stavrou, 2025). Research has also explored AI's potential as a collaborative partner, with chatbots providing real-time feedback to promote understanding of instructional design theories (Fidan and Gencel, 2022) and simulated student interactions helping to develop teaching skills (Lee and Yeo, 2022; Markel et al., 2023). However, these studies have predominantly focused on the quality of AI-generated instructional products rather than on the development of pre-service teachers' own instructional design competence through the process of human-AI collaboration.

Moreover, existing experimental studies have largely compared AI-supported learning against no-intervention or self-study conditions, leaving the question of how human-AI collaboration compares with the traditional gold standard of teacher-guided collaboration largely unexplored. This gap is particularly significant because teacher guidance has long been recognized as critical for pre-service teachers' professional skill development (Li Z. et al., 2023; Liu et al., 2023). The question of whether AI can replace teachers or complement their role in developing instructional design competence remains unanswered. As Choi et al. (2024) noted, while AI offers efficiency in generating initial solutions, it exhibits notable limitations in deep pedagogical reasoning and contextual adaptation. Lan and Chen (2024) further emphasized that teachers' agency remains crucial in the era of large language models, suggesting that AI should augment rather than replace human expertise.

The improvement of instructional design competence requires a collaborative learning environment. Generative AI can reshape smart learning environments and support efficient learning, but questions remain regarding its impact on learning experience and higher-order thinking. Teacher-guided collaborative learning also provides a high-quality learning environment, but whether this approach has unique advantages over AI-supported learning requires empirical investigation. From the perspective of information feedback theory, human-AI collaboration and teacher-student collaboration differ fundamentally in their feedback mechanisms: the former is dominated by external feedback, while the latter involves the synergy of external and internal feedback. Mind perception theory further suggests that learners' perceptions of the feedback source may lead to different learning behaviors and attitudes. These theoretical distinctions warrant systematic empirical examination.

Methodologically, previous studies have predominantly employed single-group designs, qualitative case studies, or comparisons against no-intervention conditions. These approaches lack the rigor needed to establish causal relationships between specific learning strategies and competence development. An

experimental approach with a controlled comparison between human-AI collaboration and teacher-student collaboration is therefore necessary.

Accordingly, this study compares the effectiveness of generative AI-enabled human-AI collaborative learning and teacher-guided collaborative learning in improving pre-service teachers' instructional design competence, and explores the underlying mechanisms of human-AI collaboration.

2 Literature review and theoretical framework

2.1 Research on generative artificial intelligence

As a key technology in the new round of scientific and technological revolution, generative AI's powerful conversational capabilities have become an important driver of new-quality productivity. The *Outline for Building a Leading Educational Power (2024–2035)* issued in 2025 proposes "promoting AI-enabled educational transformation" (Xinhua News Agency, 2025), further advancing the application of AI in the education sector. Recent research has adopted three distinct paradigms:

The first category is theoretical argumentation. This type of research has been continuously deepening since generative artificial intelligence entered the research field. There are two main themes in this category. (1) The first is to argue the significance, value, problems, and countermeasures of generative AI empowering education. For example, Kalota (2023) and Li S. et al. (2023) conducted an in-depth discussion on generative AI-related tools and technologies, some application scenarios, and related challenges and opportunities. Okonkwo and Ade-Ibijola (2021) systematically reviewed the diverse application scenarios of chatbots in higher education: teaching (66%), research and development (19%), assessment (6%), administration (5%), and consultation (4%). The ability of chatbots to understand and generate human-like responses (i.e., natural language processing technology) creates a more intuitive tool interaction method for students and educators (Maher et al., 2020; Rath et al., 2023), while also promoting the personalization of learning experiences based on students' individual needs (Younis et al., 2023). For example, Yan et al. argue that "generative AI enables educators to understand educational objects more efficiently and systematically" (Yan and Qin, 2023), while Wu et al. note that "generative AI can reconstruct learning spaces, redesign content, reshape capabilities, readjust processes, and redevelop assessment systems for learners" (Wu et al., 2023); (2) The second is to explore the risk governance and ethical norms of artificial intelligence. Research in this area focuses on issues such as how to prevent and govern risks. Del Valle García-Carreño's (2025) study, through a bibliometric analysis of literature on the Scopus platform, found that the high-speed adaptive capability of generative AI is reshaping the educational ecosystem. However, the application of this technology is also accompanied by ethical and equity concerns, including threats to academic integrity, data privacy risks, potential algorithmic bias, and challenges such as ensuring equitable access and teacher training. Therefore, only by properly addressing relevant risks, ethical challenges, and equity issues can its benefits be fully realized while safeguarding academic integrity and avoiding the exacerbation of educational inequality. Baidoo-Anu and Ansah's (2023) study focused on exploring the application of general artificial intelligence in educational practice, pointing out risks in the protection of privacy issues. Chocarro et al.'s (2023) study also indicated that artificial intelligence faces risks in terms of ethics, morality, and data privacy. And Yusuf et al. (2024) examines the core challenges posed by generative AI to academic integrity and its potential implications for higher education reform, analyzing these issues across three dimensions: technological advancements, legal frameworks, and practical applications. Zhu et al. (2024) propose governance pathways for ethical and moral risks, while Holmes and Tuomi (2022) emphasize the necessity of establishing broad community consensus to resolve such issues. These studies confirm that generative AI can support learning, but improper use may pose risks, providing theoretical support for this study's exploration of generative AI-enabled improvement of instructional design competence while emphasizing the need for risk prevention in research design.

The second paradigm is variable exploration, which examines the impact of generative AI on individuals as its application expands in education, business, and other industries, focusing on relationships between variables. For example, The impact of artificial intelligence on adolescent health (Awu, 2025); studies have investigated the impact of AI on student "technology anxiety" (Fang and Ma, 2025) and the impact on writing skills (Almassaad et al., 2024; Dergaa et al., 2023); the impact on learners' listening, speaking, reading, and writing skills (Monika, 2024) and the relationship between AI anxiety and AI adoption behavior (Chen and Chen, 2025). However, in extensive research on AI applications in education, the roles, perspectives, needs of teachers and educational administrators, as well as the impact of AI on their work, have been severely overlooked (Zawacki-Richter et al., 2019). Survey research has also analyzed the impact of generative AI on student learning and teaching practice. Budhathoki et al. (2024) based on the Technology Acceptance Model (TAM), explored students' or teachers' (Strzelecki et al., 2024) attitudes toward the acceptance and use of generative AI tools in education. Gabriel-Mugurel and Liliana-Luminita (2023) pointed out that parents hold positive but reserved attitudes toward AI tools, mainly due to uncertainty about the future development of this technology. Kelly et al. (2023) conducted a survey of 1,135 students at an Australian university and found that the majority of students had low levels of awareness, practical experience, and operational confidence in using these tools. A national survey of 8,563 students across seven provinces by Sun et al. found that "AI use for homework assistance is widespread but associated with problems of thinking inertia and social alienation" (Sun, 2025; Characteristics, Risks). Xiao et al. found that "teachers are willing to use generative AI but face limitations in content generation quality and deep integration into instruction" (Xiao et al., 2025). These studies provide a foundation for examining whether human-AI collaborative learning affects students' deep learning.

The third paradigm is experimental validation, which also explores variable relationships but adopts experimental research as the primary methodology. These studies treat generative AI as an experimental variable to examine its impact on enterprise employees, teachers, and students. Peikos and Stavrou (2025) explored the use of ChatGPT in assisting primary school lesson planning from the perspective of Pedagogical Content Knowledge (PCK). The results showed that incorporating PCK elements into prompts, adopting layered prompting strategies, and providing reference texts to ChatGPT may improve the quality of AI-generated lesson plans. Dong et al. conducted a study on human-AI collaborative writing with learning outcome as the dependent variable and learning style as the grouping variable (Dong et al., 2025). Zheng et al. examined variables including knowledge activation volume, contribution degree, learning performance, and critical thinking awareness, finding that generative AI improved all these variables (Zheng et al., 2024). Qian et al. (2025) systematically analyzed 39 experimental and quasi-experimental research studies on the impact of GenAI on students' learning outcomes. The study found that GenAI has a moderately strong positive effect on students' learning outcomes ($g = 0.623$). The effect of GenAI on students' learning outcomes diminished as sample size increased. Among different application types, the dialogue-interaction mode outperformed other modes—such as content generation, assessment feedback, and mixed modes—in terms of both application prevalence and effectiveness.

Experimental research provides a rigorous paradigm for verifying the impact of generative AI on learners, which this study adopts to analyze the effects of different learning modes on pre-service teachers' instructional design competence.

2.2 Research on pre-service teachers' instructional design competence

Instructional design competence is a critical component of teaching competence, and its improvement is essential for enhancing teachers' professional competence in the implementation of the "Strengthening Teacher Education Plan". As a core element of pre-service teachers' teaching skills, instructional design competence has long been a focus of subject teaching research, with continuous exploration of strategies for enhancing pre-service teachers' teaching competence. Existing research has primarily focused on various components of teaching competence, including digital teaching competence (Xue and Fang, 2024), reflective teaching competence (Liu et al., 2023), and practical teaching competence (Deng and Ren, 2023), with limited research specifically targeting instructional design competence. Furthermore, existing research is largely concentrated on theoretical discussion and case analysis, lacking rigorous experimental evidence to verify enhancement strategies. Accordingly, this study adopts an experimental research design focusing on pre-service teachers' instructional design competence to provide empirical evidence for the improvement of pre-service teachers' teaching skills.

2.3 Research on AI-assisted instructional design

Studies on generative AI in instructional design can be broadly categorized into two strands: one examining the impact of AI on the quality of instructional design products (i.e., lesson plan texts), and the other investigating the role of AI as a collaborative partner in developing pre-service teachers' instructional design competence.

At the product level, a consistent body of evidence supports the effectiveness of AI in improving lesson plan quality. Peikos and Stavrou (2025) explored the use of ChatGPT in supporting primary school lesson planning from a Pedagogical Content Knowledge (PCK) perspective, finding that incorporating PCK elements into prompts and providing reference texts improved the quality of AI-generated lesson plans. Hashem et al. (2024) found that ChatGPT-generated lesson plans aligned with the 5E model, incorporating elements such as group collaboration, classroom debates, and self-assessment activities. Kim and Ahn (2023) demonstrated that ChatGPT-generated elementary English lesson plans exhibited high effectiveness, particularly in the areas of instructional objectives and assessment strategies. Zaugg (2024) employed a quasi-experimental design and found that pre-service teachers who used an AI assistant to create lesson plans received significantly higher quality ratings from external evaluators compared to the control group, while also completing the task in substantially less time. Wickard et al. (2025) noted that AI provides novice teachers with accessible teaching examples and diverse pedagogical approaches, effectively reducing the planning pressure associated with complex instructional design tasks. Hu et al. (2024) investigated the use of GPT-4 for generating and evaluating lesson plans, revealing the substantial potential of large language models in supporting instructional design. Collectively, these studies demonstrate that AI offers clear advantages in providing structured frameworks, generating diverse teaching activities, ensuring comprehensive content coverage, and improving design efficiency.

However, these studies share a common limitation: they focus primarily on the quality of AI-generated lesson plans, rather than on how pre-service teachers develop their own instructional design competence through collaboration with AI. The emphasis has been on what AI can produce, not on what learners can learn from the process of interacting with AI. This distinction is critical. As Parsons and Curry (2024) cautioned in their investigation of whether ChatGPT could pass graduate-level instructional design assignments, if AI can independently produce high-quality instructional designs, the role of pre-service teachers in the instructional design training process must be re-examined. Xu et al. further noted that instructional design ultimately serves real, context-specific classroom situations, including students' prior knowledge, learning motivation, classroom culture, and school resources. AI-generated "generic" lesson plans often lack the capacity to perceive and adapt to these specific contexts, yet the internalization of competence precisely requires learners to engage in deep thinking and self-construction through interaction with feedback.

At the process level, a growing number of studies have begun to examine the mechanisms of human-AI collaboration and its impact

on competence development. Fidan and Gencel (2022) found that chatbot-based real-time feedback promoted pre-service teachers' understanding of instructional design theories, suggesting that AI feedback can serve as an effective tool for facilitating theoretical learning. Lee and Yeo (2022) developed an AI-based chatbot that act as a virtual student, training pre-service teachers' questioning skills and facilitating the transfer of theoretical knowledge to practical competence, revealing the potential of AI in simulated teaching contexts. Markel et al. (2023) developed GPTeach, which simulates student interactions with pre-service teachers in real time, alleviating novice teachers' instructional pressure and enriching their practical experience. Choi et al. (2024) conducted a SWOT analysis of generative AI for instructional design, noting that while AI offers efficiency in generating initial solutions, it exhibits notable limitations in deep pedagogical reasoning and contextual adaptation. Lan and Chen (2024) proposed a conceptual framework for designing "pedagogical AI agents" from both teacher and student perspectives, emphasizing that teachers' agency remains crucial in the era of large language models and that AI should serve as a tool to augment rather than replace human expertise. While these studies have revealed the potential of human-AI collaboration in teacher education from various angles, most remain exploratory in nature or focus on theoretical framework building, lacking rigorous experimental evidence for systematic comparisons between human-AI collaboration and traditional teacher guidance.

In summary, three key research gaps emerge from the existing literature. First, prior studies have primarily focused on the product quality of AI-generated lesson plans rather than on the competence development of pre-service teachers through the collaborative process. Second, among the limited studies that have examined competence development, most have employed single-group designs or qualitative case studies, lacking rigorous experimental comparisons. Third, and most critically, no prior study has directly compared human-AI collaboration with teacher-student collaboration—the traditional gold standard in teacher education—within a controlled experimental framework. This gap is particularly significant because teacher guidance has long been recognized as a critical factor in the development of pre-service teachers' professional skills (Li Z. et al., 2023; Liu et al., 2023), yet whether AI can replace or supplement this role remains empirically unexamined. To address these gaps, the present study adopts an experimental research paradigm with a crossover design to systematically compare the effects of human-AI collaboration and teacher-student collaboration on pre-service teachers' instructional design competence, while simultaneously examining both learning outcomes and learning experiences.

2.4 Theoretical framework

Information feedback is a core regulator of deep learning and a critical element of skill acquisition. Through external and internal feedback systems, information feedback acts on learners' internal monitoring and external adjustment through characteristics such as timeliness, individualization, guidance, and interactivity, collectively promoting skill improvement. Internal feedback occurs during learners' self-regulation, representing "new knowledge generated when comparing current knowledge and capabilities with reference information" (Nicol and Selvaretnam, 2021) and a learning perception that forms the basis of meta-cognition. External feedback provides "external validation" of behavioral outcomes, relying on external standards and providing explicit guidance. By providing clear information standards, external feedback helps learners establish internal "references," stimulate meta-cognition, and trigger internal feedback (Lam, 2016), with meta-cognition activation as its core mechanism. Meta-cognition serves as the "engine" of deep learning, enabling learners to continuously adjust cognitive activities during information feedback, thereby achieving autonomous construction, deep understanding, and transfer application of learning. Thus, deep learning depends on effective information feedback.

Information feedback can reinforce learners' cognition and improve learning outcomes. Both teacher feedback and generative AI feedback affect student learning performance. Generative AI feedback can provide personalized feedback based on individual information (Cadario et al., 2021), thereby enhancing task performance (Parent-Rocheleau and Parker, 2022). For example, Sharma et al. found that "AI-generated personalized revision suggestions significantly improved manuscript quality, helping students achieve higher grades" (Sharma et al., 2023). However, passive acceptance of AI feedback may hinder learning (Castelo and Ward, 2021). To address this, the theory of mind perception proposes that individuals' perceived agency and experience influence their attitudes and behaviors (Gray et al., 2007). Learners' learning initiative depends on high perception of feedback: perceived agency promotes conscious thinking (Gray and Wegner, 2012), while perceived experience facilitates active information processing. Thus, individuals' perception of feedback is a core mechanism affecting behavior and a key determinant of feedback effectiveness.

Information feedback theory provides the foundational framework for understanding how different learning strategies affect the learning process. In the human-AI collaboration mode, learners primarily obtain feedback through dialogue with artificial intelligence—they pose questions, and the AI directly provides revision suggestions. The information flow is dominated by external feedback: learners receive solutions from the AI, but the depth of their internal processing depends on the quality of their questioning and their willingness to reflect. In the teacher-student collaboration mode, teachers employ guided questioning and follow-up inquiries to prompt learners to examine and reflect on their existing instructional designs. In this process, external feedback (the teacher's guidance) and internal feedback (the learner's comparison, judgment, and self-correction based on their own knowledge in relation to the teacher's feedback) occur simultaneously. That is, in the teacher-student collaboration mode, teachers do not directly provide revision suggestions; instead, they guide learners to identify problems in their original designs and to further discover approaches for improvement. This difference in feedback type is associated with learners' learning experiences: in the former mode, learners tend to receive information passively, whereas in the latter mode, deep thinking

may activate metacognitive monitoring. The differing levels of perceived agency and experience are associated with differences in learning attitudes and behaviors.

Mind perception theory further explains why different feedback sources lead to behavioral differences among learners. The theory posits that individuals' perceptions of the agency and experience of a feedback source influence their attitudes and behaviors. In teacher-student collaboration, learners perceive the teacher as possessing higher agency and experience, and thus tend to process information more actively and adjust their behavior more proactively. In human-AI collaboration, by contrast, learners' perception of the AI may lean toward low agency, resulting in more passive information processing. This theory provides a psychological mechanism for understanding the significant differences in learning experiences under different learning strategies.

The selection of information feedback theory and mind perception theory as the core theoretical framework of this study is well justified by their direct relevance to the research variables. Information feedback theory directly explains how feedback influences learners' metacognitive monitoring mechanisms through internal and external feedback systems—a process particularly pertinent to instructional design competence, which requires iterative revision and refinement. Mind perception theory, in turn, explains why learners' perceptions of AI feedback vs. teacher feedback give rise to different learning behaviors and attitudes. Compared with existing AI learning frameworks, information feedback theory better explains how AI participates in the learning process, while mind perception theory explains the behavioral differences resulting from different learning strategies. The integration of the two theories provides a complete account of the relationship between the learning process and learning behavior. In contrast to alternative frameworks such as social presence theory (which focuses on the perception of social interaction) or cognitive load theory (which emphasizes information processing capacity), these two theories are more directly centered on the core variables of this study: feedback type and learning strategy comparison.

Building on the above theoretical integration, this study proposes the following conceptual pathway (see Figure 1): External feedback (AI or teacher) → Activation of metacognitive monitoring → Triggering of internal feedback → Facilitation of deep engagement (flow experience) → Improvement of instructional design competence. This pathway serves as a guiding framework for the experimental design and data interpretation.

2.5 Theoretical contributions and research positioning

The theoretical contributions of this study are threefold.

First, this study integrates information feedback theory and mind perception theory in a comparative investigation of human-AI collaboration and teacher-student collaboration. Existing research on AI in education has largely drawn on the Technology Acceptance Model (TAM), social presence theory, or cognitive load theory. These frameworks focus primarily on explaining users' acceptance of technology or their information processing capacity, but they seldom address the core question of how differences in feedback type affect the depth of cognitive processing during learning. Information feedback theory offers a process-level explanation for this question, while mind perception theory provides a behavioral-level explanation. The integration of the two forms a complete theoretical chain from "learning process" to "learning behavior."

Second, this study proposes a testable conceptual pathway model. The theoretical pathway of "external feedback → metacognitive monitoring → internal feedback → deep engagement → competence improvement" provides an operationalizable framework for future research to examine each mediating mechanism. Although this study did not directly measure metacognitive monitoring, the flow experience scale and self-report measures provide indirect evidence for this pathway. Future research can build on this foundation by conducting direct mediation analyses.

Third, this study provides empirical support for the theoretical framework. Through an experimental design, this study systematically compared the effects of two different feedback sources (AI vs. teacher) on learning outcomes (instructional design competence) and learning experiences (flow experience and self-reports), providing empirical evidence for the application of information feedback theory and mind perception theory in educational contexts. The findings also reveal the explanatory power and boundary conditions of existing theoretical frameworks when confronted with emerging human-AI interaction scenarios.

3 Research design

This study examines the effects of different learning strategies (human-AI collaborative learning and teacher-student collaborative learning) on pre-service teachers' instructional design competence. A 2 (learning strategy: human-AI collaboration vs. teacher-student collaboration) × 3 (time: pre-test/mid-test/post-test) mixed design was adopted. To scientifically verify competence growth under different learning modes, the study adopts multi-modal evidence for evidence-based research, including expert evaluation of instructional design texts and pre-post measurement of instructional design competence.

3.1 Research questions

1. Can the human-AI collaborative learning strategy based on generative AI effectively improve pre-service teachers' instructional design competence?
2. Can the teacher-guided collaborative learning strategy effectively improve pre-service teachers' instructional design competence?
3. What differences exist in the learning experiences (including flow states and subjective learning perceptions) of teacher education students under different learning strategies?

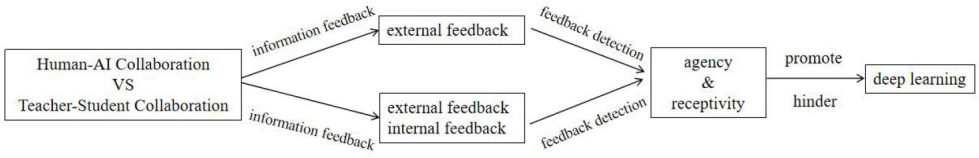

FIGURE 1
Theoretical model of the study.

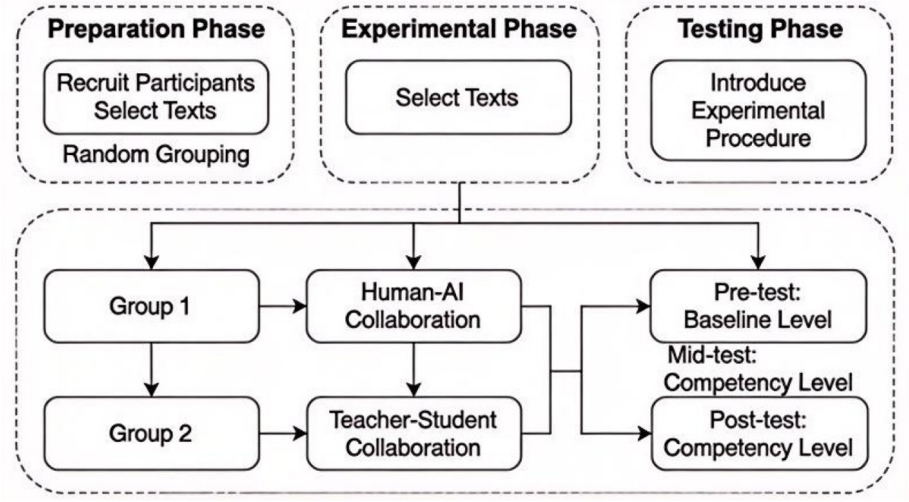

FIGURE 2
Experimental flowchart.

3.2 Participants

Forty-six pre-service teachers (10 males, 36 females) aged 21–22 in their third year of primary education major at a normal university in China were recruited for the study. All participants had completed courses in education and educational psychology, possessing a basic theoretical foundation in instructional design. Recruitment ensured that all participants had an interest in teaching and motivation to improve their instructional design skills.

Participants were randomly assigned to two experimental groups, with 23 participants in each group. Pre-test results confirmed no significant difference in initial instructional design competence between the two groups ($t = 0.79$, $p > 0.05$), indicating homogeneous groups. Although the two groups demonstrated homogeneity at pre-test, this study did not separately measure potential confounding variables such as participants' prior familiarity with AI, digital literacy, or intrinsic motivation. To mitigate the influence of these factors, the following measures were adopted: (1) random assignment ensured that these variables were balanced between the two groups; (2) pre-test results confirmed the homogeneity of the two groups in instructional design competence; (3) the crossover design allowed each participant to experience both learning strategies, thereby controlling for individual differences.

However, 10 participants were excluded from the final analysis for the following reasons: six participants failed to complete the full 8-week training program due to scheduling conflicts; four participants did not attend the post-test assessment and were therefore excluded from the analysis. The final analytical sample thus consisted of 36 participants (18 per group), with 8 males and 28 females. No significant difference was found between the excluded and retained participants in pre-test instructional design competence scores, suggesting that the exclusion did not introduce systematic bias.

A priori power analysis was conducted using G*Power 3.1 for the repeated measures ANOVA design (within-between interaction). With an assumed medium effect size ($f = 0.25$), $\alpha = 0.05$, power $(1\text{-}\beta) = 0.80$, two groups, and three measurement time points with a correlation of $r = 0.50$ among repeated measures, the analysis indicated a minimum required sample size of 28 participants. The actual sample size of 36 participants (18 per group) exceeded this minimum requirement, suggesting adequate statistical power for detecting the main effects and interaction effects in the primary analyses.

This study was approved by the Ethics Review Committee of Shaanxi Xueqian Normal University. All participants provided written informed consent. Participants' privacy and data security were strictly protected throughout the research process, and all data were anonymized.

3.3 Research instruments

Instructional design competence evaluation instrument. The instrument was developed specifically for the present study by integrating the Lesson Plan Evaluation Index by Mo and He (2007) and the Components of Pre-service Teachers' Instructional Design Skills by Jackson and Marsh (1996), comprising three first-level indicators (teaching objective design, teaching content organization, teaching process design) and 20 evaluation points, using a 5-point rating scale. Although the instrument has not been formally validated in previous studies, its dimensional structure is grounded in well-established theoretical frameworks in the field of instructional design. The content validity was established through multiple rounds of discussion among five frontline experts, who confirmed that the 20 evaluation points adequately covered the key components of instructional design competence for primary education.

To further examine the construct validity of the instrument, we examined the inter-dimensional correlations using data from all three measurement time points. The three dimensions showed low to moderate positive correlations with each other: M1 and M2 ($r = 0.373$, $p = 0.001$), M1 and M3 ($r = 0.537$, $p < 0.001$), and M2 and M3 ($r = 0.293$, $p = 0.013$). These correlations provide preliminary evidence consistent with the notion that the three dimensions represent related but distinct facets of instructional design competence. Furthermore, the instrument demonstrated adequate sensitivity to detect significant improvements in instructional design competence over the eight-week training period. The total score increased significantly from pre-test (M = 3.11, SD = 0.29) to post-test (M = 4.43, SD = 0.12), with a large effect size ($\eta p^2 = 0.932$, $p < 0.001$), indicating that the instrument was capable of capturing meaningful changes in competence over time. It should be noted that formal construct validity testing through exploratory factor analysis (EFA) or confirmatory factor analysis (CFA) was not possible due to the limited sample size ($N = 36$). Future research with larger samples is needed to further validate the dimensional structure of the instrument. Formal construct validation through EFA or CFA remains necessary in future research with larger samples to confirm the dimensional structure of the instrument.

To ensure the objectivity and fairness of the evaluation, the following measures were implemented: (1)Blind review design: Prior to submission for evaluation, all instructional design texts were uniformly anonymized by a research assistant, with participants' names, group assignments, and measurement time points removed and replaced with random identification numbers; (2)Independent scoring: The five experts completed their scoring independently, without communicating their ratings with one another; (3)Inter-rater reliability: The intraclass correlation coefficient (ICC) was used to assess the consistency of the five experts' ratings. The results showed ICC values of 0.87 (pre-test), 0.91 (mid-test), and 0.89 (post-test) across the three time points, all exceeding the threshold of 0.80, indicating good inter-rater consistency; (4) Discrepancy resolution: When the range of the five experts' scores for a given instructional design exceeded 2 points, the five experts engaged in collective discussion and re-scored the text, with the consensus score after discussion serving as the final score. In this study, texts requiring re-scoring accounted for 4.2% of the total.

The learning experience scale adopted the Flow Status Scale (FSS) developed by Jackson et al. (Xia et al., 2025), measuring 9 dimensions with 36 items using a 5-point rating scale. The scale demonstrated good reliability with Cronbach's $\alpha = 0.709$, suitable for measuring pre-service teachers' learning experiences under different learning strategies.

To further understand how different learning modes affect learning experiences, the study conducted in-depth interviews with typical cases, focusing on participants' self-reported learning experiences and self-evaluation of the two learning strategies.

3.4 Experimental materials

Instructional design content was selected from textbooks compiled by the Ministry of Education, with 20 texts (10 Chinese, 10 mathematics) sampled across grades, text types, and knowledge modules. All sampled texts were evaluated by 20 front line experts as representative of main text types for the two subjects in primary education.

For generative AI selection, instructional design revision suggestions generated by Wenxin Yiyan, Deepseek, and ChatGPT were compared using the lesson plan *"Rain in Autumn"* as an example. Ten Chinese teaching researchers rated the three sets of revision suggestions, and Deepseek was ultimately selected as the collaborative learning tool.

The AI tool used in this study was DeepSeek-V2, accessed via its web interface (chat.deepseek.com) during the period from September to November 2025. All participants received a standardized 30-min training session prior to the experiment, which included: (1) an overview of the AI tool's capabilities and limitations; (2) academic integrity in AI use; (3) how to input instructional design drafts into the AI; and (4) how to formulate instructions for the AI to assist in revising their designs. Since the guidance process involves iterative dialogue between the learner and the AI, participants were not restricted to specific prompts or provided with a fixed template.

Although participants were not restricted to a fixed prompt template, the following are representative examples of the types of prompts they used during their interactions with the AI:

(1) "I am designing a lesson plan for [subject] at [grade level]. Here is my current draft. Please provide suggestions for optimizing the instructional design."
(2) "Please review the teaching objectives in my lesson plan and analyze whether they align with the current curriculum standards. Please help me revise them."
(3) "Please review the teaching content design based on the teaching objectives and suggest revisions that better align with the objectives."
(4) "Based on the teaching objectives and content, please provide suggestions for revising the teaching process design."

TABLE 1 Correspondence between experimental elements and theoretical constructs.

Theoretical construct	Operational definition	Experimental correspondence
External feedback (Information feedback theory)	Revision suggestions and guidance directly provided by AI or the teacher	Human-AI collaboration group: revision suggestions generated through AI dialogue; Teacher-student collaboration group: teachers' guiding feedback
Internal feedback (information feedback theory)	Learners' self-correction generated through comparison between external feedback and their own knowledge	Teacher-student collaboration group: students' self-reflection triggered by the teacher's guided questioning
Metacognitive monitoring (information feedback theory)	Learners' awareness and regulation of their own learning process	Indirect measurement: dimensions such as "clear goals" and "unambiguous feedback" in the flow experience scale; "deep thinking" indicator in self-reports
Agency and experience (Mind perception theory)	Learners' perception of the feedback source	Indirectly reflected through between-strategy differences in learning experience (flow experience, self-reports)
Learning outcome	Instructional design competence	Expert-rated instructional design texts (three dimensions: teaching objective design, teaching content organization, teaching process design)
Learning experience	Flow experience and subjective learning perceptions	Flow state scale (FSS) + self-report coding analysis

These examples illustrate the general nature of the prompts used, while actual prompts varied according to individual participants' needs and the specific instructional design task.

3.5 Theoretical basis of the experimental design

The research design of this study is closely aligned with the core theoretical framework. Based on information feedback theory, the essential difference in feedback mechanisms between the two learning strategies is defined as follows: the human-AI collaboration strategy is dominated by external feedback—learners obtain solutions by posing questions to the AI, with the information flow following a "learner → AI → learner" direction, where feedback content is generated by the AI, and the depth of learners' information processing depends on the quality of their questioning and their willingness to reflect. In contrast, the teacher-student collaboration strategy involves the synergy of external and internal feedback—teachers employ guided questioning and follow-up inquiries to prompt learners to examine and reflect on their existing instructional designs, with the information flow following a bidirectional "teacher → learner ↔ teacher" pattern. As learners receive external feedback from the teacher, they simultaneously trigger internal feedback through comparison, judgment, and self-correction, thereby potentially activating metacognitive monitoring. To facilitate readers' understanding of the correspondence between the research design and the theoretical framework, Table 1 presents the mapping of experimental elements to theoretical constructs.

To account for the repeated-measures structure of the crossover design, linear mixed-effects models were used to analyze the flow experience data, with learning strategy (human-AI collaboration vs. teacher-student collaboration) as the fixed effect and participant ID as a random effect. To control for the increased risk of Type I error due to multiple comparisons across the nine flow experience dimensions and the 12 self-report indicators, the Holm correction was applied to all p values.

3.6 Experimental procedure

This study adopted a crossover experimental design to eliminate differences between participants. It should be noted that each participant only completed instructional design tasks for a single subject, not both subjects (see Figure 2). Specifically, the 36 participants were divided into a Chinese language group ($n = 18$) and a mathematics group ($n = 18$) based on their specialized direction within the primary education program. Each group only worked on the 10 texts corresponding to their assigned subject (1 pre-test, 8 weekly training texts, and 1 post-test). The crossover of learning strategies occurred within each subject group: in each subject, participants experienced both human-AI collaboration and teacher-student collaboration across the two rounds (Weeks 1–4 and Weeks 5–8). Therefore, subject discipline (Chinese vs. mathematics) served as a between-subjects variable, while learning strategy (human-AI collaboration vs. teacher-student collaboration) served as a within-subjects variable, making them separable in the experimental design. Before the experiment, all participants were required to complete an instructional design for their assigned subject (Chinese 1 for the Chinese group, Mathematics 1 for the mathematics group). Experts were invited to review these designs to ensure the baseline level of all participants, and a questionnaire survey was conducted to assess their instructional design ability.

To reduce the confounding effect of subject discipline on the learning strategy, this study employed a subject-mixing strategy. Within each experimental phase, Chinese and Mathematics subjects were arranged alternately to ensure that the two groups were

TABLE 2 Descriptive statistics of pre-service teacher' instructional design ability.

Time point	Experimental group ($n = 18$)		Control group ($n = 18$)		Total ($N = 36$)	
	M	SD	M	SD	M	SD
Pre-test	3.15	0.25	3.06	0.33	3.11	0.29
Fourth week (mid-test)	3.94	0.19	4.14	0.19	4.04	0.21
Eighth week (post-test)	4.34	0.80	4.52	0.65	4.43	0.12

TABLE 3 Repeated measures ANOVA results of pre-service teacher' instructional design ability.

Source of variation	SS	*df*	MS	F	*p*	ηp^2
Time	33.55	1.67	20.06	463.96***	<0.001	0.932
Time × group	0.44	1.67	0.27	6.12**	0.006	0.153
Group	0.25	1	0.25	4.50*	0.041	0.117
Error (time)	2.46	56.86	0.04			
Error (group)	1.90	34	0.06			

*p < 0.05.
**p < 0.01.
***p < 0.001.

exposed to completely equivalent subject content within the same phase.

The experiment lasted 8 weeks. During Weeks 1–4, the first experimental group engaged in human-AI collaborative learning, while the second experimental group engaged in teacher-student collaborative learning. During Weeks 5–8, the two groups exchanged learning modes. Weekly learning activities involved pre-service teachers drafting an instructional design, followed by revision guidance from either generative AI or front line teaching experts, with one instructional design completed per week over the 8-week training period.

Instructional design competence was evaluated at three time points: before the experiment, after Week 4, and after Week 8. Assessments involved on-site text sampling with a 2-h time limit for completing the instructional design, conducted under offline conditions without internet access to ensure authenticity. Assessment results were evaluated by front line experts.

A crossover design was adopted for two primary reasons. First, by allowing each participant to serve as their own control, the crossover design controls for individual differences between participants (e.g., prior AI familiarity, digital literacy, intrinsic motivation), thereby increasing statistical power with a relatively small sample. Second, it ensures that all participants experience both learning strategies, enabling direct within-subject comparisons of gain scores. A parallel randomized controlled design would require a larger sample to achieve equivalent statistical power and would be more susceptible to confounding by individual differences.

To mitigate potential carryover effects, the following measures were implemented. First, although each participant completed instructional design tasks within a single subject, the 10 texts within that subject covered different grades, text types, and knowledge modules. The specific texts used in the first phase (Weeks 1–4) differed from those used in the second phase (Weeks 5–8), reducing the likelihood of direct content transfer between phases. Second, a formal test of order effects was conducted to examine whether the sequence of learning strategies influenced the outcomes. Nevertheless, given the cumulative nature of instructional design competence development, the possibility of carryover effects cannot be completely excluded.

4 Results analysis

4.1 Overall improvement of pre-service teachers' instructional design competence

To ensure that the two groups of pre-service teacher were at a comparable level, an independent samples *t*-test was conducted on the pre-test data. The results showed that there was no significant difference between the two groups at the pre-test stage (experimental group: M = 3.15; control group: M = 3.06; $t = 0.862$; $p = 0.394$), indicating that the two groups were equivalent at baseline.

A 2 × 3 repeated measures ANOVA was conducted with time (pre-test, fourth week, eighth week) as the within-subjects factor and group (experimental group, control group) as the between-subjects factor. The assumption of sphericity was violated (Mauchly's W = 0.804, $p = 0.027$), so the Greenhouse-Geisser correction ($\varepsilon = 0.84$) was applied. Therefore, the corrected results are reported. Descriptive statistics and ANOVA results are presented in Tables 2, 3.

The repeated measures ANOVA results showed significant differences in pre-service teacher' instructional design ability across different time points. The main effect of time was significant, $F_{(1.67,56.86)} = 463.96$, $p < 0.001$, $\eta_p^2 = 0.932$, indicating that the instructional design ability of both groups significantly improved over the training period. Post-hoc comparisons (LSD) showed significant differences between all three time points ($p < 0.001$),

with scores increasing progressively: eighth week post-test (M = 4.43, SD = 0.12) > fourth week mid-test (M = 4.04, SD = 0.21) > pre-test (M = 3.11, SD = 0.29).

The interaction effect of Time × Group was significant, $F_{(1.67,56.86)} = 6.12$, $p = 0.006$,$\eta_p^2 = 0.153$, indicating that the two groups showed significantly different trends over time. Descriptive statistics revealed that in the first round (pre-test to fourth week), the control group showed a greater improvement ($M_{\text{difference}}$ = 1.08, SD = 0.38) than the experimental group ($M_{\text{difference}}$ = 0.79, SD = 0.28); in the second round (fourth week to eighth week), the two groups showed similar improvement (experimental group: M $_{\text{difference}}$ = 0.40, SD = 0.18; control group: $M_{\text{difference}}$ = 0.38, SD = 0.19).

The main effect of group was significant, $F_{(1,34)} = 4.50$, p = 0.041, $\eta_p^2 = 0.117$, with the control group's overall mean score (M = 3.91, SD = 0.63) being slightly higher than that of the experimental group (M = 3.81, SD = 0.54).

Content analysis of pre-service teacher' instructional designs also confirmed the quantitative results. Before training, pre-service teacher showed deficiencies in writing and implementing teaching objectives, and had difficulty designing activities aligned with teaching goals. However, after eight weeks of training, pre-service teacher' teaching objectives met the requirements of core competencies, and their activity designs were able to revolve around the teaching objectives.

Since this study adopted a crossover design in which the two groups exchanged learning strategies at Week 4, it was necessary to test for potential order effects. A mixed-design ANOVA was conducted with sequence (Group 1: human-AI collaboration followed by teacher-student collaboration vs. Group 2: teacher-student collaboration followed by human-AI collaboration) as the between-subjects factor and time (pre-test, mid-test, post-test) as the within-subjects factor. The results showed that the main effect of sequence was not significant, $F_{(1,34)} = 0.52$, p = 0.476, $\eta p^2 = 0.015$, and the sequence × time interaction was also not significant, $F_{(1.67,56.86)} = 0.38$, $p = 0.543$, η_p^2 = 0.011. These results indicate that the order in which participants experienced the two learning strategies did not significantly affect their instructional design ability improvement, supporting the validity of the crossover design."

To further examine whether the learning period influenced the improvement of instructional design competence, a paired-sample t-test was conducted with period (Round 1: Weeks 1–4 vs. Round 2: Weeks 5–8) as the independent variable and gain scores as the dependent variable. The results showed that the gain score in Round 1 (M = 0.93, SD = 0.51) was significantly higher than that in Round 2 (M = 0.39, SD = 0.19), $t_{(35)} = 6.14$, $p <$ 0.001, *Cohen's d* = 1.02. This significant period effect indicates that participants showed greater improvement in the first half of the experiment than in the second half, which is consistent with the typical learning curve pattern where initial training yields larger gains. Importantly, since both groups experienced both learning strategies across the two periods due to the crossover design, this period effect does not confound the comparison between the two strategies. The combination of non-significant order effects and significant period effects supports the validity of the crossover design while acknowledging that the timing of learning (early vs. late) influences the magnitude of improvement. However, these results do not entirely eliminate the possibility of carryover effects, as the cumulative nature of instructional design competence development may have influenced the results.

4.2 Comparison of improvement effects under different learning strategies

Since this study adopted a crossover experimental design, each participant experienced both human-AI collaboration and teacher-student collaboration learning strategies. To directly compare the improvement effects of the two strategies on pre-service teacher' instructional design ability, paired-sample t-tests were conducted using gain scores of the total instructional design ability score and its three sub-dimensions (teaching objective design, teaching content organization, teaching process design) as dependent variables. The results are presented in Table 4.

Paired-sample t-test results showed that there was no significant difference between the teacher-student collaboration strategy (M = 0.74, SD = 0.45) and the human-AI collaboration strategy (M = 0.59, SD = 0.29) in the total score of instructional design competence, $t_{(35)} = -1.257$, $p = 0.217$, Cohen's d = 0.21. Among the three sub-dimensions, in the teaching objective design dimension, the gain of the teacher-student collaboration strategy (M = 0.95, SD = 0.55) was significantly higher than that of the human-AI collaboration strategy (M = 0.52, SD = 0.48), $t_{(35)} = -2.818$, $p = 0.008$, Cohen's $d = 0.47$. In the teaching process design dimension, the gain of the human-AI collaboration strategy (M = 0.82, SD = 0.29) was significantly higher than that of the teacher-student collaboration strategy (M = 0.56, SD = 0.48), $t_{(35)} = 2.167$, $p = 0.037$, Cohen's $d = 0.36$. In the teaching content organization dimension, no significant difference was found between the two strategies ($p = 0.157$). *Post-hoc* interviews suggested that the large language models referenced by generative AI are mostly derived from instructional designs by frontline teachers, which may share commonalities with designs guided by frontline teachers, thus failing to produce substantial differences. This may partially explain the non-significant difference observed between the two strategies. It should be noted, however, that a non-significant result does not demonstrate equivalence between two interventions. The most appropriate conclusion is that no statistically significant difference was detected between the two strategies in the overall instructional design competence score.

This finding is consistent with the conclusions of Xia Liangliang et al., who found that "the impact of generative AI on students' learning agency is highest at the behavioral level but lowest at the competency and psychological trait levels."

From the perspective of information feedback theory, no statistically significant difference was detected between the two strategies in terms of overall instructional design competence gain scores, suggesting that both human-AI collaboration and teacher-student collaboration can effectively activate the feedback pathway leading to competence improvement. This may be attributable to the fact that the large language models underlying generative AI are predominantly trained on instructional designs

TABLE 4 Analysis of gain scores in pre-service teacher' instructional design ability under different learning strategies.

Dependent variable	Teacher-student collaboration gain ($n = 36$)		Human-AI collaboration gain ($n = 36$)		Paired difference		t	Cohen's d
	M	SD	M	SD	M差值	SD差值		
Teaching objective design	0.96	0.55	0.53	0.48	−0.43	0.91	−2.818**	0.47
Teaching content organization	0.66	0.59	0.43	0.42	−0.23	0.94	−1.445	0.24
Teaching process design	0.56	0.48	0.82	0.29	0.26	0.72	2.167*	0.36
Instructional design ability	0.74	0.45	0.59	0.29	−0.146	0.69	−1.257	0.21

*$p < 0.05$.
**$p < 0.01$.

developed by frontline teachers, which may partially explain the non-significant difference observed between the two strategies. However, at the dimensional level, the two strategies exhibited distinct advantages: teacher-student collaboration significantly outperformed human-AI collaboration in the teaching objective design dimension, whereas human-AI collaboration significantly outperformed teacher-student collaboration in the teaching process design dimension. These divergent patterns are consistent with the different mechanisms through which the two feedback approaches facilitate the development of specific instructional skills. Information feedback theory posits that the effectiveness of feedback depends not only on the content of external feedback but also on the extent to which it triggers internal feedback and metacognitive monitoring. The non-significant difference in the overall score does not necessarily imply equivalence in the underlying learning processes—a distinction that the flow experience and self-report findings in Section 4.4 further illuminate.

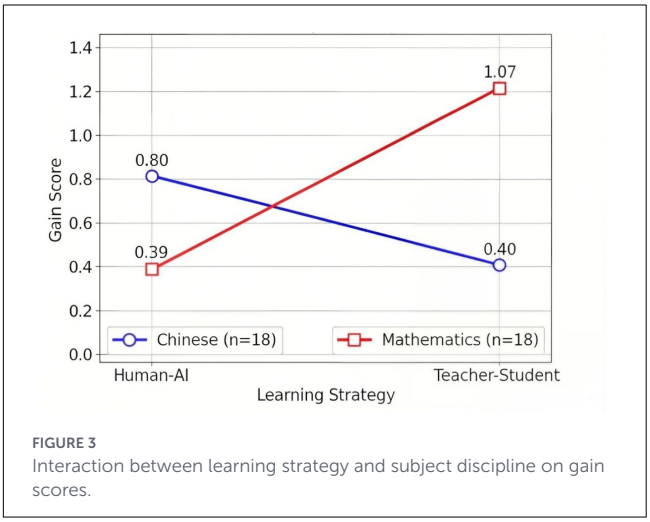

FIGURE 3
Interaction between learning strategy and subject discipline on gain scores.

4.3 Improvement effects across disciplinary backgrounds

To formally test whether the improvement effects of the two learning strategies differed across subject disciplines, a mixed-design ANOVA was conducted with learning strategy (human-AI collaboration vs. teacher-student collaboration) as the within-subjects factor and subject discipline (Chinese vs. mathematics) as the between-subjects factor, using gain scores as the dependent variable. The results revealed a significant interaction between learning strategy and subject discipline, $F_{(1,34)} = 55.946$, $p < 0.001$, $\eta p^2 = 0.622$, indicating that the difference between the two learning strategies varied significantly between Chinese and mathematics subjects. This significant interaction provides statistical justification for the subject-specific comparisons reported below.

As illustrated in Figure 3, the two learning strategies displayed a cross-over interaction pattern across the two subject disciplines. In Chinese language, the human-AI collaboration strategy (M = 0.80) yielded higher gain scores than the teacher-student collaboration strategy (M = 0.40). In contrast, in mathematics, the teacher-student collaboration strategy (M = 1.07) substantially outperformed the human-AI collaboration strategy (M = 0.39). This cross-over pattern indicates that the relative effectiveness of the two learning strategies is contingent upon subject discipline.

To further examine whether subject discipline influences the improvement effects of the two learning strategies, paired-sample t-tests were conducted on gain scores across dimensions in mathematics and Chinese language subjects. The results are presented in Table 5.

In mathematics, the teacher-student collaboration strategy significantly outperformed the human-AI collaboration strategy in the total score of instructional design competence (M = 1.07 vs. 0.39, $t_{(17)} = -5.519$, $p < 0.001$, $d = 1.30$, 95% CI [0.63, 1.98]), as well as in the teaching objective design dimension (M = 1.29 vs. 0.33, $t_{(17)} = -4.800$, $p < 0.001$, $d = 1.13$, 95% CI [0.49, 1.77]) and the teaching content organization dimension (M = 1.10 vs. 0.11, $t_{(17)} = -7.172$, $p < 0.001$, $d = 1.69$, 95% CI [0.92, 2.46]). In the teaching process design dimension, no significant difference was found between the two strategies ($p = 0.860$, $d = 0.04$, 95% CI [−0.46, 0.54]). It is noteworthy that in the mathematics group, the human-AI collaboration strategy showed very low gains in teaching objective design (M = 0.33) and teaching content organization (M = 0.11), while the teacher-student collaboration strategy maintained substantially higher gains in these dimensions.

In Chinese language, the human-AI collaboration strategy significantly outperformed the teacher-student collaboration

10.3389/fpsyg.2026.1869502

TABLE 5 Paired-sample *t*-test results of gain scores under two learning strategies across subject disciplines.

Subject	因变量	Teacher-Student collaboration gain ($n = 36$)		Human-AI collaboration Gain ($n = 36$)		Paired difference		t	*Cohen's d* [95% CI]
		M	SD	M	SD	M_{diff}	SD_{diff}		
Mathematics	Teaching objective design	1.29	0.52	0.33	0.44	−0.96	0.85	−4.800***	1.13 [0.49, 1.77]
	Teaching content organization	1.10	0.45	0.11	0.31	−0.98	0.58	−7.172***	1.69 [0.92, 2.46]
	Teaching process design	0.75	0.58	0.72	0.31	−0.04	0.84	−0.180	0.04 [−0.46, 0.54]
	Instructional design ability	1.07	0.38	0.39	0.22	−0.69	0.53	−5.519***	1.30 [0.63, 1.98]
Chinese	Teaching objective design	0.62	0.33	0.72	0.45	0.10	0.64	0.660	0.16 [−0.34, 0.66]
	Teaching Content Organization	0.22	0.32	0.75	0.25	0.53	0.52	4.300***	1.01 [0.40, 1.62]
	Teaching process design	0.37	0.25	0.92	0.22	0.55	0.42	5.665***	1.34 [0.65, 2.03]
	Instructional design ability	0.40	0.19	0.80	0.20	0.39	0.31	5.359***	1.26 [0.59, 1.93]

***$p < 0.001$

strategy in the total score of instructional design competence (M = 0.80 vs. 0.40, $t_{(17)} = 5.359$, $p < 0.001$, $d = 1.26$, 95% CI [0.59, 1.93]), as well as in the teaching content organization dimension (M = 0.75 vs. 0.22, $t_{(17)} = 4.300$, $p < 0.001$, $d = 1.01$, 95% CI [0.40, 1.62]) and the teaching process design dimension (M = 0.92 vs. 0.37, $t_{(17)} = 5.665$, $p < 0.001$, $d = 1.34$, 95% CI [0.65, 2.03]). In the teaching objective design dimension, no significant difference was found between the two strategies ($p = 0.518$, $d = 0.16$, 95% CI [−0.34, 0.66]).

The significant interaction between learning strategy and subject discipline ($p < 0.001$, $\eta p^2 = 0.622$) confirmed that the difference between the two learning strategies varied significantly across subject disciplines. Specifically, a cross-over pattern was observed: in mathematics, teacher-student collaboration significantly outperformed human-AI collaboration ($d = 1.30$, 95% CI [0.63, 1.98]), whereas in Chinese language, human-AI collaboration significantly outperformed teacher-student collaboration ($d = 1.26$, 95% CI [0.59, 1.93]). However, the wide confidence intervals for the effect sizes indicate considerable uncertainty in the estimates due to the small subsample size (n = 18 per subject). Therefore, the precise magnitude of these disciplinary differences requires further validation with larger samples.

4.4 Learning experiences under different learning strategies

Learning experience refers to an individual's perception and experience during the learning process, encompassing nine dimensions: challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, concentration on the task at hand, sense of control, loss of self-consciousness, transformation of time, and autotelic experience (enjoyable experience). To account for the repeated-measures structure of the crossover design, a linear mixed model was used to analyze the flow experience data, with learning strategy (human-AI collaboration vs. teacher-student collaboration) as the fixed effect and participant ID as a random effect. The results are presented in Table 6.

The linear mixed model results revealed significant differences between the two learning strategies across all nine dimensions and the overall flow experience score (all $p < 0.01$), with large to extremely large effect sizes (Cohen's $d = 1.04$–2.82). Specifically, the teacher-student collaboration strategy significantly outperformed the human-AI collaboration strategy in all dimensions. The largest effect sizes were observed in transformation of time ($d = 2.82$), autotelic experience ($d = 2.55$), and unambiguous feedback ($d = 2.36$), followed by challenge-skill balance ($d = 2.21$), sense of control ($d = 1.63$), loss of self-consciousness ($d = 1.51$), action-awareness merging ($d = 1.31$), concentration on task ($d = 1.20$), and clear goals ($d = 1.04$). These results indicate that face-to-face interaction with expert teachers provides pre-service teachers with clearer feedback, stronger immersion, a greater sense of control, and better alignment between task demands and personal skills. To address the issue of multiple comparisons, the p values were adjusted using the Holm correction. All nine dimensions remained

TABLE 6 Linear mixed model results of flow experience dimensions under different learning strategies.

Flow experience dimension	Human-AI collaboration		Teacher-student collaboration		$F_{(1,35)}$	P	Cohen's d
	M	SD	M	SD			
Challenge-skill balance	2.91	0.61	4.03	0.81	42.723	<0.001	2.21
Action-awareness merging	3.29	0.80	4.02	0.63	14.932	<0.001	1.31
Clear goals	3.43	0.92	4.07	0.53	9.504	0.004	1.04
Unambiguous feedback	3.11	0.66	4.20	0.53	48.773	<0.001	2.36
Concentration on task	3.23	0.83	4.08	0.53	12.706	0.001	1.20
Sense of control	3.49	0.99	4.31	0.47	23.208	<0.001	1.63
Loss of self-consciousness	3.19	0.75	4.24	0.63	19.945	<0.001	1.51
Transformation of time	3.11	0.90	4.40	0.52	69.638	<0.001	2.82
Autotelic experience	3.29	0.80	4.31	0.49	56.926	<0.001	2.55
Overall flow experience	3.23	0.74	4.12	0.45	68.791	<0.001	2.80

All p values are adjusted using the Holm correction for multiple comparisons across the nine flow experience dimensions.

statistically significant after correction, indicating that the observed differences are robust.

This finding presents an interesting contrast with the results of instructional design ability: in terms of improving instructional design competence, the two strategies showed no significant difference in the overall score; however, in terms of learning experience, teacher-student collaboration demonstrated a clear and substantial advantage across all dimensions of flow experience. This aligns with the perspective of Luckin et al. (2016), who argued that AI's optimal role in education is not to replace teachers, but rather to serve as a powerful tool for enhancing human intelligence.

To explore how different learning strategies affect learners' learning experiences, this study employed a self-report method to track typical cases, with reports focusing on concentration, intrinsic interest, learning willingness, and sense of achievement. The collected data were coded using NVIVO 12.0 and analyzed using the Mann-Whitney U non-parametric test. The results are presented in Table 7.

As shown in the table above, there were significant differences in students' learning experiences under different learning strategies. Based on the chi-square values, the greatest differences in learning experience between the two strategies were in "intrinsic interest" and "learning willingness." Compared to teacher-student collaboration, human-AI collaboration was less effective in stimulating students' intrinsic interest and willingness to learn. Students reported in their self-reports: "During expert guidance, I would think step by step about how to revise and improve based on the expert's guidance, but when using AI, I generally did not engage in deep thinking and simply replaced content according to the AI's revision suggestions." However, in terms of goal attainment within the sense of achievement dimension, the two learning strategies did not show significant differences in learning experience, indicating that both strategies achieved the goal of revising instructional designs. The results of the non-parametric test were highly consistent with the linear mixed model results of flow experience, with teacher-student collaboration significantly outperforming human-AI collaboration in the vast majority of self-report indicators.

After applying the Holm correction for multiple comparisons, 11 out of 12 indicators remained statistically significant, with only goal achievement showing no significant difference.

It is noteworthy that in the goal attainment indicator, the mean ranks of the two groups were identical (36.50), $p = 1.000$. This finding echoes the non-significant difference between the two groups in total flow experience scores ($p = 0.587$), indicating that the two learning strategies were comparable in terms of learning outcome achievement but showed significant differences in the quality of the learning process experience.

These differences are consistent with the predictions of mind perception theory. Learners perceived the teacher as possessing higher agency and experience, which led to more active information processing, deeper cognitive engagement, and elevated levels of concentration, control, and intrinsic interest. In contrast, although the AI provided functionally equivalent feedback, learners perceived it as a lower-agency source, resulting in a more passive learning stance. This pattern is consistent with the theoretical proposition that perceived agency and experience may mediate the relationship between feedback source and learning behavior. Notably, the comparable "goal attainment" ratings between the two conditions further corroborate this interpretation: both strategies achieved the instrumental goal of revising instructional designs (comparable outcomes), but they differed substantially in how learners engaged with the feedback (process divergence). This outcome-process dissociation may illustrate the complementary roles of AI and teachers in instructional design training.

5 Conclusions and recommendations

Through experimental research, this study compared the differences in pre-service teacher' teaching skill improvement under different learning strategies and at different time points. The study found that the main effect of time was extremely significant ($\eta_p^2 = 0.932$), indicating that after eight weeks of training, pre-service teacher' instructional design ability had

TABLE 7 Mann-Whitney *U* test results of learner self-reports under different learning strategies.

Dimension	Indicator	Human-AI collaboration		Teacher-student collaboration		Mann-Whitney *U*	*Z*	*P*
		M	SD	M	SD			
Concentration	Instructional design	11	28.50	28	44.50	360.00	−3.77	<0.001
	Problem improvement	16	31.00	27	42.00	450.00	−2.63	0.009
	Expert thinking	6	24.60	30	47.08	231.00	−5.30	<0.001
Intrinsic Interest	Inquiry motivation	3	27.50	21	45.50	324.00	−4.47	<0.001
	Flow experience	2	24.50	26	48.50	216.00	−5.76	<0.001
	Sense of control	4	26.00	25	47.00	270.00	−5.01	<0.001
Learning willingness	Deep learning	6	25.50	28	47.50	252.00	−5.16	<0.001
	Continuous learning	4	25.50	26	47.50	252.00	−5.22	<0.001
	Active learning	3	25.50	25	47.50	252.00	−5.28	<0.001
Sense of achievement	Goal achievement	15	36.50	19	36.50	648.00	0.00	1.000
	Ability improvement	6	26.50	26	46.50	288.00	−4.71	<0.001
	Overcoming difficulties	6	26.50	27	46.50	288.00	−4.69	<0.001

p values are adjusted using the Holm correction for multiple comparisons across the 12 self-report indicators.

substantially improved. The interaction effect of time and group was significant ($\eta_p^2 = 0.153$), indicating differences in the change trends between the two groups: in the first round, the control group (teacher-student collaboration) showed greater improvement than the experimental group (human-AI collaboration); in the second round, the improvement rates of the two groups converged. Paired-sample *t*-tests further showed that in the total score of instructional design ability and its three sub-dimensions, the mean gains of the teacher-student collaboration strategy were higher than those of the human-AI collaboration strategy, but none of the differences reached statistical significance ($p > 0.05$), with effect sizes ranging from small to medium ($d = 0.24$–0.26).The results showed that no statistically significant difference was detected between the two strategies in improving normal students' instructional design ability, with small to medium effect sizes ($d = 0.24$–0.26). The most appropriate conclusion is that no statistically significant difference was found between the two strategies in the overall score. Subject-specific analysis revealed a significant interaction between learning strategy and subject discipline ($p < 0.001$, $\eta p^2 = 0.622$), confirming that the effects of the two learning strategies differed across subject disciplines. A cross-over pattern was observed: in mathematics, teacher-student collaboration significantly outperformed human-AI collaboration ($d = 1.30$, 95% CI [0.63, 1.98]), whereas in Chinese language, human-AI collaboration significantly outperformed teacher-student collaboration ($d = 1.26$, 95% CI [0.59, 1.93]). This pattern suggests that the relative effectiveness of the two approaches depends on subject-specific characteristics.

This difference may be related to subject characteristics. Chinese language instructional design places greater emphasis on text interpretation, emotional experience, and language expression, which are highly open and subjective. Generative AI (such as Deepseek) has natural advantages in language generation, text analysis, and multi-perspective interpretation, enabling it to provide pre-service teacher with rich teaching ideas and diverse modes of expression. In contrast, mathematics instructional design focuses more on logical reasoning, formula derivation, and problem-solving, which are highly structured, leaving relatively limited room for AI assistance. Significant differences were found in students' learning experiences under different learning strategies, with the main differences lying in "intrinsic interest" and "learning willingness."

These findings are consistent with international research. Zawacki-Richter et al. (2019), in their systematic review of AI in higher education, identified four key application areas and emphasized that the transformation of the teacher's role is a core challenge of AI integration in education. Holmes and Tuomi (2022) argued that AI should serve as a "cognitive partner" for teachers rather than a replacement. The present study provides empirical support for these perspectives: while generative AI can serve as an effective instructional assistant, it remains less effective than teacher guidance in promoting deep learning experiences and sustained engagement.

These findings also have implications for teacher professional identity. Rather than viewing AI as a substitute for teachers, the findings suggest that the most productive educational model involves complementary human-AI collaboration. In this model, AI serves as an efficient feedback provider and instructional design assistant, while teacher-student collaboration demonstrated particular advantages in facilitating metacognitive activation, promoting deep engagement, and providing emotional support. This complementary partnership aligns with the contemporary perspective that AI should augment rather than replace human expertise (Holmes and Tuomi, 2022; Lan and Chen, 2024). Teacher-student collaboration demonstrated unique advantages that are consistent with the facilitation of learners' metacognitive monitoring and internal feedback processes, roles that AI has yet to fully replicate.

It should be noted that the discussion of "deep learning" and "metacognitive activation" in the following sections represents theoretical interpretations derived from information feedback theory, rather than directly measured outcomes. While the flow experience scale and self-report measures provide indirect evidence of engagement, future research should employ dedicated metacognition scales (e.g., the Metacognitive Awareness Inventory) and learning transfer tests to directly verify these underlying mechanisms.

The quantitative findings of this study can be interpreted through the conceptual pathway proposed in the theoretical framework. The significant main effect of time ($\eta_p^2 = 0.932$) is consistent with the interpretation that both learning strategies may have activated the feedback → metacognition → engagement → competence improvement pathway. The non-significant difference between the two strategies in overall instructional design ability gains suggests that both human-AI collaboration and teacher-student collaboration can effectively trigger this pathway at the outcome level. However, the significant differences in flow experience and self-report measures indicate that the two strategies differ substantially in the quality of the engagement phase of this pathway. Teacher-student collaboration produced higher levels of unambiguous feedback, concentration, and sense of control, suggesting that teacher feedback may be more conducive to metacognitive monitoring and internal feedback processes. In contrast, human-AI collaboration, while achieving comparable learning outcomes, was less effective in promoting deep engagement during the learning process. This pattern is consistent with the mind perception theory: learners perceive higher agency and experience in teacher feedback, leading to more active cognitive engagement, whereas AI feedback, despite its utility, may be processed more passively.

Information feedback theory suggests that the effectiveness of information feedback depends on the mutual reinforcement of internal and external feedback. Individuals' perceived agency and receptivity to information feedback are core factors influencing behavioral differences. Under different learning strategies, learners' demonstrated learning interest and willingness are influenced by their perceived feedback, which in turn affects their mastery and improvement of skills. Therefore, in enhancing pre-service teacher' instructional design ability, it is necessary to construct a dual feedback mechanism, activate students' metacognitive monitoring mechanisms, and employ diverse strategies to promote pre-service teacher' agency and receptivity to information feedback, thereby facilitating their deep learning.

The findings of this study also provide empirical support for a division of labor between AI and teachers in instructional design training. Specifically, AI demonstrated particular strengths in supporting teaching process design, where it significantly outperformed teacher guidance (see Section 4.2). In contrast, teacher guidance was more effective in fostering teaching objective design and, most notably, in cultivating a richer learning experience across all dimensions of flow experience (see Section 4.4). This suggests that an optimal instructional model may involve integrating both approaches — leveraging AI's efficiency in generating structured suggestions and simulating diverse teaching scenarios, while relying on teachers' expertise in guiding deep reflection, activating metacognitive processes, and providing emotional and motivational support. Future research could explore how to design hybrid instructional models that strategically combine human-AI collaboration and teacher-student interaction to maximize both learning outcomes and learning experiences.In summary, AI can effectively support pre-service teachers in developing certain aspects of instructional design competence, particularly in generating and organizing teaching process components. However, in terms of deep learning engagement, metacognitive reflection, and positive learning experience, teacher guidance demonstrated particular advantages.Specific recommendations are as follows:

5.1 Establish an interactive dual information feedback loop to activate the core layer of deep learning

The results confirmed that both learning strategies can improve pre-service teachers' instructional design competence, but with different areas of strength. Teacher-student collaboration significantly outperformed human-AI collaboration in the teaching objective design dimension ($p = 0.008$), while human-AI collaboration significantly outperformed teacher-student collaboration in the teaching process design dimension ($p = 0.037$). The teacher-student collaboration strategy provides pre-service teachers with a holistic learning pathway, in which teachers employ guided questioning, follow-up inquiries, and counter-questions to prompt learners to reflect on their own instructional designs, thereby activating metacognition and facilitating knowledge internalization. The human-AI collaboration strategy, in contrast, provides pre-service teachers with rich instructional design examples and basic design guidance, laying a foundation for creative instructional design. The two learning strategies influence learning through different information feedback channels. Therefore, it is necessary to establish a dual feedback loop of internal and external feedback, using external feedback to stimulate internal feedback and using internal feedback to reduce dependence on external feedback, thereby achieving self-monitoring of knowledge learning.

An important question emerging from these findings is why the substantial differences in learning experience did not translate into corresponding differences in learning outcomes. One possible explanation is that the flow experience differences primarily reflect the process quality of learning, whereas the improvement of instructional design competence may require a longer intervention period for process-level advantages to manifest at the outcome level. The eight-week intervention may not have been sufficient for process differences to fully translate into outcome differences. Another possibility is that the expert rating instrument, which focuses on the final quality of instructional design texts, may not fully capture changes in learners' thinking processes. Future research could incorporate process-oriented assessment indicators, such as the number of revision iterations, the depth of modifications,

and the quality of reflective journals, to provide a more comprehensive evaluation of the differences between the two strategies.

It is important to note that the absence of a statistically significant difference between the two strategies does not necessarily imply equivalence. While the effect sizes were small to medium (d = 0.24–0.26), suggesting that the magnitude of any true difference is likely limited, a non-significant result may also reflect insufficient statistical power to detect a small effect given the relatively modest sample size (N = 36). Future research with larger samples could provide more precise estimates of the difference between the two strategies. Additionally, the comparable outcomes at the product level should be interpreted alongside the significant differences in learning experience at the process level, which suggest that the two strategies differ substantially in how learners engage with feedback, even if the final instructional design products appear similar in quality.

5.2 Enhance feedback interactivity and emotional experience to activate the external layer of deep learning

The results confirmed that teacher-student collaboration significantly outperformed human-AI collaboration across all nine dimensions of flow experience (all $p < 0.01$), indicating a substantial gap between AI feedback and teacher feedback in terms of interactivity and emotional experience. Perceived agency and experience influence individuals' learning behaviors and attitudes, thereby affecting deep learning. Previous research has noted that "when faced with feedback on the same behavior from humans and AI, individuals may respond differently due to different perceptions of agency and experience" (Guan et al., 2025). Compared with generative AI feedback, teacher feedback provides learners with higher levels of perceived agency and experience through guided thinking and emotional engagement, which further explains the significant advantage of teacher-student collaboration in learning experience. Therefore, when using generative AI to support learning, feedback interactivity and anthropomorphism can be enhanced through task design and role positioning. Existing research indicates that "when AI feedback includes more interactivity and sociality, acceptance increases, and AI anthropomorphism helps enhance individuals' perception of AI experience" (Nussberger et al., 2022). Enhancing interactivity, sociality, and anthropomorphism can improve learners' perceived agency and experience of feedback, thereby promoting purposeful thinking and enhancing individual initiative.

5.3 Focus on disciplinary knowledge logic to activate the intermediate layer of deep learning

The results revealed a significant interaction between learning strategy and subject discipline ($p < 0.001$, $\eta p^2 = 0.622$), indicating that the effectiveness of learning strategies depends on subject-specific knowledge logic. Different disciplines have unique knowledge formation logics, and feedback should be tailored to disciplinary characteristics and knowledge formation logic to provide targeted guidance for deep learning. Therefore, the appropriate guidance approach should be selected based on subject characteristics — in Chinese language, which emphasizes text interpretation and emotional experience, AI's strengths in language generation and multi-perspective analysis can be fully utilized; in mathematics, which requires logical reasoning and structured problem-solving, teacher-guided feedback is more effective. Effective information feedback should focus on knowledge logic, starting from knowledge formation mechanisms to help learners establish underlying cognitive structures.

In summary, the three strategies discussed above — establishing a dual feedback loop, enhancing AI feedback interactivity, and tailoring guidance to disciplinary knowledge logic — collectively point toward a hybrid instructional model. In this model, AI is responsible for providing efficient external feedback and rich instructional examples, while teachers focus on stimulating internal feedback and cultivating deep learning experiences. The relative emphasis on each approach can be adjusted flexibly according to subject-specific characteristics, the targeted dimension of instructional design competence, and the desired learning experience. This hybrid model leverages the complementary strengths of both approaches rather than treating them as competing alternatives.

5.4 Limitations

Several limitations of this study should be acknowledged. First, the sample size was relatively small (N = 36), particularly for the subject-specific analyses (n = 18 per subject), which produced wide confidence intervals for the effect size estimates. Second, all participants were recruited from a single institution in China, limiting the generalizability of the findings. Third, the eight-week intervention period, while sufficient to detect improvements, does not speak to the long-term retention of the observed effects. Fourth, the study employed only one AI platform (DeepSeek-V2), and the novelty effect of AI interaction may have influenced participants' engagement. Fifth, the proposed mechanisms of metacognitive monitoring and feedback processing were not directly measured, and the conceptual pathway was only indirectly supported. Furthermore, the instructional design competence measurement instrument has not been subjected to formal construct validity testing through EFA or CFA, and this remains a limitation of the present study. Sixth, although the crossover design was adopted to control for individual differences and the order effect test was non-significant, the cumulative nature of instructional design competence development means that carryover effects between the two learning phases cannot be completely ruled out. The significant period effect, in which participants showed greater improvement in the first half of the experiment than in the second half, further underscores this concern. Future research should address these limitations through larger and more diverse samples, longitudinal designs, comparisons across AI platforms, and direct measurement of the proposed mechanisms.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving humans were approved by the Shaanxi Preschool Normal University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

BW: Writing – review & editing, Validation, Investigation, Supervision, Data curation, Methodology, Writing – original draft, Project administration. QZ: Writing – review & editing, Software. WZ: Writing – review & editing, Data curation, Project administration. YJ: Writing – original draft, Funding acquisition, Methodology. JG: Writing – review & editing, Investigation. ML: Project administration, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was supported by The Shaanxi Provincial Social Science Fund project "Research on the Training Model for Outstanding Science Teachers in Shaanxi Province under the Context of Chinese-Style Modernization" (Project No.: 2024P024); "The Youth Innovation Team of Shaanxi Universities" (Shaanxi Education Letter [2022] No. 943); Key Research Project of the 2025 Shaanxi Provincial Higher Education Teaching Reform Research Program: "Research and Practice on the Adaptation Mechanism between Discipline and Major Configuration Adjustment in Local Applied Universities and Regional Development" (Project No. 25BG086); Project of the 2026 Shaanxi Universities Outstanding Young Talent Support Program: "A Cross-Country Comparison on the Establishment of High-Quality Science and Technology Education Systems in Primary and Secondary Schools" (Project No. 2026QJ-02(N)).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahn, J., Lee, J., and Son, M. (2024). ChatGPT in ELT: disruptor? Or well-trained teaching assistant? *MELT J.* 78, 345–355. doi: 10.1093/elt/ccae017

Almassaad, A., Alajlan, H., and Alebaikan, R. (2024). Student perceptions of generative artificial intelligence: investigating utilization, benefits, and challenges in higher education. *Systems* 12:385. doi: 10.3390/systems12100385

Awu, I. O. (2025). Can generative artificial intelligence improve adolescent health? *J. Adolesc. Health* 77:790. doi: 10.1016/j.jadohealth.2025.05.020

Baidoo-Anu, D., and Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. *J. AI* 7, 52–62. doi: 10.61969/jai.1337500

Budhathoki, T., Zirar, A., Njoya, E. T., and Timsina, A. (2024). ChatGPT adoption and anxiety: a cross-country analysis utilising the unified theory of acceptance and use of technology (UTAUT). *Stud. Higher Educ.* 49, 831–846. doi: 10.1080/03075079.2024.2333937

Cadario, R., Longoni, C., and Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nat. Hum. Behav.* 5, 1636–1642. doi: 10.1038/s41562-021-01146-0

Castelo, N., and Ward, A. F. (2021). Conservatism predicts aversion to consequential artificial intelligence. *PLoS One* 16:261467. doi: 10.1371/journal.pone.0261467

Chen, P., and Chen, Y. (2025). The mechanism of artificial intelligence anxiety on generative AI adoption behavior: An empirical study based on the technology acceptance model. *New Media Soc.* 101-118+480-481.

Chocarro, R., Cortiñas, M., and Marcos-Matás, G. (2023). Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educ. Stud.* 49, 295–313. doi: 10.1080/03055698.2020.1850426

Choi, G. W., Kim, S. H., Lee, D., and Moon, J. (2024). Utilizing generative AI for instructional design: exploring strengths, weaknesses, opportunities, and threats. *TechTrends*, 68, 832–844. doi: 10.1007/s11528-024-00967-w

Dalgiç, A., Yaşar, E., Demir, M. (2024). ChatGPT and learning outcomes in tourism education: the role of digital literacy and individualized learning. *J. Hospitality Leisure Sport Tourism Educ.* 34:100481. doi: 10.1016/j.jhlste.2024.100481

Del Valle García-Carreño, I. (2025). Exploring risks and benefits in generative artificial intelligence through systematic review and bibliometric analysis. *Eur. Educ. Res.* 8, 57–94. doi: 10.31757/euer.834

Deng, X., and Ren, Z. (2023). Research on the development of teaching practice competencies of English teacher education students based on "Internet + full-process mentorship system + UGSIO." China Electro-Educ 5, 121–128. doi: 10.3969/j.issn.1006-9860.2023.05.016

Dergaa, I., Chamari, K., Zmijewski, P., and Saad, H. B. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biol. Sport* 40, 615–622. doi: 10.5114/biolsport.2023.125623

Dong, W., Zhao, F., Yu, N., et al. (2025). Research on cognitive and behavioral patterns of human-machine collaborative programming learning: an analysis based on learning style differences. *J. Distance Educ.* 43, 86–96. doi: 10.15881/j.cnki.cn33-1304/g4.2025.04.010

Fang, X., and Ma, H. (2025). Research on the impact of artificial intelligence learning on "technical anxiety" among college students. Res. High. Educ. Eng. 77–83.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., and Wals, A. (2023). A SWOT analysis of ChatGPT: implications for educational practice and research. *Innovations Educ. Teach. Int.* 61, 460–474. doi: 10.1080/14703297.2023.2195846

Fidan, M., and Gencel, N. (2022). Supporting the instructional videos with chatbot and peer feedback mechanisms in online learning: the effects on learning performance and intrinsic motivation. *J. Educ. Comput. Res.* 60, 1716–1741. doi: 10.1177/07356331221077901

Gabriel-Mugurel, D., and Liliana-Luminita, T. (2023). Students' perceptions of the impact of generative artificial intelligence. *J. Univ. Teach. Learn. Practice* 20, 451–471. doi: 10.18662/rrem/17.3/1030

Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., et al. (2023). *Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education [EB/OL]*. Stuttgart, Germany: University of Hohenheim.

Gou, J., Yang, J., Li, R., et al. (2024). Enhancement of digital teaching competencies among teacher trainees supported by intelligent technology: the role of expectation value beliefs. *China Electric Educ.* 8, 56–61+86. doi: 10.3969/j.issn.1006-9860.2024.08.008

Gray, H. M., Gray, K., and Wegner, D. M. (2007). Dimensions of mind perception. *Science* 315, 619–619. doi: 10.1126/science.1134475

Gray, K., and Wegner, D. M. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* 125, 125–130. doi: 10.1016/j.cognition.2012.06.007

Großmann, L., and Krüger, D. (2024). Assessing the quality of science teachers' lesson plans: evaluation and application of a novel instrument. *Sci. Educ.* 108, 153–189. doi: 10.1002/sce.21832

Guan, J., Li, W., He, G., and Zhang, X. (2025). Artificial intelligence feedback: a literature review and research prospects. *Foreign Econ. Manag.* 47, 83–100. doi: 10.16538/j.cnki.fem.20240622.301

Hashem, R., Ali, N., El Zein, F., Fidalgo, P., and Abu Khurma, O. (2024). AI to the rescue: exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Res. Practice Technol. Enhanced Learn.* 19:23. doi: 10.58459/rptel.2024.19023

Holmes, W., and Tuomi, I. (2022). State of the art and practice in AI in education. *Eur. J. Educ.* 57, 542–570. doi: 10.1111/ejed.12533

Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y., Gu, X., et al. (2024). Teaching plan generation and evaluation with GPT-4: unleashing the potential of LLM in instructional design. *IEEE Trans. Learn. Technol.* 17, 1471–1485. doi: 10.1109/TLT.2024.3384765

Huang, T., Huang, W., and Zhang, Z. (2025). How artificial intelligence empowers teachers' professional development: Theoretical models and practical directions. *Mod. Distance Educ. Res.* 37, 35–44. doi: 10.3969/j.issn.1009-5195.2025.01.004

Ibrahim, H., Asim, R., Zaffar, F., Rahwan, T., and Zaki, Y. (2023). Rethinking homework in the age of artificial intelligence. *IEEE Intell. Syst.* 38, 24–27. doi: 10.1109/MIS.2023.3255599

Jackson, S. A., and Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: the flow state scale. *J. Sport Exercise Psychol.* 18, 17–35. doi: 10.1123/jsep.18.1.17

Kalota, F. (2023). A primer on generative artificial intelligence. *Educ. Sci.* 14:172. doi: 10.3390/educsci14020172

Kelly, A., Sullivan, M., and Strampel, K. (2023). Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines. *J. Univ. Teach. Learn. Pract.* 20:12. doi: 10.53761/1.20.6.12

Kim, E., and Ahn, K. (2023). Exploring the potential of utilizing ChatGPT in elementary English lesson planning and English language teaching. *Korea Assoc. Primary English Educ.* 29, 59–87. doi: 10.25231/pee.2023.29.4.59

Lam, R. (2016). Assessment as learning: examining a cycle of teaching, learning, and assessment of writing in the portfolio-based classroom. *Stud. Higher Educ.* 41, 1900–1917. doi: 10.1080/03075079.2014.999317

Lan, Y.-J., and Chen, N.-S. (2024). Teachers' agency in the era of LLM and generative AI: designing pedagogical AI agents. *Educ. Technol. Soc.* 27, I-XVIII. doi: 10.30191/ETS.202401_27(1).PP01

Lee, D., and Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Comput. Educ.* 191, 1–17. doi: 10.1016/j.compedu.2022.104646

Li, S., Gu, X., Li, R., et al. (2023). Artificial intelligence education development index: connotation, framework, and measurement. *Open Educ. Res.* 29, 21–33.

Li, X., Jing, J., Yin, Y., et al. (2020). Utilizing experimental micro-videos to enhance teaching competencies of teacher education students. *Lab. Res. Explor.* 39, 224–227. doi: 10.3969/j.issn.1006-7167.2020.05.049

Li, Z., Chen, N., and Wu, G. (2023). Research on the "4H" mirror model for enhancing teaching skills of teacher education students under the "Strong Teachers Plan." Heilongjiang Higher *Educ. Res.* 9, 154–160. doi: 10.13445/j.cnki.t.e.r.2023.04.004

Liu, L., Cao, S., and Wu, X. (2023). The impact of feedback strategies in microteaching on the development of teaching reflection skills among teacher education students: a multimodal data-evidence-based perspective. *Teach. Educ. Res.* 35, 41–49.

Luckin, R., Holmes, W., Griffiths, M., and Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. London: Pearson.

Maher, S., Kayte, S., and Nimbhore, S. (2020). Chatbots its techniques using AI: a review. *Int. J. Res. Appl. Sci. Eng. Technol.* 8, 503–508. doi: 10.22214/ijraset.2020.32537

Markel, J. M., Opferman, S. G., Landay, J. A., et al. (2023). "GPTeach: interactive TA training with GPT-based students," in *Proceedings of the Tenth ACM Conference on Learning@Scale* (New York: ACM), 226–236. doi: 10.1145/3573051.3593393

Mo, C., and He, X. (2007). Development of a lesson plan compilation and evaluation indicator system. *Educ. Vocational* 23, 110–112.

Monika, M. (2024). A study on analyzing the role of ChatGPT in English acquisition among ESL learners during English language classroom. *Bodhi Int. J. Res. Humanit. Arts Sci.* 8, 75–84. doi: 10.13140/RG.2.2.28252.56961

Nicol, D., and Selvaretnam, G. (2021). Making internal feedback explicit: harnessing the comparisons students make during two-stage exams. *Assess. Eval. Higher Educ.* 2, 1–16. doi: 10.31234/osf.io/jrc6y

Nussberger, AM., Luo, L., Celis, LE., and Crockett, MJ. (2022). Public attitudes value interpretability but prioritize accuracy in artificial intelligence. *Nat. Commun.* 13:5821. doi: 10.1038/s41467-022-33417-3

Okonkwo, C. W., and Ade-Ibijola, A. (2021). Chatbots applications in education: a systematic review. *Comput. Educ. Artif. Intell.* 2:100033. doi: 10.1016/j.caeai.2021.100033

OpenAI (2023). *Teaching with AI [EB/OL]*. OpenAI.

Parent-Rocheleau, X., and Parker, S. K. (2022). Algorithms as work designers: how algorithmic management influences the design of jobs. *Hum. Resour. Manag. Rev.* 32:100838. doi: 10.1016/j.hrmr.2021.100838

Parsons, B., and Curry, J. H. (2024). Can ChatGPT pass graduate-level instructional design assignments? *Potential implications of artificial intelligence in education and a call to action. TechTrends* 68, 67–78. doi: 10.1007/s11528-023-00912-3

Peikos, G., and Stavrou, D. (2025). ChatGPT for science lesson planning: an exploratory study based on pedagogical content knowledge. *Educ. Sci.* 15:338. doi: 10.3390/educsci15030338

Qian, L., Li, W.-H., Gu, T.-X., and Wang, X.-H. (2025). Does the use of generative artificial intelligence help improve students' learning outcomes? A meta-analysis based on 39 experimental and quasi-experimental research literatures. *Mod. Educ. Technol.* 35, 36–45. doi: 10.3969/j.issn.1009-8097.2025.08.004

Rath, S., Pattanayak, A., Tripathy, S., Priyadarshini, S. B. B., Tripathy, A., Tanvi, S., et al. (2023). Prediction of a novel rule-based chatbot approach (RCA) using natural language processing techniques. *Int. J. Intell. Syst. Appl. Eng.* 11, 318–325.

Sharma, A., Lin, I. W., Miner, A. S., et al. (2023). Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat. Mach. Intell.* 5, 46–57. doi: 10.1038/s42256-022-00593-2

Strzelecki, A., Cicha, K., Rizun, M., and Rutecka, P. (2024). Acceptance and use of ChatGPT in the academic community. *Educ. Inf. Technol.* 29, 22943–22968. doi: 10.1007/s10639-024-12765-1

Sun, H. (2025). Characteristics, risks, and countermeasures of generative AI usage among primary and secondary school students: a survey analysis based on 8,563 students from seven provinces across China. *Primary Secondary School Manag.* 10, 29–33.

Wickard, C. J., Fyfe, E. R., and Cur, D. L. (2025). A new lesson plan design and development assistant? Using ChatGPT to support pre-service teachers. *Int. J. Digit. Soc.* 16, 1972–1981. doi: 10.20533/ijds.2040.2570.2025.0246

Wu, F., Xia, Z., and Gao, S. (2025). Reimagining smart learning environments with generative artificial intelligence: from element optimization to ecosystem reconstruction. *Res. Electro-Educ.* 1, 54–63. doi: 10.13811/j.cnki.eer.2025.01.008

Wu, N., Chen, X., and Feng, Y. (2023). From "disorder" to "order": the shift in educational applications of generative artificial intelligence and its generative mechanism. *J. Distance Educ.* 41, 42–51. doi: 10.15881/j.cnki.cn33-1304/g4.2023.06.005

Xia, L., Shen, K., Wang, Y., and Dong, Y. (2025). Student learning agency in generative artificial intelligence contexts: Current status and influencing factors. *Mod. Distance Educ*. 37–47. doi: 10.13927/j.cnki.yuan.20250416.005

Xiao, W., Pian, Y., and Ma, X. (2025). Current status and training optimization recommendations for artificial intelligence education applications in primary and secondary school teachers: an empirical analysis based on large-scale survey data. *China Electric Edu.* 10, 114–122.

Xinhua News Agency (2025). *The Central Committee of the Communist Party of China and the State Council issued the "Outline for Building a Leading Country in Education (2024-2035)*". Beijing, China: Xinhua News Agency.

Xue, A., and Fang, F. (2024). Framework structure, basic manifestations and training strategies for digital teaching capabilities of English teacher education students. *China Electric Educ*. 12, 123–129. doi: 10.3969/j.issn.1006-9860.2024.12.017

Yan, L., and Qin, P. (2023). The value, challenges, and countermeasures of generative artificial intelligence in enhancing ideological and political education. *Guangxi Soc. Sci*. 9, 201–206. doi: 10.3969/j.issn.1004-6917.2023.09.026

Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., and Nasser, M. (2023). A systematic literature review on the applications of robots and natural language processing in education. *Electronics* 12:2864. doi: 10.3390/electronics12132864

Yusuf, A., Pervin, N., and Román-González, M. (2024). Generative AI and the future of higher education: a threat to academic integrity or reformation? Evidence from multicultural perspectives. *Int. J. Educ. Technol. High. Educ*. 21:21. doi: 10.1186/s41239-024-00453-6

Zaragoza, A., Seidel, T., and Hiebert, J. (2024). Exploring preservice teachers' abilities to connect professional knowledge with lesson planning and observation. *Eur. J. Teach. Educ.* 47, 120–139. doi: 10.1080/02619768.2021.1996558

Zaugg, T. (2024). *EL-Education and learning in an inclusive environment [Large language model]*. Innoved Consulting. Available online at: https://innoved.consulting/EL

Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education-Where are the educators? *Int. J. Educ. Technol. High. Educ*. 16:39. doi: 10.1186/s41239-019-0171-0

Zheng, L., Gao, L., and Huang, Z. (2024). Can conversational robots based on generative AI technology enhance online collaborative learning performance? *Res. Electro-education*, 45, 70–76+84.

Zhu, Z., Dai, L., Zhao, X., et al. (2024). Cultivating new-type talents: the new mission of education in the digital and intelligent era. *Res. Electro-education* 45, 52–60. doi: 10.13811/j.cnki.eer.2024.01.007