Teacher Education and Curriculum Studies
2026, Vol. 11, No. 3, pp. 127–136
https://doi.org/10.11648/j.tecs.20261103.13

Research Article

AI-Empowered Cloud Mentoring for Cross-Institutional Teacher Development: A Triadic Synergy Model and Its Effectiveness

Kaidan Deng[1, 2, *], Yantian Xu[1]

[1]School of Education Science, Nannjing Normal University, Nanjing, China

[2]Institute of EduInfo Science & Technology, Nanjing Normal University, Nanjing, China

Abstract

The Collaborative Quality Improvement Program for Teacher Education in China, launched by the Ministry of Education in 2022, relies primarily on cross-institutional mentoring, a mechanism that is often constrained by geographical distance and insufficient individualized guidance, as well as a lack of systematic effectiveness evaluation. Focusing on AI-empowered "cloud mentoring", this paper constructs a triadic synergy paradigm (expert teacher, AI system, and novice teacher) that reengineers cross-institutional mentoring into four data-driven phases: diagnosis, matching, mentoring, and reflection. To evaluate the effectiveness of this model, a quasi-experimental mixed-methods design with 23 mentoring pairs (12 in the experimental group and 11 in the control group, supported by 10 expert mentors) was employed over an 18-week intervention period. Data were collected through multimodal classroom video analysis, TPACK scale assessments, and coding of mentoring dialogue transcripts. The results demonstrate that the AI-enhanced model significantly improved novice teachers' classroom effectiveness overall ($F = 5.43$, $p = 0.029$, $\eta^2 p = 0.21$) and deepened mentoring discourse quality, with higher-order cognitive engagement rising from 41.3% in the control group to 62.4% in the experimental group. Technological knowledge (TK) also showed significant improvement ($F = 5.89$, $p = 0.023$), teachers perceived the AI tools as useful (PUM = 4.31) but raised legitimate concerns regarding data privacy and algorithmic transparency. The study contributes a triadic synergy theoretical model that extends teacher professional development research and provides a scalable, empirically grounded blueprint for high-quality cross-institutional mentoring, while also discussing critical tensions between technological empowerment and humanistic care that must be managed for ethical implementation.

Keywords

Teacher Education Collaboration, AI-empowered Mentoring, Cross-institutional Mentoring, Intelligent Tutoring Systems, Effectiveness Evaluation

1. Introduction

Teachers are the primary resource for educational development. In 2022, China's Ministry of Education launched the Collaborative Quality Improvement Program for Teacher Education, aiming to raise teacher education standards through

*Correspondence: Kaidan Deng (msdengkd@163.com)

Received: 26 July 2026; **Accepted:** 6 August 2026; **Published:** 2 September 2026

cluster-based assistance from high-level normal universities to less-developed ones [1]. Personnel exchange, particularly cross-institutional mentoring, has become the core implementation pathway. However, a survey of 387 participating teachers in 2024 revealed three structural dilemmas: 62.3% reported insufficient interaction due to spatial–temporal separation; 55.7% cited a lack of precise diagnosis of mentees' needs; and 48.9% noted the absence of systematic documentation for effectiveness evaluation [2]. These findings expose the fundamental challenge: traditional cross-institutional mentoring, constrained by geography and reliant on subjective experiential judgment, struggles to achieve precision, depth, and accountability.

Concurrently, artificial intelligence (AI) technologies are maturing rapidly. Natural language processing now enables automated analysis of classroom discourse; computer vision captures non-verbal teaching behaviors; and learning analytics allows fine-grained modeling of teacher development trajectories [3, 4]. These advances offer transformative potential for overcoming the limitations of conventional mentoring. This study therefore explores how AI can be deeply embedded in the cross-institutional mentoring process to realize model innovation and enhanced effectiveness. Three questions guide this inquiry: 1) What new form should AI-empowered cross-institutional mentoring take? 2) Can it effectively improve novice teachers' competencies, and through what mechanisms? (3) What challenges arise in practical application, and how can routine implementation be ensured?

2. Literature Review

2.1. Teacher Mentoring and Professional Learning

Teacher mentoring, as one of the most enduring support models in professional development, has been consistently shown to enhance self-efficacy, instructional quality, and retention [5, 6]. Its effectiveness, however, hinges on high-frequency interaction, diagnostic precision, and relational trust-conditions severely tested in cross-institutional contexts. A meta-analysis of distance mentoring programs found that only 34% of purely online pairs maintained substantive exchanges more than twice a month, compared with 78% for same-institution pairs [7]. Beyond reduced frequency, cross-institutional mentors often lack sufficient knowledge of the mentee's teaching context, student demographics, and institutional culture, which diminishes the relevance and specificity of their guidance [8]. These limitations highlight the difficulty of transplanting traditional mentoring models directly into remote, inter-institutional settings without substantial adaptation.

Early attempts to address these challenges leveraged information and communication technologies (ICT), such as video conferencing and online forums, to facilitate remote communication. These technology-assisted models partially alleviated temporal and spatial constraints, yet they largely positioned technology as a mere communication conduit [9]. The core cognitive processes of mentoring—observation, diagnosis, feedback, and reflection—remained essentially unchanged, with limited gains in the depth of professional learning [10]. Researchers have thus called for more transformative integration of technology that can directly augment the intellectual work of mentoring rather than simply mediating interaction [11].

The theoretical foundations of effective mentoring are deeply rooted in the nature of teacher professional learning. Reflective practice, as articulated by Schön [12], emphasizes the importance of moving beyond technical rationality to engage in reflection-in-action and reflection-on-action. For novice teachers, this reflective capacity is best developed through sustained dialogue with more experienced others who can make tacit knowledge explicit. Korthagen [13] extended this perspective by proposing a multi-level model of reflection that spans environment, behavior, competencies, beliefs, identity, and mission. Similarly, sociocultural theories of learning emphasize that professional growth occurs through mediated activity and dialogic interaction within communities of practice [14, 15]. In mentoring, the quality of dialogue—particularly the extent to which it moves from information exchange toward critical co-construction is a key determinant of learning outcomes [16]. These perspectives suggest that effective mentoring must not only transmit knowledge but also scaffold higher-order thinking and reflective capacity, a requirement that places considerable demands on mentors' diagnostic and dialogic skills.

2.2. Feedback and Reflection in Mentoring

A critical function of mentoring is the provision of feedback that enables mentees to close the gap between current and desired performance. Hattie and Timperley's [17] synthesis of feedback research identifies three key questions that effective feedback must address: Where am I going? How am I going? Where to next? For feedback to be effective, it must be specific, timely, and focused on the task rather than the person. However, in traditional cross- institutional mentoring, feedback often suffers from being too general, too delayed, and insufficiently grounded in concrete evidence from the mentee's actual teaching. When mentors rely solely on memory and general impressions formed during a classroom observation, their feedback tends toward vague exhortations rather than actionable, evidence- based guidance.

The role of reflection in transforming feedback into professional learning cannot be overstated. Meaningful reflection involves not merely recalling what happened in a lesson, but critically examining the assumptions, values, and reasoning that underpin pedagogical decisions [18]. Yet, novice teachers often lack the meta-cognitive frameworks and observational acuity to reflect deeply on their own practice without structured support. Video-based reflection has been shown to be more powerful than memory-based reflection, as it allows teachers to observe their own teaching from an external perspective [19]. AI-enhanced video analysis can potentially amplify this benefit by automatically identifying critical moments that teachers might otherwise overlook, thus scaffolding the development of more nuanced professional vision.

2.3. AI Applications in Teacher Education

Recent years have witnessed a paradigm shift in AI applications within teacher education, moving from auxiliary tools toward cognitive partners [3]. Three directions are particularly relevant. First, AI-driven classroom analysis leverages speech recognition, natural language processing, and computer vision to automatically identify discourse structures (e.g., IRE sequences), question types, and student engagement distributions, enabling scalable, objective instructional diagnosis [20, 21]. Second, intelligent tutoring systems (ITS) for teacher learning simulate expert cognitive processes to generate personalized feedback based on teaching behavior data and to recommend targeted learning resources; while effective for skill acquisition, ITS faces limitations in handling complex pedagogical decisions and providing emotional support [22]. Third, learning analytics integrates multi-source data to construct multidimensional teacher development profiles, making growth trajectories visible and actionable [23, 24].

A particularly promising but underexplored area is the use of AI to support teacher reflection. Existing research on video-based teacher reflection has demonstrated that structured viewing protocols and guided noticing frame- works can enhance the quality of reflection [25]. AI technologies can automate aspects of this scaffolding by highlighting pedagogically significant moments, generating descriptive statistics about instructional patterns, and posing reflective questions based on observed data. This capacity positions AI as not merely an analytical tool but as a potential "reflection partner" that can prompt deeper engage- ment with the complexities of teaching. However, the design of such AI-mediated reflection must carefully balance automation with teacher agency, ensuring that the technology supports rather than supplants teachers' own meaning-making processes [26].

To realize this potential, however, it is necessary to examine how AI-enhanced video analysis concretely outperforms conventional video-based approaches in facilitating accurate teacher reflection. A critical advantage lies in its capacity to mitigate behavioral biases that commonly compromise teacher reflection—namely, selective attention, confirmation bias, and retrospective reconstruction [25]. Whereas conventional video analysis leaves these biases largely unchecked, as it relies on manual judgment and the viewer's existing professional vision [19], AI counteracts them through systematic, data-driven, and timestamped analytical capabilities that flag critical moments, provide neutral evidence, and anchor discussion to observable classroom events. In this way, AI-enhanced video analysis not only automates data extraction but also introduces an objective analytical perspective that augments human perception [26].

2.4. Ethical and Critical Perspectives

Alongside technological advances, a growing body of literature has begun to examine the ethical and practical challenges of AI in education. Concerns have been raised about algorithmic bias, data privacy, and the potential deprofessionalization of teaching when AI systems are deployed without adequate human oversight [27, 28]. Selwyn [26] cautions against technological solutionism, arguing that AI should be understood not as a neutral tool but as a socio-technical system that reshapes power relations and professional identities. Holmes et al. [29] similarly emphasize the need for human-centered AI that augments rather than replaces teacher agency. These critical perspectives are essential for ensuring that AI-empowered mentoring models remain aligned with the core values of teacher professionalism.

Despite these advances, two notable gaps persist. First, there is a disconnect between technology demonstration and theoretical grounding: many studies showcase the technical feasibility of AI tools without adequately explaining why they work from the perspective of teacher learning mechanisms [26]. Second, a dichotomous mindset prevails that positions AI as a replacement for, rather than an augmentation of, human mentors, leaving the potential of "human–human–machine" triadic synergy largely unexplored [29]. This study enters precisely at these gaps, constructing a triadic synergy model that reconceptualizes the relationship among expert teacher, AI system, and novice teacher, and empirically evaluating its effectiveness.

3. The Triadic Synergy Model

3.1. Triadic Synergy Architecture

Traditional mentoring operates through expert-driven dyadic actuation, with the mentor as the exclusive knowledge source. The introduction of AI enables a triadic synergy architecture in which three actors assume differentiated and complementary roles [30]. The expert teacher serves as the "wisdom core," providing context-sensitive pedagogical insight, value judgment, and emotional support—capabilities that AI cannot replicate. The AI system functions as the "data engine," undertaking multimodal data collection, automated analysis, and personalized resource recommendation, thereby freeing the mentor from repetitive observational labor to concentrate on higher-order guidance [31]. The novice teacher becomes an "active constructor", empowered to independently retrieve diagnostic data, compare it against personal perceptions, and engage in reflective dialogue under the mentor's guidance.

The operational logic follows a closed loop: AI continuously collects and analyzes the mentee's teaching behavior data to generate a precise diagnostic profile; the mentor interprets the AI diagnosis through the lens of professional judgment to determine guidance focus and strategy; the mentee applies suggestions in practice, and new behavioral data are captured and analyzed by AI, enabling a "diagnosis–guidance–practice–re-diagnosis" iterative cycle. Through this loop, the three parties' synergy deepens: the AI system improves diagnostic accuracy with accumulating data; the mentor enhances guidance efficiency; and the mentee accelerates professional growth through precision feedback [32].

3.2. Four-Phase Mentoring Process

The core mentoring process is reengineered into four data-driven phases. Diagnosis: Multimodal analysis intgrates speech (teacher–student speech ratio, speech rate), text (question type distribution according to Bloom's taxonomy, feedback strategy, sentiment polarity), and behavior (positional movement, gaze distribution) channels to construct a digital teaching portrait and generate a diagnostic radar chart [21]. Matching: A two-way algorithm considers multidimensional competency profiles—mentee profiles encompass diagnostic results, self-reported development needs, and disciplinary background; mentor profiles encompass expertise areas, mentoring experience, style (directive, elicitive, or collaborative), and avai- lability—yielding recommendations from which pairs autonomously confirm their partnership [33]. Mentoring: AI-powered lesson episode analysis automatically annotates key clips from classroom videos based on teaching phases or critical pedagogical events, generating preliminary data reports. These reports serve as an evidence base for focused, data-informed remote lesson polishing, shifting feedback from impressionistic judgment to precise, evidence-grounded dialogue [17]. Reflection: The system dynamically updates teaching competency growth curves and identifies "suggestion–improvement" correspondences, scaffolding mentees' self-regulated learning and helping them transition from being guided to becoming proactive developers of their own practice [24].

3.3. Platform Architecture

The intelligent collaborative mentoring platform adopts a browser/server architecture with four layers: data collection (recording systems, speech recognition, video analysis algorithms); intelligent analysis (discourse analysis, behavior recognition, competency diagnosis models); application services (lesson management, clip annotation, remote polish- ing, growth portfolios, resource recommendation); and user interaction (Web and mobile access). The technological stack is specified as follows. For speech-to-text transcription, the platform employs the iFLYTEK voice recognition SDK (Chinese Mandarin model, version 5.0), which achieves an accuracy rate exceeding 95% for standard classroom Mandarin and automatically generates timestamped transcripts. For discourse analysis, a custom Python-based natural language processing (NLP) pipeline using the THULAC (THU Lexical Analyzer for Chinese) toolkit performs segmentation and part-of-speech tagging, followed by a rule-based classifier that identifies IRE sequences, classifies questions according to Bloom's taxonomy, and analyzes sentiment polarity. For behavioral recognition, OpenPose (version 1.7) is used for skeletal tracking of teacher positional movement, combined with a gaze-tracking algorithm developed in-house using OpenCV to capture attention distribution patterns. These multimodal indicators are integrated into a rule-based inference engine that generates competency diagnoses aligned with CLASS and TPACK rubrics, prioritizing interpretability over black-box machine learning to ensure transparency and professional oversight. The backend is built on Python 3.9 with Django framework, MySQL database, and Nginx server, hosted on a private cloud infrastructure. Data security and privacy are protected through de-identification of video data, tiered access permissions, and informed consent protocols [34].

4. Effectiveness Evaluation

4.1. Evaluation Framework

A three-dimensional framework of "competency development, process quality, and subjective perception" was constructed. Competency development: Classroom teaching effectiveness was measured using a simplified Classroom Assessment Scoring System (CLASS) across Emotional Support, Classroom Organization, and Instructional Support (inter-rater Cohen's κ=0.82) [35]; Technological Pedagogical Content Knowledge (TPACK) was assessed via a localized and validated scale (α=0.91) [36]; teaching reflection quality was coded as technical, contextual, or critical based on established reflection hierarchy frameworks [18]. Process quality: Behavioral indicators included interaction frequency, duration, and feedback timeliness; dialogue content was coded into information exchange, interpretive analysis, and critical co-construction, with higher-order dialogue proportion serving as the core depth indicator [16]. Subjective perception: A modified Technology Acceptance Model questionnaire measured Perceived Usefulness (PU) and Perceived Ease of Use (PEOU), supplemented by open- ended questions on user experience and ethical concerns.

4.2. Research Design

A quasi-experimental design was employed with 23 "cloud mentoring" pairs between C Normal University and L Normal College. Ten expert teachers with associate senior or higher titles and over 15 years of teaching experience served as mentors, and 23 early-career teachers with less than five years of experience as mentees. Participants were randomly assigned to an experimental group (N=12) adopting the AI-empowered model and a control group (N=11) continuing traditional online mentoring (video-conference lesson polishing with written feedback). Baseline equivalence between groups was confirmed on teaching experience, subject taught, and pre-test CLASS scores ($p>0.05$). The 18-week study comprised: pre-test (baseline classroom videos, TPACK pre-test, platform training for the experimental group); intervention (four 3–4 week cycles of diag- nosis–polishing–practice–re-diagnosis for the experimental group, with equivalent frequency requirements for the control group); and post-test (classroom videos, TPACK post-test, satisfaction questionnaire, semi-structured interviews). To eliminate observer bias in CLASS coding, all raters were trained researchers who were blind to group assignments and the study's hypotheses. Classroom videos were de-identified (with teacher names, school logos, and any contextual references to group assignment removed) and randomly ordered for assessment. Pre-test and post-test videos were coded in separate batches to prevent recall bias. The reported inter-rater reliability (Cohen's $\kappa = 0.82$) was calculated under these blinded conditions. It is important to note that all AI-generated annotations and diagnostic reports served as reference inputs for the mentoring conversations; the final feedback and guidance decisions were made by the expert mentors, who reviewed and validated the AI outputs against their professional judgment to ensure human oversight and pedagogical appropriateness. Data collected included 92 classroom videos (≈3,974 minutes), complete TPACK scale returns, 92 mentoring records (≈213,000 words of transcribed dialogue), and 17 interview transcripts (≈115,000 words).

A post-hoc statistical power analysis was conducted using G*Power 3.1 to assess the sensitivity of the study design. For an ANCOVA with one covariate and two groups, with $\alpha = 0.05$ and power $(1-\beta) = 0.80$, the minimum detectable effect size is $f = 0.61$ (equivalent to $\eta^2p \approx 0.27$). The observed effect sizes for the primary outcomes ranged from $\eta^2p = 0.19$ to 0.24, indicating that the study is underpowered to reliably detect effects smaller than this threshold. This limitation is acknowledged in the discussion and conclusions.

4.3. Findings

Classroom teaching effectiveness. ANCOVA with pre-test scores as covariates revealed that the experimental group scored significantly higher on Emotional Support, Instructional Support, and Overall Mean (see Table 1). Classroom Organization showed a positive but non-significant trend, possibly requiring a longer intervention to manifest measurable change.

Table 1. *ANCOVA results for post-test classroom teaching effectiveness scores.*

Dimension	Exp. Group M (SD)	Control Group M (SD)	F	p	η^2p
Emotional Support	5.42 (0.78)	4.89 (0.92)	4.67	0.041	0.19
Classroom Organization	5.31 (0.85)	4.94 (0.88)	3.82	0.063	0.16
Instructional Support	5.18 (0.73)	4.61 (0.95)	6.14	0.021	0.24
Overall Mean	5.30 (0.71)	4.81 (0.84)	5.43	0.029	0.21

Note. Pre-test scores statistically controlled as covariates.

TPACK. The experimental group significantly outperformed the control group on Technological Knowledge (TK) and Technological Pedagogical Content Knowledge (TPCK); differences on Pedagogical Knowledge (PK) and Content Knowledge (CK) were not significant (Table 2). This pattern is consistent with the expectation that AI-empowered mentoring directly strengthens technology-related competencies, while broader pedagogical and content knowledge requires longer-term accumulation.

Reflection quality. Higher-order reflection (contextual plus critical) constituted 65.3% of experimental group reflective texts, significantly higher than the control group's 42.7% ($\chi^2 = 9.82$, $p < 0.01$). Interview data corroborated that AI diagnostics helped mentees become aware of previously unnoticed teaching habits and better understand the reasons behind performance gaps.

Table 2. *ANCOVA results for post-test TPACK dimension scores.*

Dimension	Exp. Group M (SD)	Control Group M (SD)	F	p	η^2p
TK	4.61 (0.58)	4.12 (0.67)	5.89	0.023	0.22
PK	4.83 (0.55)	4.67 (0.63)	1.42	0.245	0.07
CK	4.91 (0.52)	4.85 (0.58)	0.86	0.363	0.04
TPACK	4.57 (0.61)	4.23 (0.72)	4.77	0.030	0.19

Process quality. Experimental group sessions averaged 57.4 minutes (vs. 41.8 minutes, $t = 3.17$, $p<0.01$), with shorter feedback intervals (2.1 vs. 3.8 days, $t=3.54$, $p<0.01$). The distribution of cognitive engagement in mentoring dialogues (Table 3) shows that higher-order dialogue reached 62.4% in the experimental group versus 41.3% in the control group ($\chi^2 = 14.37$, $p<0.001$). AI-annotated clips reduced the time cost of video review for mentors and provided objective anchors for discussion, facilitating deeper cognitive engagement.

Table 3. *Distribution of cognitive engagement levels in mentoring dialogue content (%)x`.*

Cognitive Level	Experimental Group	Control Group
Information Exchange	37.6	58.7
Interpretive Analysis	41.2	29.8
Critical Co-construction	21.2	11.5

Technology acceptance. PU mean was 4.31 (SD=0.52) and PEOU 3.97 (SD=0.68); overall satisfaction reached 4.18 for mentees and 4.45 for mentors. Nevertheless, qualitative data revealed concerns centered on data privacy, algorithmic transparency, and the potential undermining of professional intuition by data-driven judgments.

5. Discussion

5.1. Theoretical Contributions and Dialogue with the Literature

The empirical findings of this study contribute to multiple strands of educational research. First, the triadic synergy model extends the traditional dyadic mentoring paradigm by introducing AI as a functional third agent with a distinct role in the cognitive system. This extension resonates strongly with Salomon's [30] distributed cognition theory, which posits that cognitive processes are spread across individuals, artifacts, and social interactions. In the present model, the AI system functions precisely as such a "cognitive artifact," augmenting the perceptual and analytical capacities of both mentor and mentee in ways that enable forms of joint cognitive engagement impossible in conventional settings. This aligns with the Vygotskian [15] notion that psycho- logical tools mediate human activity and fundamentally transform the nature of that activity.

Second, the three mechanisms identified—cognitive offloading, shared reference, and feedback precision—collectively illuminate why and how AI augmentation can enhance mentoring effectiveness. The shared reference mechanism, in particular, offers a novel contribution by showing that AI-generated data does not merely inform the mentor but actively reshapes the dialogic space between mentor and mentee. When AI diagnosis converges with or diverges from human judgment, it creates what expansive learning theory [37] terms "contradictions" that can trigger productive re-examination of practice. In this sense, AI serves as both a mirror reflecting practice back to participants and a wedge that opens up taken-for-granted assumptions for critical scrutiny.

Third, the significant improvement in mentoring dialogue quality—specifically the shift from information exchange toward critical co-construction—can be understood through the lens of dialogic pedagogy and feedback theory. Hattie and Timperley's [17] model emphasizes that effective feed- back must reduce the gap between current and desired performance. The AI-enhanced model achieved this by making the gap visible and specific, transforming vague mentoring conversations into evidence-based, targeted discussions. This finding extends prior work on video-based reflection [19] by demonstrating that AI-powered clip annotation can amplify the reflective affordances of video, directing attention to pedagogically significant moments that might otherwise escape notice.

5.2. Mechanisms of Effectiveness

The three mechanisms underpinning the model's effectiveness warrant further elaboration. Cognitive offloading operates at the level of the mentor's attentional and analytical resources: by automating the labor-intensive tasks of behavioral coding, timing analysis, and pattern recognition, AI frees mentors to engage in the higher-order interpretive work that constitutes expert mentoring. This is consistent with cognitive load theory, which suggests that offloading extraneous cognitive processing enables deeper engagement with essential content. Shared reference transforms the epistemological basis of mentoring dialogue. Instead of relying on the mentor's recollection and the mentee's self-report—both subject to significant bias—the pair co-examines an externalized, data-based representation of teaching. This shared artifact mediates their interaction, grounding it in evidence rather than impression. Feedback precision addresses a persistent weakness of traditional mentoring: the tendency for feedback to be general, delayed, and decoupled from specific behavioral evidence. The AI system's capacity to link feedback directly to timestamped video clips and quantified behavioral indicators made feedback more actionable and growth more visible to mentees, thereby enhancing self-efficacy and motivation.

5.3. Tensions and Ethical Considerations

Important tensions must be managed to prevent technological empowerment from inadvertently undermining professional values. The risk of "data occlusion" refers to the systematic neglect of tacit, relational, and affective dimensions of teaching that resist quantification but are central to quality instruction [26]. When a mentoring conversation becomes overly centered on metrics such as question types per minute or speech ratios, there is a danger of reducing teaching to a set of discrete, measurable behaviors at the expense of holistic pedagogical judgment. Similarly, confusion over professional subjectivity may arise when AI-generated diagnoses conflict with teachers' experiential self-perceptions, potentially engendering a form of epistemic deference that undermines professional autonomy. This study therefore insists on positioning AI as "augmented intelligence," with the expert teacher retaining inviolable authority as the wisdom core.

Ethical governance must adhere to the principles of informed consent, data minimization, algorithmic transparency and explainability, and teacher data sovereignty [34]. The observed trajectory of mentors' relationships with AI—from initial skepticism to instrumental rationality and, eventually, synergistic complementarity—highlights the importance of developing "human–machine collaborative competence". This competence includes the ability to critically interpret AI-generated data, to calibrate between algorithmic suggestions and professional intuition, and to maintain agentic control over pedagogical decisions. Teacher education and mentor preparation programs should therefore incorporate explicit training in AI literacy and human–machine collaboration skills [28].

5.4. Pathways to Routine Implementation

For routine implementation, systemic considerations are paramount. The technical threshold can be lowered through a "cloud services plus lightweight terminals" model, allowing less-resourced institutions to access sophisticated AI analysis without substantial local infrastructure investment. However, sustainable adoption requires more than technological access; it demands institutional embedding. AI-empowered mentoring activities should be formally recognized within teacher evaluation and professional title promotion systems, creating structural incentives for sustained engagement. Furthermore, dedicated budgetary provisions for AI service subscriptions, mentor training, and ongoing technical support are necessary to prevent the initiative from becoming a short-term project rather than an enduring practice. Drawing on lessons from educational change literature, the successful scaling of such innovations depends on alignment across technological, organizational, and policy dimensions, with particular attention to the sensemaking processes through which practitioners come to understand and value new practices [37].

Before turning to broader implications, it is important to consider whether the 18-week (four-cycle) intervention duration was sufficient to capture meaningful changes across all measured dimensions. The non-significant trends observed in Classroom Organization ($p = 0.063$, $\eta^2p = 0.16$) and in Pedagogical Knowledge and Content Knowledge (PK: $p = 0.245$, $\eta^2p = 0.07$; CK: $p = 0.363$, $\eta^2p = 0.04$) warrant careful interpretation. Classroom Organization encompasses time management, behavior management, and instructional pacing—competencies that typically require sustained, iterative practice over multiple semesters to fundamentally reshape [35]. While our 18-week intervention showed positive directional trends, the lack of statistical significance likely reflects the need for extended scaffolding beyond a single semester. Similarly, PK and CK represent deep disciplinary knowledge and pedagogical content knowledge that are generally accumulated over years of teaching experience and broader professional learning activities, such as curriculum study, subject-specific professional development, and peer collaboration [36]. In contrast, the significant improvements in Emotional Support, Instructional Support, TK, and TPACK are best understood as short-to-medium term gains—competencies that are more amenable to immediate, data-driven feedback and behavioral adjustment. These distinctions suggest that different dimensions of teacher competence operate on different developmental timescales. We therefore frame the current findings as evidence of the model's capacity to catalyze proximal outcomes, while acknowledging that distal outcomes—particularly those related to classroom management and deep content knowledge—may require longer-term interventions and complementary professional learning experiences. Future longitudinal research with extended intervention periods is needed to assess the sustainability and growth trajectory of these competencies.

Two practical implications for teacher preparation programs follow from these findings. First, training should prioritize critical interpretation over technical operation. Our data showed that teachers' relationships with AI evolved from skepticism to instrumental rationality and ultimately to synergistic complementarity—a trajectory that required sustained engagement with AI-generated outputs. Rather than focusing on how to use AI tools, preparation programs should emphasize how to interrogate them: identifying potential biases in automated diagnoses, recognizing the limits of quantified teaching indicators, and practicing “human override” decisions when algorithmic suggestions conflict with professional intuition. Second, training should embed structured comparison exercises that directly confront teachers with divergences between AI-generated diagnoses and expert mentor judgments. Our findings suggest that productive learning occurred precisely at moments of tension—when AI diagnosis converged with or diverged from human judgment, it created “contradictions” that triggered productive re-examination of practice. Teacher preparation programs can simulate such tensions through case-based activities where teachers compare AI reports with expert evaluations of the same lesson, articulate rationales for accepting or rejecting algorithmic suggestions, and develop calibrated professional judgment. These implications directly address the risk of “data occlusion” discussed above, positioning AI as a stimulus for critical reflection rather than a substitute for pedagogical reasoning.

6. Conclusion

This study constructed and empirically validated a triadic synergy paradigm of AI-empowered cross-institutional mentoring within the context of the Collaborative Quality Improvement Program. The model significantly enhanced novice teachers’ classroom teaching effectiveness, deepened the cognitive quality of mentoring dialogue, and strengthened their capacity to integrate technology with pedagogy. The underlying mechanisms are cognitive offloading, shared reference, and feedback precision. Theoretically, the study extends the traditional dyadic mentoring framework by introducing AI as a functional third agent, thereby contributing a novel conceptual model for understanding teacher professional learning in the AI era. Practically, it provides an empirically grounded, scalable blueprint for precision teacher assistance that can inform policy and practice within the Collaborative Quality Improvement Program and similar large-scale teacher development initiatives.

Teachers’ reasonable concerns regarding data privacy, algorithmic fairness, and the preservation of professional autonomy underscore the imperative of aligning technological advancement with ethical reflection. The triadic synergy model’s insistence on the expert teacher as the wisdom core, with AI positioned as augmented intelligence, offers a principled framework for navigating the tension between technological empowerment and humanistic care. Limitations include a modest sample size (N = 23 mentoring pairs, with 12 in the experimental group and 11 in the control group), which constrains statistical power (the study is powered to detect only large effects of $f \geq 0.61$) and limits generalizability; a single-semester intervention period that precludes examination of long-term sustainability of the observed effects; and a single-case institutional context that restricts the applicability of findings to other educational settings. Therefore, the findings should be interpreted as preliminary evidence from an exploratory proof-of-concept study, and future research with larger, more diverse samples is needed to confirm and extend these results. Future research should incorporate larger, more diverse samples, extended longitudinal designs that track the durability of observed effects, and explorations of integrating generative AI tools—such as large language models—into the collaborative mentoring process. Ultimately, AI-empowered mentoring is not about replacing human wisdom with machine intelligence, but about using technology to amplify the distinctly human capacities for pedagogical judgment, relational attunement, and professional growth.

Abbreviations

AI	Artificial Intelligence
CK	Content Knowledge
CLASS	Classroom Assessment Scoring System
ICT	Information and Communication Technologies
ITS	Intelligent Tutoring Systems
PEOU	Perceived Ease of Use
PK	Pedagogical Knowledge
PU	Perceived Usefulness
TK	Technological Knowledge
TPACK	Technological Pedagogical Content Knowledge

Author Contributions

Kaidan Deng: Conceptualization, Formal Analysis, Funding acquisition, Validation, Writing – review & editing

Yantian Xu: Data curation, Formal Analysis, Methodology, Resources, Validation

Funding

The work in this paper is supported by the Key Project of the "Collaborative Quality Improvement Plan for Teacher Education" of the Chinese Society of Education (Grant No. 202400002207ZXB).

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Ministry of Education, P. R. C. Implementation Plan for the Collaborative Quality Improvement Program for Teacher Education. 2022. (in Chinese)

[2] Lu, X., Liu, X. School-Based Teaching Research Empowered by Big Data: Value, Models, and Implementation Strategies. Journal of Teacher Education, 2024, 11(4): 47-57. http://doi.org/10.13718/j.cnki.jsjy.2024.04.006

[3] Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. Intelligence Unleashed: An Argument for AI in Education. London: Pearson, 2016.

[4] Zawacki-Richter, O., Marín, V., Bond, M., et al. Systematic Review of Research on Artificial Intelligence Applications in Higher Education. International Journal of Educational Technology in Higher Education, 2019, 16(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

[5] Hobson, A., Ashby, P., Malderez, A., et al. Mentoring Beginning Teachers: What We Know and What We Don't. Teaching and Teacher Education, 2009, 25(1), 207–216. http://dx.doi.org/10.1016/j.tate.2008.09.001

[6] Ingersoll, R., Strong, M. The Impact of Induction and Mentoring Programs for Beginning Teachers: A Critical Review of the Research. Review of Educational Research, 2011, 81(2), 201–233. https://doi.org/10.3102/0034654311 403323

[7] Dawson, P., Henderson, M., Mahoney, P., et al. What Makes for Effective Feedback: Staff and Student Perspectives. Assessment & Evaluation in Higher Education, 2019, 44(1), 25–36. https://doi.org/10.1080/02602938.2018.1467877

[8] Bullough, R. V. Mentoring and New Teacher Induction in the United States: A Review and Analysis of Current Practices. Mentoring & Tutoring: Partnership in Learning, 2012, 20(1), 57–74. https://doi.org/10.1080/13611267.2012.645600

[9] Kram, K. E., Isabella, L. A. Mentoring Alternatives: The Role of Peer Relationships in Career Development. Academy of Management Journal, 1985, 28(1), 110–132. https://doi.org/10.2307/256064

[10] Ensher, E. A., Heun, C., & Blanchard, A. Online Mentoring and Computer-Mediated Communication: New Directions in Research. Journal of Vocational Behavior, 2003, 63(2), 264–288. https://doi:10.1016/S0001-8791(03)00044-7

[11] Dede, C. Comparing Frameworks for 21st Century Skills. In J. Bellanca & R. Brandt (Eds.), 21st Century Skills: Rethinking How Students Learn. Bloomington: Solution Tree Press, 2010, pp. 51–76.

[12] Schön, D. The Reflective Practitioner: How Professionals Think in Action. New York: Basic Books, 1983.

[13] Korthagen, F. In Search of the Essence of a Good Teacher: Towards a More Holistic Approach in Teacher Education. Teaching and Teacher Education, 2004, 20(1), 77–97. http:// dx.doi.org/10.1016/j.tate.2003.10.002

[14] Lave, J., Wenger, E. Situated Learning: Legitimate Peripheral Participation. Cambridge: Cambridge University Press, 1991. https://doi.org/10.1017/CBO9780511815355

[15] Vygotsky, L. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press, 1978.

[16] Garrison, D., Anderson, T., Archer, W. Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. American Journal of Distance Education, 2001, 15(1), 7–23. https://doi.org/10.1016/S1096-7516(00) 00016-6

[17] Hattie, J., Timperley, H. The Power of Feedback. Review of Educational Research, 2007, 77(1), 81–112. https://doi.org/ 10.3102/003465430298487

[18] Hatton, N., Smith, D. Reflection in Teacher Education: Towards Definition and Implementation. Teaching and Teacher Education, 1995, 11(1), 33–49. https://doi.org/10. 1016/0742-051x(94)00012-u

[19] Gaudin, C., Chaliès, S. Video Viewing in Teacher Education and Professional Development: A Literature Review. Educational Research Review, 2015, 16, 41–67. https://doi.org/10.1016/j.edurev.2015.06.001

[20] Chen, G., Clarke, S., Resnick, L. Classroom Discourse Analysis: A Tool for Critical Reflection. International Journal of Educational Research, 2015, 72, 80–93.

[21] Wang, D., Liu, H., Qiu, M. Analysis Method and Application Verification on Teacher Behavior Data in Smart Classroom. Instruction and Teacher Professional Development, 2020, (05): 120-127. (in Chinese)

[22] VanLehn, K. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 2011, 46(4), 197–221. http://doi.org/10.1080/00461520.2011.611369

[23] Siemens, G. Learning Analytics: The Emergence of a Discipline. American Behavioral Scientist, 2013, 57(10), 1380–1400. http://doi.org/10.1177/0002764213498851

[24] Zhu, Z., Peng, H. Technology-enabled Pathways to Smart Education Practice. Journal of the Chinese Society of Education, 2020, (10), 1–8. (in Chinese)

[25] van Es, E., Sherin, M. Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. Journal of Technology and Teacher Education, 2002, 10(4), 571–596.

[26] Selwyn, N. Should Robots Replace Teachers? AI and the Future of Education. Cambridge: Polity Press, 2019.

[27] Jobin, A., Ienca, M., Vayena, E. The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence, 2019, 1(9), 389–399. http://doi.org/10.1038/s42256-019-0088-2

[28] Mittelstadt, B., Allo, P., Taddeo, M., et al. The Ethics of Algorithms: Mapping the Debate. Big Data & Society, 2016, 3(2), 1–21. http://doi.org/10.1177/2053951716679679

[29] Holmes, W., Bialik, M., & Fadel, C. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston: Center for Curriculum Redesign, 2019.

[30] Salomon, G. Distributed Cognitions: Psychological and Educational Considerations. Cambridge University Press, 1993.

[31] Ren, C., Luo, Z. Strategy Construction for Digital Empowerment of Rural Teachers' Professional Development: A SWOT Analysis Perspective. Contemporary Teacher Education, 2026, 19(01): 75-80. (in Chinese) https://doi.org/10.16222/j.cnki.cte.2026.01.009

[32] Chen, X., Ding, K. The Connotation Characteristics and Practical Path of Collaborative Quality Improvement in Teacher Education from the Perspective of Synergetics. Theory and Practice of Education, 2024, 44 (24): 51-55. (in Chinese) http://doi.org/10.13532/j.cnki.cn11-1314/g4.2024.24.012

[33] Lin, M., Wu Y., Song H. Teacher Education Transformation in the Age of Artificial Intelligence: Theoretical Stance, Transformation Pathways, and Potential Challenges. Open Education Research, 2024, 30(04): 28-36. (in Chinese) https://doi.org/10.13966/j.cnki.kfjyyj.2024.04.004

[34] Floridi, L., Cowls, J. A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 2019, 1(1): 2-15. https://doi.org/10.1162/99608f92.8cd550d1

[35] Pianta, R., La Paro, K., Hamre, B. Classroom Assessment Scoring System (CLASS) Manual. Baltimore: Brookes, 2008.

[36] Mishra, P., Koehler, M. Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. Teachers College Record, 2006, 108(6), 1017–1054. http://doi.org/10.1111/j.1467-9620.2006.00684.x

[37] Engeström, Y. Learning by Expanding: An Activity- Theoretical Approach to Developmental Research. Helsinki: Orienta-Konsultit, 1987. http://doi.org/10.1017/CBO9781139814744