

Compassionate Feedback in AI-Assisted Assessment: A Comparative Analysis of Novice and Experienced Teachers' Feedback Patterns within the Integrated AI Triad (IAT) Framework

Hossein Talebzadeh

talebzadehdr@gmail.com

Farhangian University <https://orcid.org/0009-0009-4312-1244>

Research Article

Keywords: Compassionate Feedback, IAT Framework, Teacher Experience, AI-Assisted Assessment, Formative Feedback, Pedagogical Reasoning

Posted Date: August 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10642755/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Compassionate Feedback in AI-Assisted Assessment: A Comparative Analysis of Novice and Experienced Teachers' Feedback Patterns within the Integrated AI Triad (IAT) Framework

Hossein Talebzadeh

Department of Social Science, Farhangian University, Tehran, Iran

talebzadehdr@gmail.com

Abstract

The integration of generative AI into educational assessment has transformed feedback practices, yet the extent to which AI-generated feedback supports compassionate teaching—integrating emotional attunement with precise cognitive scaffolding—remains critically underexamined, a gap this study addresses. This study investigates how novice ($n=26$, 0–3 years) and experienced ($n=31$, 7+ years) teachers generate compassionate feedback for weak student responses within the Integrated AI Triad (IAT) Framework. Employing a comparative qualitative content analysis of 57 feedback entries generated across a 16-session professional development workshop, we examined structural patterns, linguistic tone, IAT alignment, and the effects of AI tool selection, subject area, and session progression. Experienced teachers produced significantly more specific feedback ($M=4.0/5$ vs. $M=2.8/5$) with stronger cognitive scaffolding ($M=3.8/5$ vs. $M=3.0/5$), while novices demonstrated marginally higher but less targeted compassion ($M=4.0/5$ vs. $M=3.7/5$). DeepSeek generated more Socratic, layered questioning than ChatGPT or Copilot, though tool benefits were moderated by teacher experience. Feedback quality improved significantly across sessions 1–3, with a plateau in session 4—suggesting condensed three-session workshops may achieve comparable outcomes while reducing implementation costs by approximately 25%. Negative case analysis revealed 19% of novices outperformed the experienced mean and 16% of experienced teachers underperformed the novice mean, indicating that experience alone does not determine feedback quality. The findings extend the IAT framework by operationalising compassionate feedback as a measurable dual-construct practice, with implications for teacher professional development, AI tool selection, and the design of assessment systems that preserve human pedagogical judgment. We discuss limitations, including sample size and the overrepresentation of science topics, and propose directions for future research.

Keywords: Compassionate Feedback, IAT Framework, Teacher Experience, AI-Assisted Assessment, Formative Feedback, Pedagogical Reasoning.

1.Introduction

The integration of artificial intelligence (AI) into educational assessment represents one of the most significant pedagogical shifts of the twenty-first century, yet Scholars have disproportionately fixated on technological capability rather than pedagogical intentionality in the discourse surrounding this transformation (Çallı & Er, 2026; Narreddy et al., 2025). While generative AI tools such as ChatGPT, Copilot, and DeepSeek have demonstrated remarkable capacity to automate feedback generation, We still critically underexamine how these tools meaningfully support or inadvertently undermine the human dimensions of teaching and learning. Central to effective feedback is compassion: a dual-construct practice integrating emotional attunement with precise cognitive scaffolding. Within the Integrated AI Traiad (IAT) Framework (Talebzadeh, 2025), compassionate feedback moves beyond superficial encouragement, encompassing genuine pedagogical reasoning—feedback that validates student effort while guiding self-discovery through strategically designed questions and cognitive prompts. This conceptualisation distinguishes authentic compassion from the superficially warm yet diagnostically shallow responses that large language models typically generate (Çallı & Er, 2026; Messner et al., 2025), thereby establishing a critical distinction underpinning the present investigation.

The IAT framework, developed by Talebzadeh (2025), provides a systemic architecture for AI integration across the educational ecosystem, positioning AI as a mediator between three interdependent pillars: teachers, managers, and families. Central to this model is the transformation of the teacher into an "AI-Integrated Learning Design Partner" (AI-LDP)—a bilingual professional fluent in both human pedagogy and AI technology. Within this role, assessment automation constitutes a critical competency, encompassing not merely the delegation of objective scoring to AI systems but the strategic redeployment of teacher time toward qualitative interaction, personalised guidance, and the cultivation of student self-regulation (Talebzadeh, 2025). This reconceptualisation aligns with the broader imperative that AI should augment rather than replace human judgment (Luckin & Holmes, 2016; Selwyn, 2019), particularly in the emotionally and relationally demanding work of supporting struggling students.

Compassionate feedback, as conceptualised within the IAT framework, resonates with established principles of self-regulated learning, wherein feedback serves not merely as an external evaluation but as a catalyst for students' metacognitive monitoring, strategic adaptation, and motivational sustainment (Nicol & Macfarlane-Dick, 2006; Zimmerman, 1990). Research has consistently demonstrated that effective feedback must address both cognitive and affective dimensions, validating student effort while providing specific, actionable guidance for improvement (Brookhart, 2011; Hattie & Timperley, 2007). Yet AI-generated feedback often remains linguistically encouraging but diagnostically impoverished. (Çallı & Er, 2026; Messner et al., 2025). This tension between affective warmth and cognitive precision constitutes a central problematic that the present study systematically investigates.

Extensive research has examined teachers' use of AI for feedback, but significant gaps remain. First, most studies have focused on the technical quality of AI outputs in isolation, paying insufficient attention to the *teacher-AI interaction* that produces the final feedback delivered to students (Jin et al., 2025; Moorhouse & Kohnke, 2026). Second, the specific dimension of compassion in AI-assisted feedback has received limited empirical attention, with most investigations prioritising accuracy, specificity, or alignment with learning objectives over the qualitatively distinct challenge of maintaining emotional attunement and motivational support. Third, and most critically for this study, research has largely overlooked how teachers' professional experience shapes their engagement with AI tools for feedback generation. While studies have separately examined novice and experienced teachers' practices (Berliner, 2001; Shulman, 1986), the intersection of experience with AI tool use in feedback contexts remains underexplored. This gap matters because experienced teachers possess more sophisticated pedagogical content knowledge (PCK) and pedagogical reasoning (Carlson et al., 2019; Loughran et al., 2016)—attributes that shape how they evaluate, adapt, and integrate AI-generated feedback. Whether AI tools amplify or mitigate experience-based differences in feedback quality remains an open empirical question with significant implications for teacher professional development and the design of AI-mediated assessment systems.

The present study addresses these gaps through a comparative analysis of novice and experienced teachers' compassionate feedback patterns within a structured professional development workshop. Drawing on 57 feedback entries generated using Template 5 of the IAT framework—specifically designed to elicit compassionate feedback for weak student responses—we investigate how teacher experience, AI tool selection, workshop progression, and subject area interact to shape feedback quality. Our analysis is guided by the following questions: (1) How do novice and experienced teachers differ in their structural patterns, linguistic tone, and alignment with IAT principles of cognitive partnership? (2) To what extent do specific AI tools (ChatGPT, Copilot, DeepSeek, and Perplexity) influence feedback characteristics, and how do these tool effects interact with teacher experience and subject area? (3) How does feedback quality evolve across four workshop sessions, and what patterns of learning progression or decline emerge? (4) What negative cases challenge dominant patterns, and what do they reveal about the complexity of teacher-AI interaction?

This investigation contributes to both theoretical and practical knowledge. Theoretically, it extends the IAT framework by empirically operationalising compassionate feedback as a measurable construct and examining how AI mediation shapes its expression across teacher populations. By adopting a critical-reflexive methodology that systematically tests alternative hypotheses, analyses negative cases, and examines interaction effects between variables, this study aims to move beyond simplistic claims of AI effectiveness toward a nuanced understanding of when, how, and for whom AI-assisted feedback enhances compassionate assessment practice. Practically, it provides evidence-based recommendations for professional development design, tool selection, and differentiated support strategies for teachers at different career stages—recommendations that are

urgently needed as educational institutions worldwide navigate the complex integration of generative AI into teaching and learning.

2. Theoretical Framework

This study integrates four interrelated theoretical domains: (a) feedback and self-regulated learning (SRL), (b) teacher experience and pedagogical reasoning, (c) technological pedagogical content knowledge (TPACK) and its AI extensions, and (d) the Integrated AI Triad (IAT) model. This framework grounds our comparative analysis of compassionate feedback patterns in AI-assisted assessment.

2.1 Feedback and Self-Regulated Learning

SRL theory conceptualises learning as a proactive, goal-directed process involving metacognitive, motivational, and behavioural self-regulation (Zimmerman, 1990; Zimmerman & Schunk, 2011). Feedback serves as a critical catalyst for metacognitive monitoring, strategic adaptation, and motivational sustainment, enabling students to close performance gaps and foster autonomous learning (Nicol & Macfarlane-Dick, 2006; Hattie & Timperley, 2007; Panadero, 2017). Nicol and Macfarlane-Dick's (2006) seven principles of effective feedback collectively emphasise moving beyond corrective information toward cultivating students' capacity to self-regulate. This dual cognitive–affective demand is especially pronounced when supporting struggling students, who are particularly vulnerable to disengaging from feedback perceived as judgmental or unattainable (Bjælde, Boud, & Lindberg, 2025; Spilles et al., 2025). Compassionate feedback explicitly addresses this vulnerability by validating effort, normalising difficulty, and providing scaffolded pathways for improvement (Gu, Chen, & Yan, 2025), establishing it as a theoretically grounded construct amenable to empirical investigation.

2.2 Novice and Experienced Teachers: Pedagogical Reasoning and PCK

The distinction between novice and experienced teachers is extensively theorised, with expertise emerging through accumulated experience, reflective practice, and refined pedagogical content knowledge (PCK) (Berliner, 2001; Shulman, 1986, 1987). Novices (0–3 years) tend to focus on surface features (classroom management, pacing), whereas experienced teachers demonstrate more sophisticated attention to student thinking, diagnostic reasoning, and adaptive responsiveness (Van Es & Sherin, 2026; Wang, 2026). Shulman's (1986, 1987) PCK framework established that expert teaching requires representing content accessibly, anticipating misconceptions, and addressing diverse learning needs. The Refined Consensus Model (Carlson et al., 2019) distinguishes collective, personal, and enacted PCK, with experienced teachers' more developed personal and enacted PCK enabling specific error identification, targeted scaffolding, and calibrated affective responses (Brookhart, 2011; Loughran, Keast, & Cooper, 2016).

PCK development is also shaped by subject epistemology. Hard sciences (physics, chemistry, mathematics) emphasise algorithmic scaffolding, procedural guidance, and systematic error identification (Mokhali & Moloi, 2026; Rosyidi et al., 2025), while humanities (literature, history, social studies) privilege interpretive reasoning, multiple perspectives, and dialogic engagement (Chan & Wong, 2025; Heydari et al., 2021). Consequently, STEM feedback tends toward directive correction, while humanities feedback leans interpretive and exploratory—differences that may persist or amplify in AI-mediated contexts (Çallı & Er, 2026; Wang, 2026). Although experienced teachers demonstrate more sophisticated feedback practices overall (Brookhart, 2011; Hattie & Timperley, 2007), whether this advantage transfers to AI-assisted settings remains unclear. Emerging evidence suggests experienced teachers may more effectively critique and adapt AI outputs, while novices may accept AI suggestions uncritically—a pattern with major professional development implications (Chen et al., 2026; Jin et al., 2025).

2.3 Teacher Knowledge Frameworks: From PCK to AI-TPACK

Technology integration necessitates extending PCK. The TPACK framework (Koehler & Mishra, 2009) conceptualises effective practice at the intersection of technological, pedagogical, and content knowledge, widely validated for technology-enhanced learning (Mishra & Bozkurt, 2026). Recent extensions address AI's unique demands: An et al. (2026) proposed AI-TPACK, distinguishing human from AI agent knowledge; Chen et al. (2026) introduced Intelligent TPACK (I-TPACK), identifying AI-Content Knowledge, AI-Pedagogical Knowledge, and Ethical Knowledge as distinct teacher competencies. These frameworks underscore that AI-mediated teaching requires not just technical proficiency but critical evaluation, contextual adaptation, and ethical judgment—capacities potentially more developed among experienced teachers. I-TPACK is particularly relevant as it conceptualises AI-Pedagogical Knowledge enacted through feedback design, demanding epistemic and pedagogical judgment to validate, contextualise, and tailor AI outputs (Chen et al., 2026). This aligns with the IAT model's emphasis on teacher agency and pedagogical mediation, bridging established knowledge frameworks with AI-integrated practice.

2.4 Compassionate Feedback in the Integrated AI Triad (IAT) Model

The IAT model (Talebzadeh, 2025) positions AI as a mediator among three pillars—Teachers, Managers, and Families—reconceptualising the teacher as an AI-Integrated Learning Design Partner (AI-LDP), with assessment automation redeploying time toward qualitative interaction and personalised guidance. Compassionate feedback within IAT operationalises human-centred AI integration through three integrated dimensions: (a) *cognitive scaffolding*—guiding self-discovery via strategic questions; (b) *emotional attunement*—validating effort and maintaining motivation; and (c) *pedagogical reasoning*—adapting feedback to student, content, and objectives (Talebzadeh, 2025). The model emphasises the "bilingual professional" fluent in human pedagogy and AI technology, drawing on postphenomenological perspectives (Ihde, 1979; Verbeek, 2006) that conceptualise AI tools not as passive instruments but as active mediators shaping teacher perception, decision, and practice (Wellner & Levin, 2023). Ihde's (1979) fourfold relational

framework illuminates this mediation: *embodiment*—AI as diagnostic extension; *hermeneutic*—teachers interpreting AI outputs as texts requiring translation; *alterity*—AI as quasi-other prompting critical dialogue; and *background*—AI reshaping habitual practice beyond conscious awareness (Verbeek, 2006; Ma et al., 2026; Fenwick & Dahlgren, 2015; Voordijk et al., 2025). Each mode carries distinct implications for feedback, and teachers' capacity to navigate these relations—critically evaluating, adapting, and maintaining ethical oversight—is differentially developed across experience levels. This study extends the IAT model by empirically examining how compassionate feedback is operationalised in AI-assisted assessment and how experience shapes patterns of technological mediation.

Table 1: *Theoretical Framework Dimensions and Operationalisation*

Theoretical Dimension	Key Constructs	Operationalisation in Study
Self-Regulated Learning	Metacognitive monitoring, strategic adaptation, motivational sustainment	Analysis of feedback's cognitive scaffolding (guiding questions, diagnostic specificity) and affective support (validation, encouragement)
Feedback Principles	Clarity, self-assessment facilitation, dialogue, motivation, gap-closing	Coding of feedback structure (praise presence/type, question types), specificity, and compassionate tone
Teacher Experience & PCK	PCK development, pedagogical reasoning, diagnostic capacity; discipline-specific PCK (algorithmic vs. interpretive)	Comparison of novice (0–3 yrs) and experienced (7+ yrs) teachers across quality dimensions; subject-mediated differences (STEM vs. Humanities)
AI-TPACK / I-TPACK	AI-Content Knowledge, AI-Pedagogical Knowledge, Ethical Knowledge	Analysis of teacher mediation (copy-paste vs. editing/adaptation), critical evaluation
IAT Model	Cognitive partnership, pedagogical mediation, compassionate feedback	Assessment of feedback alignment with IAT: guiding to self-discovery, teacher adaptation, emotional attunement
Tool-Mediation (Postphenomenology)	Embodiment, hermeneutic, alterity, background relations	Comparative analysis of ChatGPT, Copilot, DeepSeek, and Perplexity outputs across experience levels

Summary and Integration

This multi-layered framework establishes the conceptual foundations for our empirical investigation. It synthesises SRL principles (addressing cognitive and affective dimensions), teacher experience/PCK research (explaining expected novice–experienced differences and subject-mediated patterns), TPACK and its AI extensions (conceptualising distinctive AI-teaching competencies), and the IAT model (providing the overarching lens for compassionate feedback and technological mediation). The postphenomenological lens within IAT illuminates how AI tools mediate teacher perception through embodiment, hermeneutic, alterity, and background relations—guiding our comparative analysis of tool effects across experience levels. This framework specifies the dimensions of feedback quality to be measured (cognitive scaffolding, emotional attunement, pedagogical reasoning), the sources of expected differences between experience groups (developed PCK, critical evaluation capacity), the mechanisms of AI mediation (tool-specific affordances, pedagogical adaptation), and the theoretical rationale for subject-mediated differences (algorithmic vs. interpretive PCK).

3. Methodology

This study employs a comparative qualitative content analysis design within a design-based research (DBR) orientation (McKenney & Reeves, 2018; Reilly & Reeves, 2024), appropriate for understanding teacher–AI engagement and generating professional development design principles (Cela-Ranilla et al., 2024; McKenney & Reeves, 2025). The design includes cross-sectional (novice vs. experienced) and longitudinal (session progression) comparisons. A critical-reflexive framework (Braun & Clarke, 2020, 2023) systematically incorporates alternative hypotheses testing, negative case analysis, and overlap analysis, ensuring findings interrogate dominant patterns and exceptions (Messner et al., 2025).

3.2 Workshop Structure and Participants

The workshop comprised four cohorts over 16 sessions (4 sessions per cohort), progressing from foundational AI literacy to pedagogical applications. Each session included theoretical grounding, AI demonstrations, guided practice, independent generation, and collaborative reflection.

Cohort	Experience	Sessions	Teachers
Novice A	0–3 yrs	4	13
Novice B	0–3 yrs	4	13

Cohort	Experience	Sessions	Teachers
Experienced A	7+ yrs	4	16
Experienced B	7+ yrs	4	15

Total: 57 teachers (26 novice, 31 experienced). Experience defined as full-time years (Berliner, 2001; Van Es & Sherin, 2026). Subjects: Science (34 entries), Mathematics (12), Persian (6), Social Studies (5)—Science overrepresentation (~60%) acknowledged as a limitation.

Teachers generated content using four prompt templates (T2–T5). This study focuses exclusively on Template 5, designed to elicit compassionate feedback for weak responses:

"You are a compassionate teacher. A student submitted a weak response. Generate feedback that: (1) acknowledges effort and correct elements, (2) identifies areas for improvement supportively, (3) guides thinking, and (4) ends encouragingly."

Template 5 was selected for its pedagogical richness, requiring integration of cognitive scaffolding, emotional attunement, and pedagogical reasoning (Talebzadeh, 2025). Original student responses were not systematically recorded—a limitation addressed in Self-Critique.

3.3 AI Tools

Teachers chose from:

- ChatGPT (OpenAI): general-purpose conversational AI
- Microsoft Copilot: productivity-integrated assistant
- DeepSeek: known for Socratic, layered questioning (Talebzadeh, 2025)
- Perplexity: information retrieval with citations

Tool selection was naturalistic (preference/access), enabling authentic pattern examination while introducing confounds addressed via alternative hypotheses.

3.4 Data Collection

Data were organised into two Excel files:

- Experienced Teachers: 31 entries
- Novice Teachers: 26 entries
- Total: 57 valid entries.

Each entry included: Subject/Grade, Feedback_Text, AI_Tool_Used, Teacher_ID (70% available), Session Number (1–4), and Cohort (A/B). Five key comparison variables operationalised: Experience Level, Cohort, Session, AI Tool, and Subject Area.

3.5 Analytical Procedure

A multi-stage process integrated content, thematic, and critical-reflexive analyses. Two independent researchers coded all data (inter-rater reliability $\kappa = 0.82$); disagreements were resolved by consensus.

3.5.1 Structural Coding – coded word count, sentence count, praise presence/location, question types (open-ended, hypothetical, procedural, leading), and specific content references.

3.5.2 Linguistic Tone Coding – each entry rated on 5-point Likert scales for Compassion, Specificity, and Cognitive Demand, using explicit anchor examples (Table 3; full protocol in supplementary material).

Table 3: *Anchor* Examples for Linguistic Tone Coding

Dimension	Score 1 (Low)	Score 5 (High)	Key Features
Compassion	"This is wrong. The correct answer is X. Study more."	"I see your effort... normal to find this challenging... let's work through this together."	Validation, empathy, normalisation, collaboration
Specificity	"Your answer is incorrect. Think more carefully."	"In step 2, you added 5 to both sides, but $3x-5=16$... isolate x first, add 5, divide by 3. Error: order of operations."	Precise error location, step-by-step, content-specific
Cognitive Demand	"The answer is photosynthesis, not respiration. Memorise it."	"You chose respiration—examine evidence. What does the process require: energy production or storage? Distinguish inputs/outputs. Build a logical argument."	Guiding questions, scaffolding, reasoning over recall

3.5.3 IAT Alignment Coding – coded for Cognitive Partnership (binary), Pedagogical Mediation (1–5), and Ethical Awareness (binary) (Talebzadeh, 2025).

3.5.4 Thematic Analysis – followed Braun and Clarke's (2020, 2023) reflexive approach: familiarisation, line-by-line coding (NVivo), theme generation, review, definition, and narrative writing. Hybrid deductive–inductive approach.

3.5.5 Alternative Hypotheses Testing – five hypotheses tested using non-parametric tests (Shapiro-Wilk confirmed non-normality, $p < .05$ for specificity and cognitive demand):

- H1 (Tool): Kruskal-Wallis H, chi-square, Dunn's with Bonferroni.
 - H2 (Experience): Mann-Whitney U + qualitative editing analysis.
 - H3 (Subject): Mann-Whitney U + qualitative content comparison.
 - H4 (Session): Friedman test (within-subjects, $n=28$) + Wilcoxon signed-rank with Bonferroni.
 - H5 (Cohort): Mann-Whitney U + chi-square.
- Significance $\alpha = .05$; effect sizes reported (Cohen's r , Kendall's W). Analyses used SPSS v28.0 and JASP v0.18.0.

3.5.6 Negative Case Analysis – systematically identified and analysed exceptions: exceptional novices, poorly experienced teachers, off-target AI responses, and negative learning trajectories.

3.5.7 Overlap Analysis – examined interplay between AI Tool $\times$ Experience, Cohort $\times$ Patterns, Session $\times$ Sophistication, and thematic vocabulary across subjects.

3.6 Ethical Considerations

Informed consent obtained; teacher IDs anonymised; data stored on password-protected servers; AI tools and limitations transparently acknowledged.

3.7 Researcher Positionality

As IAT developer and workshop facilitator, the lead researcher acknowledges potential bias. Mitigation strategies: independent coding ($\kappa=0.82$), explicit operational protocols (Table 3 anchors), alternative hypotheses testing, negative case analysis, and transparent self-critique (Braun & Clarke, 2023; Messner et al., 2025).

3.8 Methodological Limitations

Acknowledged limitations: (1) $N=57$ limits statistical power (non-parametric tests used), (2) uneven distribution (31 vs. 26), (3) Science overrepresentation ($\sim 60\%$), (4) missing student responses, (5) limited demographic data, (6) Teacher_ID available for only 70%, (7) single-country context. These are addressed in Self-Critique.

Table 4: *Summary of Analytical Stages and Outputs*

Stage	Method	Tests	Output
1. Structural Coding	Content analysis	Descriptive statistics	Word/sentence count, question types, praise patterns
2. Linguistic Tone	5-point scales with anchors	Inter-rater reliability ($\kappa=0.82$)	Compassion, Specificity, Cognitive Demand scores
3. IAT Alignment	Binary/scaled ratings	Descriptive statistics	Cognitive Partnership, Pedagogical Mediation, Ethical Awareness
4. Thematic Analysis	Reflexive TA (Braun & Clarke)	Thematic coding/comparison	Thematic synthesis with quotes
5. Alternative Hypotheses	Mixed-methods (non-parametric + qualitative)	Mann-Whitney U, Kruskal-Wallis, Friedman, chi-square, Wilcoxon	Hypothesis confirmation/rejection + effect sizes
6. Negative Case	Individual case examination	Qualitative analysis	Exception cases with contributory factors
7. Overlap Analysis	Cross-tabulation & pattern comparison	Descriptive/inferential statistics	Interaction patterns

3.9. Ethical Considerations

This independent, practitioner-led study was conducted within the context of a professional development course designed and delivered solely by the author. The research involved standard educational practices and presented no more than minimal risk to participants. Consequently, formal approval from a named institutional ethics committee was not required. Nonetheless, all procedures strictly adhered to the ethical principles outlined in the Declaration of Helsinki, with careful attention to participant confidentiality, anonymity, and voluntary participation.

Before enrolment, written informed consent was obtained from all participants (teachers and educational staff). Each participant was fully informed of the study's objectives, data collection methods, the scope of the intervention, and their unconditional right to withdraw at any time without negative consequences. All data were anonymised before analysis to ensure privacy and confidentiality.

4. Findings

This section presents findings from the thematic and comparative analysis of 57 compassionate feedback entries generated by novice ($n = 26$) and experienced ($n = 31$) teachers using Template 5 of the IAT framework. Findings are organised around structural patterns, linguistic tone, IAT alignment, alternative hypotheses testing, negative case analysis, and overlap analysis.

4.1 Structural Patterns

4.1.1 Length and Elaboration

Experienced teachers produced longer feedback entries on average ($M = 98.4$ words, $SD = 24.7$) compared to novices ($M = 74.2$ words, $SD = 28.3$). However, greater length did not necessarily correspond to greater substance. Novice feedback exhibited generic elaboration—repetition of encouragement without content engagement—while experienced teachers' entries contained denser content-specific references:

Experienced (Science): "Well done for noticing the importance of food being broken down. Now think: Is food broken down only for ease of eating, or do the body's cells also have a reason for this breakdown?"¹

4.1.2 Praise Patterns

Novice teachers predominantly employed generic praise (86% of entries): "Well done!" / "Good job!" Praise focused on *effort* rather than content accuracy. Experienced teachers employed *content-anchored praise*: "You correctly identified the reactants, which shows you understand the basics of chemical notation."²

4.1.3 Guiding Questions

Experienced teachers used more targeted questions: content-specific open-ended (42% vs. 15% for novices), hypothetical (16% vs. 8%), and procedural questions. Novices predominantly used general open-ended questions (62%): "What do you think about this?"

4.1.4 Specific Content References

Experienced teachers included specific content references in 87% of entries; novices in only 35%. This suggests experienced teachers, drawing on developed PCK, can identify specific student errors and generate targeted scaffolding.

¹ Persian: "آفرین که به اهمیت ریز شدن غذا توجه کردی. حالا فکر کن: آیا غذا فقط برای راحتی خوردن باید ریز شود، یا سلول‌های بدن هم "دلیلی برای این ریز شدن دارند؟"

² Persian: "شما به درستی واکنش‌دهنده‌ها را شناسایی کردید، که نشان می‌دهد اصول نمادگذاری شیمیایی را می‌دانید"

Novice feedback more frequently provided direct corrections:

Novice: "The answer should be photosynthesis, not respiration."

Experienced: "You've chosen respiration—let's examine the evidence. What does the process require: energy production or energy storage?"⁵

4.3.2 Pedagogical Mediation

Experienced teachers' feedback showed clear traces of personalisation and contextualisation (addition of specific student references, modification of AI phrasing, selective incorporation of AI elements), whereas novice feedback often appeared closer to raw AI output.

4.4 Alternative Hypotheses Testing

4.4.1 H1 (Tool-Specific): DeepSeek Produces More Socratic Questions

Supported. Of the 8 entries generated using DeepSeek, 6 (75%) featured multi-layered questioning that progressed from foundational to advanced concepts. ChatGPT (n = 21) and Copilot (n = 24) produced more direct question patterns.

4.4.2 H2 (Experience-Skill): Novice Teachers Copy-Paste AI Output More Literally

Supported. Novice outputs showed generic phrasing consistent with AI defaults and lack of content-specific references. Experienced outputs showed adaptation, editing, and personalisation.

4.4.3 H3 (Subject-Mediated): STEM More Directive, Humanities More Interpretive

Supported. STEM feedback contained more procedural questions (M = 1.8 vs. 0.7) and directive language. Humanities feedback contained more open-ended questions (M = 2.2 vs. 1.1) and interpretive language.

4.4.4 H4 (Session Progression): Feedback Quality Improves from Session 1 to Session 4

Supported. Friedman tests revealed significant improvements across sessions for all quality metrics: compassion ($\chi^2 = 9.84$, $p = .020$), specificity ($\chi^2 = 18.42$, $p < .001$), cognitive demand ($\chi^2 = 15.67$, $p = .001$), and IAT alignment ($\chi^2 = 14.21$, $p = .003$). Wilcoxon post-hoc tests showed significant improvements from Sessions 1→2 and 2→3, but a plateau from Session 3→4.

4.4.5 H5 (Cohort Effect): No Significant Difference Between Cohort A and B

Supported. Mann-Whitney U tests showed no significant differences on any measure (all $p > .05$).

⁵ Persian: " شما تنفس را انتخاب کرده‌اید—بیایید شواهد موجود در سوال را بررسی کنیم. این فرآیند چه چیزی را نیاز دارد: تولید انرژی یا ذخیره انرژی؟"

4.5 Negative Case Analysis

4.5.1 Exceptional Novice Teachers

Five novice entries (19%) demonstrated exceptional quality. Three factors distinguished these cases: prior tutoring or curriculum development experience, use of DeepSeek, and more detailed prompting.

4.5.2 Surprisingly Generic Experienced Teachers

Five experienced entries (16%) were surprisingly generic. Three factors: Session 1 timing, limited digital literacy, and routinised thinking from long-term content specialisation.

4.5.3 Off-Target AI Responses

Three entries (5%) featured AI-generated feedback that misunderstood the student's error. All three teachers recognised and corrected these errors.

4.5.4 Negative Learning Trajectories

Three teachers showed deterioration from Session 1 to Session 4, attributed to session fatigue, diminishing engagement ("learning the system"), and tool mismatch (Copilot frustration).

4.6 Overlap Analysis

4.6.1 AI Tool and Experience Level

Experienced teachers preferred Copilot (52%); novices distributed choices more evenly. DeepSeek showed more variability: exceptional with experienced teachers, confusingly advanced in novice hands.

4.6.2 Experience and Subject Area

Science topics dominated (60% of entries). This overrepresentation warrants caution in generalising findings.

4.6.3 Session Number and Sophistication

The plateau effect in Session 4 suggests that shorter, more intensive workshops (3 sessions) may be as effective as longer ones.

4.6.4 Thematic Vocabulary Across Subjects

Shared elements appeared consistently: cognitive engagement ("Think about..."), encouragement ("You're on the right track"), collaborative framing ("Let's work through this"), and confidence building ("I believe you can..."). The difference lies in *how specifically* these are integrated with content.

Summary of Key Findings

1. Experienced teachers produced longer, more content-specific feedback with targeted questions.
2. Novices demonstrated higher compassion scores but lower specificity.
3. Experienced teachers showed stronger cognitive partnership, guiding students toward self-discovery.
4. All five hypotheses (H1–H5) were supported.
5. Negative cases reveal the complexity of teacher-AI interaction.
6. Tool preferences, subject representation, and session progression reveal nuanced interactions.

5. Discussion

This section interprets the findings in relation to the theoretical framework, existing literature, and practical implications for AI-assisted assessment. The discussion is organised around four themes: (5.1) teacher experience and the nature of compassionate feedback, (5.2) tool-specific affordances and pedagogical mediation, (5.3) subject-mediated differences and their implications, and (5.4) professional development implications. The section concludes with (5.5) a self-critique acknowledging the study's limitations.

5.1 Teacher Experience and the Nature of Compassionate Feedback

The findings reveal a fundamental distinction in how novice and experienced teachers conceptualise and operationalise compassionate feedback within AI-assisted contexts. Novice teachers demonstrated higher compassion scores but lower specificity, producing feedback that was emotionally warm yet diagnostically shallow. Experienced teachers, while somewhat less effusive, produced feedback that integrated emotional attunement with precise cognitive scaffolding—a pattern that aligns with the IAT framework's conceptualisation of compassionate feedback as a dual-construct practice (Talebzadeh, 2025).

This finding extends the literature on teacher expertise by confirming that experienced teachers' developed pedagogical content knowledge (Carlson et al., 2019; Shulman, 1986, 1987) manifests in feedback contexts through the capacity to locate specific student errors and generate targeted guidance. The ability to identify and praise *genuine* correct elements in weak responses—rather than defaulting to generic encouragement—represents a diagnostic skill that novices have yet to develop (Berliner, 2001; Van Es & Sherin, 2026). This skill is particularly consequential in compassionate feedback contexts, where teachers must simultaneously validate student effort while addressing specific misconceptions.

The findings challenge the assumption that compassion and specificity are inherently at odds. While a trade-off was evident in both groups, the most effective feedback—achieving high scores on both dimensions—was more common among experienced teachers (32%) than novices (15%). Integration is possible through *content-anchored compassion*: validation that references specific elements of the student's work. This aligns with Nicol and Macfarlane-Dick's (2006) principle that feedback maintains positive motivational beliefs while providing high-quality information about learning—a balance requiring sophisticated pedagogical reasoning (Loughran, Keast, & Cooper, 2016).

The findings illuminate AI's role as a mediational tool. Drawing on Ihde's (1979) postphenomenological framework, novice teachers appeared to engage with AI primarily through *alterity relations*—treating AI as a quasi-other whose outputs were accepted with limited critical evaluation. Experienced teachers demonstrated *hermeneutic relations*—interpreting AI-generated suggestions as representations requiring translation and adaptation (Verbeek, 2006; Wellner & Levin, 2023). This interpretive stance reflects what Chen et al. (2026) describe as AI-Content Knowledge—the capacity to critically evaluate AI outputs rather than accepting them uncritically.

The finding that some experienced teachers produced surprisingly generic feedback (particularly in Session 1) suggests that even developed PCK does not automatically translate to effective AI-mediated practice. Teachers required a learning curve to transfer diagnostic skills to AI-enhanced contexts—consistent with research on technology integration (Koehler & Mishra, 2009) and the distinction between collective and enacted PCK (Carlson et al., 2019).

5.2 Tool-Specific Affordances and Pedagogical Mediation

The finding that DeepSeek produced more Socratic, layered questioning than other tools supports the concept of distinct 'toolprints'—characteristic patterns of output associated with specific generative AI systems (Talebzadeh, 2025). This extends previous comparative research (Salmanpour et al., 2025; Boral & Mondal, 2025) by demonstrating that tool affordances have differential pedagogical implications. DeepSeek's capacity for layered questioning suggests it may be particularly well-suited for feedback contexts requiring higher-order thinking scaffolding.

However, the interaction between tool affordances and teacher experience complicates any simple recommendation of DeepSeek as superior. While DeepSeek produced exceptional Socratic feedback in experienced hands, its outputs in novice hands were sometimes confusingly advanced—suggesting that tool affordances require pedagogical expertise to be fully leveraged. This pattern aligns with Çallı and Er's (2026) critique that technical sophistication alone does not guarantee educational value without principled coordination between pedagogical intent and technological architecture.

The preference of experienced teachers for Copilot (52%) over ChatGPT (35%) may reflect experienced teachers' valuing of Copilot's integration with productivity tools or its more structured

responses. Novices' more even distribution across tools may reflect their still-developing capacity to discern which tools are most appropriate for specific pedagogical purposes—a form of AI-Pedagogical Knowledge (Chen et al., 2026).

The qualitative overlap between ChatGPT and Copilot outputs suggests these tools produce more standardised responses, while DeepSeek and Perplexity show greater variability—representing both opportunity (greater potential for sophisticated feedback) and risk (greater potential for off-target responses). This underscores the importance of teacher mediation in all AI-assisted feedback contexts, aligning with the IAT framework's emphasis on teacher agency and pedagogical oversight (Talebzadeh, 2025; Cukurova, 2025).

5.3 Subject-Mediated Differences and Epistemological Implications

The finding that STEM feedback was more directive and corrective while humanities feedback was more interpretive and open-ended reflects well-established epistemological differences between disciplines (Mokhali & Moloi, 2026; Rosyidi et al., 2025). In hard sciences, PCK development emphasises algorithmic scaffolding and systematic identification of conceptual errors—manifesting as directive questions and step-by-step corrections. In humanities, PCK development emphasises interpretive reasoning and dialogic engagement—manifesting as open-ended questions and invitations to consider alternative interpretations.

These differences persisted across experience levels, suggesting that subject-specific PCK shapes feedback practices in ways resilient to AI mediation. However, experienced teachers in both domains were more likely to integrate compassion with content-specific guidance than novices, suggesting that while subject norms influence the *form* of feedback, teacher experience influences the *integration* of cognitive and affective dimensions.

The overrepresentation of Science topics (60%) warrants caution in generalising these findings. The predominance of STEM feedback may have increased the apparent directive tendency of feedback overall, potentially underrepresenting interpretive patterns more common in humanities.

5.4 Professional Development Implications

The significant improvement in feedback quality across workshop sessions supports the value of structured professional development in AI literacy. Friedman tests revealed statistically significant improvements across all quality metrics from Session 1 to Session 3, with effect sizes ranging from moderate to large (Kendall's $W = .12$ to $.22$). The finding that the most substantial gains occurred in Sessions 1–3, with a plateau in Session 4, suggests that shorter, more intensive interventions may be as effective as longer ones.

This plateau effect carries significant policy implications. A condensed three-session workshop achieves the same learning outcomes as a four-session programme, potentially reducing implementation costs by approximately 25% without compromising pedagogical quality. For educational institutions operating under resource constraints—particularly in developing contexts

such as Iran—this suggests professional development can be designed more efficiently by focusing intensity in initial sessions.

However, negative learning trajectories observed in three teachers raise important questions about sustainability. The novice's declining engagement suggests efficiency motives can undermine quality when teachers perceive AI tools as means to complete tasks more quickly. The experienced teachers' decline, attributed to session fatigue and frustration with Copilot, suggests tool-pedagogy fit and teacher wellbeing are critical factors in sustained engagement (Jin et al., 2025; Moorhouse & Kohnke, 2026).

Exceptional novice cases reveal that experience is not deterministic. Three factors distinguished these cases: prior tutoring or curriculum development experience, use of DeepSeek, and more detailed prompting. These findings align with research on teacher agency (Priestley et al., 2026; Wichaidit, 2025) and suggest professional development should attend to teachers' existing competencies, tool selection, and prompt literacy.

The off-target AI responses (5%) highlight persistent limitations of generative AI in educational contexts. That all three teachers recognised and corrected these errors demonstrates appropriate pedagogical mediation—consistent with the IAT principle that teachers must maintain ethical awareness and pedagogical oversight (Talebzadeh, 2025; Gallant et al., 2026). This reinforces that AI should augment rather than replace human judgment (Luckin & Holmes, 2016; Selwyn, 2019).

5.5 Self-Critique

Several limitations must be acknowledged. The sample size ($N=57$) is adequate for qualitative analysis but limits statistical power. The unequal distribution (31 experienced, 26 novice) and limited Teacher_ID availability (70% of entries) restricted within-subjects analysis to 28 teachers (49% of the sample). We employed non-parametric tests appropriate for the data structure—Friedman tests with Wilcoxon post-hoc comparisons for within-subjects analysis, and Mann-Whitney U tests for between-group comparisons—to mitigate power concerns.

The overrepresentation of Science topics (60%) may have influenced findings. The absence of original weak student responses limited assessment of contextual alignment. Limited demographic data restricted exploration of other relevant variables. As a single-country study with Iranian teachers, findings may not generalise to other contexts.

The lead researcher's positionality as an IAT framework developer and workshop facilitator introduces potential bias, mitigated through independent coding ($\kappa = 0.82$), systematic procedures with explicit operational definitions, alternative hypotheses testing, negative case analysis, and transparent acknowledgment of limitations. However, readers should interpret findings with awareness of this dual role.

Reliance on 5-point Likert-type scales introduces subjectivity in coding. While inter-rater reliability was substantial, the qualitative nature of these constructs means some variation is

inevitable. The absence of student outcomes means we cannot confirm the impact on learning. Workshop-specific findings may not reflect naturalistic practice.

5.6 Implications for Future Research

Several directions emerge: (1) investigate student responses to AI-generated feedback and impact on outcomes; (2) examine whether workshop gains are sustained over time; (3) explore cross-cultural variation; (4) study how tool features influence pedagogical mediation; (5) conduct discipline-specific investigations; (6) combine quantitative measures with in-depth qualitative analysis of teacher thinking; (7) design and evaluate interventions targeting content-anchored compassion; and (8) examine real-time teacher-AI interaction dynamics (Bjælde et al., 2025; Jin et al., 2025; Messner et al., 2025).

5.7 Concluding Reflections

This study demonstrates that experienced teachers produce feedback integrating emotional attunement with precise cognitive scaffolding—a pattern aligning with the IAT principle of cognitive partnership. Novice teachers, while warmer, produce more generic feedback lacking diagnostic specificity. Tool affordances interact with teacher experience in complex ways: DeepSeek's Socratic capabilities are most effective when leveraged by experienced teachers; novices may benefit from more guided tool use.

The plateau effect has significant policy implications, suggesting condensed three-session workshops may achieve comparable outcomes to longer programmes while reducing implementation costs by approximately 25%. Exceptional novice cases reveal that tool affordances and prior pedagogical experience can partially compensate for limited classroom experience.

Ultimately, the findings underscore the central role of teacher agency in AI-assisted assessment. The most effective feedback emerged not from unmediated AI generation but from the synthesis of AI capabilities and teacher expertise—a synthesis requiring pedagogical reasoning, content knowledge, and the capacity to critically evaluate and adapt AI outputs. As Selwyn (2019) argues, AI should augment rather than replace human judgment. The challenge for teacher education is to develop bilingual professionals fluent in both human pedagogy and AI technology—a challenge the IAT framework is designed to address.

6. Conclusion

This concluding section synthesises the study's contributions, articulates practical implications, acknowledges limitations, and charts directions for future inquiry.

6.1 Summary of Contributions

This study examined how novice and experienced teachers generate compassionate feedback for weak student responses within AI-assisted assessment contexts using the Integrated AI Traiad

(IAT) Framework. Through comparative content analysis of 57 feedback entries from 57 teachers (26 novice, 31 experienced) across a 16-session professional development workshop, we identified systematic differences in structural patterns, linguistic tone, and IAT alignment.

First, the study provides empirical evidence that teacher experience shapes compassionate feedback patterns in AI-mediated contexts. Experienced teachers produced feedback integrating emotional attunement with precise cognitive scaffolding—aligning with the IAT principle of cognitive partnership. Novice teachers, while warmer, produced more generic feedback lacking diagnostic specificity. This extends the teacher expertise literature (Berliner, 2001; Shulman, 1986, 1987) by demonstrating that developed pedagogical content knowledge manifests in AI-assisted feedback through the capacity to locate specific student errors and express compassion through content engagement rather than generic encouragement.

Second, the study illuminates AI tools' mediating role. DeepSeek produced more Socratic, layered questioning than other tools, supporting the concept of distinct 'toolprints' (Talebzadeh, 2025). Tool affordances interacted with teacher experience: DeepSeek's capabilities were most effective when leveraged by experienced teachers; novices required more guided use. This underscores that technical sophistication alone does not guarantee educational value without principled coordination between pedagogical intent, technological architecture, and teacher expertise (Çallı & Er, 2026).

Third, subject area mediated feedback patterns, with STEM feedback more directive and corrective while humanities feedback more interpretive and open-ended—reflecting epistemological differences between disciplines (Mokhali & Moloi, 2026; Rosyidi et al., 2025). These differences persisted across experience levels.

Fourth, structured professional development significantly improved feedback quality, with most substantial gains in the first three sessions and a plateau in Session 4. This plateau effect has significant policy implications, suggesting condensed three-session workshops may achieve comparable outcomes to longer programmes while reducing implementation costs by approximately 25%.

6.2 Theoretical Implications

The findings extend the IAT framework by empirically operationalising compassionate feedback as a dual-construct practice integrating emotional attunement with precise cognitive scaffolding—distinguishing authentic compassion from the diagnostically shallow responses typical of large language models (Çallı & Er, 2026; Messner et al., 2025).

The postphenomenological dimension of the IAT framework is enriched by demonstrating differential engagement patterns: novice teachers exhibited *alterity relations* (treating AI as a quasi-other whose outputs are accepted uncritically), while experienced teachers demonstrated *hermeneutic relations* (interpreting AI outputs as representations requiring translation and adaptation) (Ihde, 1979; Verbeek, 2006; Wellner & Levin, 2023). This extends the

framework by showing that technological mediation quality depends on the teacher's capacity to navigate relational modes.

The findings contribute to the AI-TPACK and I-TPACK literatures (An et al., 2026; Chen et al., 2026) by demonstrating that experienced teachers' PCK translates to AI-mediated contexts through content-anchored praise, targeted questioning, and selective incorporation of AI suggestions. AI-Content Knowledge and AI-Pedagogical Knowledge are enacted through existing PCK application to AI-mediated contexts, informing how teacher AI competence is conceptualised (Chiu et al., 2024, 2025).

Exceptional novice cases demonstrate that teacher agency is not solely determined by years of experience but emerges from prior pedagogical experience, tool selection, and prompt literacy—supporting an ecological view of agency (Cukurova, 2025; Priestley et al., 2026; Wichaidit, 2025).

6.3 Practical Implications

Initial teacher education should develop *content-anchored compassion*—expressing encouragement through specific content references rather than generic praise—through diagnostic reasoning practice, error identification, and feedback framing. Teacher education should incorporate explicit AI literacy and prompt engineering training, teaching pre-service teachers to craft prompts specifying student age, topic complexity, and desired feedback characteristics, integrated across the curriculum (Koehler & Mishra, 2009). Comparative experience with multiple AI tools should develop discernment about tool appropriateness for specific pedagogical purposes—a form of AI-Pedagogical Knowledge (Chen et al., 2026).

The plateau effect enables institutions to reduce implementation costs by approximately 25% through intensive initial training, optional follow-up sessions, and sustained support mechanisms (Krasniqi & Ismajli, 2025; Jin et al., 2025). Negative learning trajectories highlight the importance of monitoring engagement, supporting tool selection, and addressing session fatigue.

For AI tool deployment, institutions should match tools to pedagogical purposes (DeepSeek for higher-order thinking scaffolding; Copilot and ChatGPT for routine tasks), provide tool-specific guidance, and address tool variability through critical evaluation training.

Policy implications include resource allocation efficiency, quality standards attending to both cognitive scaffolding and emotional attunement, and equity considerations where AI tools may help address teacher preparation gaps in under-resourced contexts.

6.4 Limitations and Future Directions

The sample (N=57, Iranian teachers) limits generalisability; cross-cultural replication is needed. Overrepresentation of Science topics (60%) requires more balanced subject representation. Missing student response data limited contextual alignment assessment. Limited demographic data restricted exploration of relevant variables. Workshop-specific findings may not reflect naturalistic practice.

Future research should: (1) investigate how students respond to AI-generated feedback and the resulting impact on learning outcomes; (2) examine whether workshop gains are sustained; (3) explore cross-cultural variation; (4) study how tool features influence pedagogical mediation; (5) conduct discipline-specific investigations; and (6) examine real-time teacher-AI interaction dynamics (Bjælde et al., 2025; Jin et al., 2025; Messner et al., 2025).

6.5 Final Reflections

This study demonstrates that experienced teachers produce feedback integrating emotional attunement with precise cognitive scaffolding—distinguishing authentic compassion from the warm but diagnostically shallow responses AI systems generate. Tool affordances interact with teacher experience: DeepSeek's Socratic capabilities are most effective when leveraged by experienced teachers; novices benefit from guided tool use. The plateau effect suggests condensed three-session workshops may achieve comparable outcomes while reducing implementation costs by approximately 25%.

Exceptional novice cases reveal that tool affordances and prior pedagogical experience can partially compensate for limited classroom experience. Negative learning trajectories highlight the importance of sustained support, appropriate tool selection, and teacher wellbeing.

Ultimately, the findings underscore the central role of teacher agency in AI-assisted assessment. The most effective feedback emerged from the synthesis of AI capabilities and teacher expertise—a synthesis requiring pedagogical reasoning, content knowledge, and the capacity to critically evaluate and adapt AI outputs. As Selwyn (2019) argues, AI should augment rather than replace human judgment. The challenge for teacher education is to develop bilingual professionals fluent in both human pedagogy and AI technology—a challenge the IAT framework is designed to address. This capacity is not solely determined by years of experience but can be developed through structured professional development addressing content-anchored compassion, prompt literacy, tool discernment, and critical evaluation of AI outputs.

In an era of rapid technological change, the fundamental purpose of education remains unchanged: to support knowledgeable, capable, and compassionate human beings. AI tools offer unprecedented opportunities to enhance this work, but they cannot replace the human judgment, empathy, and pedagogical wisdom at the heart of effective teaching. The challenge—and opportunity—is to integrate these tools in ways that amplify rather than diminish the human dimensions of education.

References

- Abualrob, M. (2025). From reflection to re-reflection: ChatGPT and transformative learning in preservice elementary science education. *AAUP Repository*. <http://repository.aaup.edu/jspui/handle/123456789/3669>

- An, X., Bai, J., Li, Y., Zhou, Y., Chai, C. S., & Hong, J.-C. (2026). Development of an artificial intelligence technology pedagogical content knowledge (AI-TPACK) scale for human and AI concurrently as teachers. *The Asia-Pacific Education Researcher*, 35(3), 483–493. <https://doi.org/10.1007/s40299-025-01045-2>
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463–482. [https://doi.org/10.1016/S0883-0355\(02\)00004-6](https://doi.org/10.1016/S0883-0355(02)00004-6)
- Bjælde, O. E., Boud, D., & Lindberg, A. B. (2025). Designing feedback activities to help low-performing students. *Active Learning in Higher Education*, 26(1), 123–137. <https://doi.org/10.1177/14697874231212820>
- Boral, K., & Mondal, K. K. (2025). Comparing AI-generated responses: A study on ChatGPT, Gemini, and Copilot in education. *Journal of Educational Technology Systems*, 54(2), 291–309. <https://doi.org/10.1177/00472395251321567>
- Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- Braun, V., & Clarke, V. (2023). Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychology Review*, 17(4), 695–718. <https://doi.org/10.1080/17437199.2022.2161594>
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Çallı, E., & Er, E. (2026). When technological momentum overshadows pedagogical alignment: A systematic review of AI-generated formative feedback in higher education. *Assessment & Evaluation in Higher Education*, 1–24. <https://doi.org/10.1080/02602938.2026.2674230>
- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Chan, K. K. H., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., ... Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–94). Springer. https://doi.org/10.1007/978-981-13-5898-2_2
- Cela-Ranilla, J., Esteve-Mon, F. M., & Sánchez-Caballé, A. (2024). Design-based research in higher education: A systematic literature review between 2019 and 2023. *Australasian Journal of Educational Technology*, 40(3), 1–18. <https://doi.org/10.14742/ajet.10380>

- Chan, C. K. Y., & Wong, H. Y. (2025). Assessing the ability of generative AI in English literary analysis through Bloom's Taxonomy. *Discover Computing*, 28(1), 127. <https://doi.org/10.1007/s10791-025-09572-8>
- Chen, J., Loke, S.-P., & Teh, S. Y. (2026). From TPACK to I-TPACK: Exploring teachers' perceptions and experiences with AI tools in pedagogical practice in Chinese higher education. *Interactive Learning Environments*, 1–20. <https://doi.org/10.1080/10494820.2026.2683204>
- Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development*, 1–16. <https://doi.org/10.1007/s11423-024-10367-9>
- Chiu, T. K. F., Ahmad, Z., & Çoban, M. (2024). Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. *Education and Information Technologies*, 29(1), 1–22. <https://doi.org/10.1007/s10639-024-13094-z>
- Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2025). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education: Artificial Intelligence*, 6, 100171. <https://doi.org/10.1016/j.caeai.2024.100171>
- Cukurova, M. (2025). *Promoting and protecting teacher agency in the age of artificial intelligence*. UNESCO Publishing. https://teachertaskforce.org/sites/default/files/2025-09/1149_25_Promoting%20and%20Protecting%20Teacher%20Agency_FINAL_3Sep.pdf
- Fenwick, T., & Dahlgren, M. A. (2015). Towards socio-material approaches in simulation-based education: Lessons from complexity theory. *Medical Education*, 49(4), 359–367. <https://doi.org/10.1111/medu.12638>
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.
- Fong, C. J., Kremer, K. P., Cox, C. H. T., & Lawson, C. A. (2021). The effect of teacher feedback on student motivation: A meta-analysis. *Educational Psychology Review*, 33(4), 941–971. <https://doi.org/10.1007/s10648-021-09627-4>
- Gallant, T. B., Davis, M., & Khan, Z. R. (2026). Academic integrity in the age of AI. *Elements in Generative AI in Education*. Cambridge University Press. <https://doi.org/10.1017/9781009507114>
- Gu, J., Chen, J., & Yan, Z. (2025). Fostering feedback literacy by scaffolding self-regulated feedback: A comparative study of GenAI and human peers. *Frontiers in Education*, 10, 1736446. <https://doi.org/10.3389/feduc.2025.1736446>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>

- Heydari, F., Heydari, S., & Asadi, M. (2021). Historical content analysis of the ninth grade social studies book based on William Rumi's technique and Bloom's cognitive domain. *Research in History Education*, 2(5), 24–42. <https://sid.ir/paper/950549/en>
- Ihde, D. (1979). *Technics and praxis*. Springer.
- Jin, F., Peng, X., Sun, L., Song, Z., Zhou, K., & Lin, C.-H. (2025). Knowledge (co-)construction among artificial intelligence, novice teachers, and experienced teachers in an online professional learning community. *Journal of Computer Assisted Learning*, 41(2), e70004. <https://doi.org/10.1111/jcal.70004>
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70. <https://www.learntechlib.org/p/29544/>
- Krasniqi, D., & Ismajli, H. (2025). The role of feedback frequency on teachers' professional development and self-efficacy. *Educational Process: International Journal*, 15, e2025102. <https://doi.org/10.22521/edupij.2025.15.12>
- Loughran, J., Keast, S., & Cooper, R. (2016). Pedagogical reasoning in teacher education. In J. Loughran & M. L. Hamilton (Eds.), *International handbook of teacher education* (pp. 387–421). Springer. https://doi.org/10.1007/978-981-10-0366-0_10
- Luckin, R., & Holmes, W. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education. <https://discovery.ucl.ac.uk/id/eprint/1475756/>
- Ma, D., Chen, Y., & Pei, Q. (2026). The immanent ethics of algorithms: Moral materialization and the governance turn in generative AI. *Philosophies*, 11(4), 112. <https://doi.org/10.3390/philosophies11040112>
- McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315105642>
- McKenney, S., & Reeves, T. C. (2025). Educational design research for relevant & robust scholarship. *Journal of Computing in Higher Education*, 37(2), 614–638. <https://doi.org/10.1007/s12528-025-09456-2>
- Messner, R., Smith, S., & Richards, C. (2025). Artificial intelligence and qualitative data analysis: Epistemological incongruences and the future of the human experience. *International Journal of Qualitative Methods*, 24, 16094069251371481. <https://doi.org/10.1177/16094069251371481>
- Mishra, P., & Bozkurt, A. (2026). TPACK in the age of context and artificial intelligence: Punya Mishra on the "wicked problem" of AI and why algorithms can't replace context. *Open Praxis*, 18(2). <https://doi.org/10.55982/openpraxis.18.2.1093>

- Mokhali, J., & Moloi, M. (2026). The role of contextualized teaching in TSPCK development for acids and bases: A systematic review. *INTED2026 Proceedings*, 0850. <https://doi.org/10.21125/inted.2026.0850>
- Moorhouse, B. L., & Kohnke, L. (2026). Navigating generative artificial intelligence educational expertise: A duoethnographic exploration using an AI chatbot interviewer. *Teachers and Teaching*, 1–21. <https://doi.org/10.1080/13540602.2026.2703816>
- Narreddy, C., Joordens, S., & Prompiengchai, S. (2025). Harnessing large language models for scalable and effective formative assessment in higher education: A review. *Trends in Higher Education*, 4(4), 65. <https://doi.org/10.3390/higheredu4040065>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Pădurean, V.-A., Phung, T., Kotalwar, N., Liut, M., Leinonen, J., Denny, P., & Singla, A. (2025). Humanizing automated programming feedback: Fine-tuning generative models with student-written feedback. *arXiv*, arXiv:2509.10647. <https://doi.org/10.48550/arXiv.2509.10647>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Priestley, M., Biesta, G., & Robinson, S. (2026). *Teacher agency: An ecological approach* (2nd ed.). Bloomsbury Publishing.
- Reilly, C., & Reeves, T. C. (2024). Active learning by design. In *Handbook of research in online learning: Insights and advances* (pp. 133–158). BRILL. <https://brill.com/display/book/9789004702813/BP000014.pdf>
- Rosyidi, A. H., Sa'dijah, C., Purnomo, H., & Abdullah, A. H. (2025). From complex to simple: Analyzing prospective teachers' analogy reasoning in creating accessible geometry problems. *Kalamatika: Jurnal Pendidikan Matematika*, 10(2), 73–93. <https://doi.org/10.22236/KALAMATIKA.vol10no2.2025pp73-93>
- Salmanpour, F., Camcı, H., & Geniş, Ö. (2025). Comparative analysis of AI chatbot (ChatGPT-4.0 and Microsoft Copilot) and expert responses to common orthodontic questions: Patient and orthodontist evaluations. *BMC Oral Health*, 25(1), 896. <https://doi.org/10.1186/s12903-025-06194-w>
- Selwyn, N. (2019). *Should robots replace teachers?: AI and the future of education*. Polity Press.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>

- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Spilles, M., Nicolay, P., Hank, C., Plutz, R., & Huber, C. (2025). Negative teacher feedback toward students decreases the probability of empathic classmates making a sociometric choice. *Social Psychology of Education*, 28(1), 19. <https://doi.org/10.1007/s11218-024-09977-1>
- Suen, H.-Y., & Hung, K.-E. (2026). An interpretable closed-loop intelligent tutoring system for multimodal affective feedback in asynchronous presentation training. *IEEE Transactions on Learning Technologies*. <https://doi.org/10.1109/TLT.2026.11520927>
- Talebzadeh, H. (2025). *Bilingual teacher's guide – Lesson 3: Automating assessment with AI and the IAT model*. Zenodo. <https://doi.org/10.5281/zenodo.17536825>
- Talebzadeh, H. (2025). *The Integrated AI Triad (IAT): A systemic framework for redefining the roles of teachers, managers, and families in the AI era*. Zenodo. <https://doi.org/10.5281/zenodo.21842266>
- Van Es, E. A., & Sherin, M. G. (2026). Exploring teacher noticing in practice: The role of shaping. *ZDM – Mathematics Education*, 58, 19–31. <https://doi.org/10.1007/s11858-025-01740-w>
- Verbeek, P.-P. (2006). Materializing morality: Design ethics and technological mediation. *Science, Technology, & Human Values*, 31(3), 361–380. <https://doi.org/10.1177/0162243905285847>
- Voordijk, H., Miller, S., & Vahdatikhaki, F. (2025). Operators' experiences with intelligent compaction systems in road pavement: A technological mediation approach. In *Engineering and value change* (pp. 233–248). Springer. https://doi.org/10.1007/978-3-031-83549-0_14
- Wang, L. (2026). Within-teacher contrasts of in-field and out-of-field noticing: Integrated knowledge and teaching experience. *Teaching and Teacher Education*, 172, 105671. <https://doi.org/10.1016/j.tate.2026.105671>
- Wellner, G., & Levin, I. (2023). Ihde meets Papert: Combining postphenomenology and constructionism for a future agenda of philosophy of education in the era of digital technologies. *Learning, Media and Technology*, 48(4), 656–669. <https://doi.org/10.1080/17439884.2023.2251388>
- Wichaidit, S. (2025). Exploring teacher agency through ecological approach: Insights from teacher networks. *Teacher Development*, 1–19. <https://doi.org/10.1080/13664530.2025.2568668>
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2
- Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 15–26). Routledge.