International Journal of Learning, Teaching and Educational Research
Vol. 25, No. 7, pp. 1063-1084, July 2026
https://doi.org/10.26803/ijlter.25.7.48
Received Mar 31, 2026; Revised Jul 7, 2026; Accepted Jul 14, 2026

Agentic Generative AI as a Dialogic Scaffold for Problem-Posing Pedagogy

Agnes M. Taveros*
Faculty of Veterinary Medicine,
Visayas State University, Philippines

Joedel Peñaranda
Faculty of Education,
The University of Melbourne, Australia

Marianne Agnes T. Mendoza
Biliran Province State University,
Philippines

Abstract. Dialogic problem posing requires learners to generate, justify, critique, and revise curriculum-aligned problems, yet these routines are difficult to sustain when an instructor must support several groups simultaneously. This study examined whether an answer-constrained agentic generative artificial intelligence (AI) workspace could function as a dialogic scaffold without displacing student agency. An explanatory sequential mixed-methods quasi-experimental design involved 96 students enrolled in four intact undergraduate course sections during an eight-week intervention. Two sections used an instructor-managed GPT-class AI workspace composed of planning, critique, and dialogue-coaching agents, while two comparison sections completed the same routines through instructor-guided printed prompts. Quantitative data included pretest and posttest problem-posing quality scores, peer dialogic engagement ratings from transcribed collaborative work, and substantive revision-cycle counts. Qualitative data came from focus groups, draft histories, final artefacts, and an instructor interview. Controlling for pretest performance and prior course achievement, the AI group outperformed the comparison group in problem-posing quality (partial eta squared =.13) and dialogic engagement (partial eta squared =.17) and completed more substantive revisions (d = 1.81). Qualitative findings indicated that the workspace broadened planning, surfaced coherence problems earlier, and prompted more equitable peer participation. These findings suggest that generative AI can support inquiry-oriented learning when it is bounded, answer-constrained, and

Citation :
Taveros, A. M., Peñaranda, J., & Mendoza, M. A. T. (2026). Agentic Generative AI as a Dialogic Scaffold for Problem-Posing Pedagogy. *International Journal of Learning, Teaching and Educational Research*, 25(7), 1063–1084.
https://doi.org/10.26803

*Corresponding author: Joedel Peñaranda; *joedel.penaranda@student.unimelb.edu.au*

positioned as a scaffold for critique and dialogue rather than a source of finished answers.

Keywords: agentic generative artificial intelligence; dialogic pedagogy; human-artificial intelligence collaboration; mixed-methods research; problem posing

1. Introduction

Problem posing repositions learners from answerers of ready-made questions to authors of purposeful inquiries. In critical pedagogy, asking learners to name, examine, and question situations is a humanising act because it treats students as meaning makers rather than passive recipients of knowledge (Freire, 2000). Although a substantial empirical tradition on problem posing has developed in mathematics education, the pedagogical logic is not limited to a single subject. Across language, science, social studies, vocational, and interdisciplinary classrooms, students who pose problems must decide what matters, specify conditions, anticipate how others will interpret a task, and make a problem worth discussing (Brown & Walter, 2005; Silver, 1994, 2013).

Collaborative problem posing is especially dependent on classroom dialogue. The quality of the final task rests not only on wording but also on the peer talk that tests whether the problem is coherent, meaningful, and intellectually demanding. Dialogic teaching emphasises reciprocal, cumulative, and purposeful talk in which students take up one another's ideas, ask for reasons, and refine meanings over time (Alexander, 2018; Howe & Abedin, 2013; Mercer & Littleton, 2007; Wegerif, 2007). In such routines, students do not simply generate questions; they learn to justify why a context matters, whether information is sufficient, and how a responder might reason from the available conditions.

Despite its value, dialogic problem posing is difficult to orchestrate in course settings. An instructor must monitor several groups, maintain curricular coherence, support quieter students, and move learners beyond superficial rewriting. Recent syntheses on problem-posing instruction show that stronger interventions tend to include explicit opportunities for critique, reflection, and revision (Cai et al., 2024; Zhang et al., 2024; Zhang & Cai, 2021). These conditions are demanding because they require sustained feedback while students are still forming ideas. Without such support, problem posing can become a one-time production task rather than a cycle of inquiry.

Generative artificial intelligence (AI) has intensified this design problem. Contemporary AI tools can provide immediate responses, adaptive prompts, and conversational support, but they can also invite answer substitution, weaken productive struggle, and obscure the boundary between student reasoning and machine output (Adiguzel et al., 2023; Bearman et al., 2023; Chan & Hu, 2023; Dwivedi et al., 2023; Kasneci et al., 2023; Lo, 2023). Recent research in IJLTER similarly shows that learners increasingly use generative AI tools and must evaluate chatbot responses critically rather than accept them as authoritative (Jere, 2026). The key question is therefore not whether AI can respond, but whether its response role can be pedagogically bounded.

This study uses the term agentic AI to describe a coordinated, goal-oriented arrangement of specialised AI functions that sustain activity across multiple steps. In contrast to a free-form chatbot that responds to any prompt, the agentic workspace in this study used separate planning, critique, and dialogue-coaching roles. These roles were configured to ask, clarify, and challenge rather than generate finished problems or final answers. The study therefore sits in current debates on human-AI collaboration, where scholars call for learning systems that preserve human judgement, make accountability visible, and align technology with pedagogical purpose (Kostopoulos et al., 2025; Markauskaite et al., 2022; Molenaar, 2022; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021).

The study contributes to IJLTER's scope by examining an innovative methodology for digital classrooms in an authentic course setting. It links learning theory, dialogic pedagogy, formative assessment, and responsible AI integration. Rather than treating AI as a general productivity tool, the study investigates whether an answer-constrained, instructor-managed AI workspace can function as a dialogic scaffold that strengthens peer reasoning and revision while keeping students accountable for their own problem designs.

1.1 Problem Statement and Research Objectives

Problem posing is widely recommended in inquiry-oriented curricula, yet its classroom enactment often stops at producing isolated questions. Students may create problems that are contextually interesting but underspecified, technically answerable but thin in intellectual purpose, or linguistically clear but weak in peer discussion value. The difficulty is compounded in group work because participation can become uneven: confident students may dominate task decisions while quieter learners remain peripheral to the reasoning process (Asterhan et al., 2020; Howe et al., 2019).

AI introduces both opportunity and risk into this setting. If students use AI as an answer engine, it may replace the very reasoning that problem posing is meant to cultivate. If AI is constrained as a scaffold, however, it may help distribute prompts for planning, critique, and participation more consistently than a single instructor can provide in real time. The present study addressed this instructional problem by testing an instructor-managed, answer-constrained agentic workspace in a dialogic problem-posing routine.

The study had four objectives: to determine whether the AI-supported routine improved the quality of student's posed problems, to examine whether it strengthened peer-to-peer dialogic engagement, to compare the number and meaning of substantive revision cycles across conditions, and to explain how students and the instructor experienced the benefits and constraints of the approach. These objectives foregrounded student agency because the workspace was not designed to compose finished problems, solve tasks, or assess students autonomously.

1.2 Research Questions

The study was guided by the following research questions:

1. To what extent does answer-constrained agentic AI-supported dialogic problem posing improve students' problem-posing quality?
2. To what extent does the intervention improve students' peer-to-peer dialogic engagement during collaborative work?
3. How do revision cycles differ between the agentic AI condition and the comparison condition?
4. How do students and the course instructor explain the benefits, constraints, and orchestration demands of the approach?

1.3 Significance and Originality of the Study

The study is significant for three reasons. First, it extends problem-posing research by focusing on revision and peer dialogue as mechanisms of problem quality rather than treating the final written problem as the only outcome. Second, it contributes to dialogic pedagogy by examining whether AI can support uptake, justification, generative questioning, and participation equity in group work. Third, it contributes to educational technology research by comparing a bounded agentic arrangement with a non-AI routine that used the same instructional sequence.

The originality of the design lies in the combination of three features: an answer-constrained AI guardrail, differentiated agent roles, and visible revision histories. Prior chatbot studies often examine broad learner-AI interaction or automated feedback. This study instead positions AI as a dialogue infrastructure that supports students in talking to one another. In doing so, it offers a practical model for using generative AI without surrendering authorship, assessment judgement, or instructional authority to the system.

2. Literature Review and Conceptual Framework

This section synthesises four bodies of literature that informed the intervention. It first situates problem posing as a learning and formative assessment process across disciplines. It then reviews dialogic pedagogy and participation equity, examines AI as a bounded scaffold, and presents the conceptual model that guides the intervention.

2.1 Problem Posing as Learning and Assessment Across Disciplines

Problem posing refers to generating new problems, reformulating existing ones, or modifying given conditions so that learners shape the problem space rather than respond only to a teacher-given task (Brown & Walter, 2005; Silver & Cai, 1996; Stoyanova & Ellerton, 1996). Much of the formal empirical literature comes from mathematics education, but the activity is pedagogically broader. Across disciplines, students identify relevant information, make assumptions explicit, define purpose, and design a task that another learner can interpret and respond to.

Because learners must select relations, constraints, evidence, or conditions, problem posing can reveal understanding in ways that conventional answer-focused assessments often miss (Silver, 1994, 2013). A well-posed classroom problem is coherent, contextually meaningful, cognitively purposeful, and open enough to invite explanation, inquiry, or extension (Cai et al., 2024; Singer et al., 2013). In this sense, problem posing functions simultaneously as a learning activity, a record of reasoning, and a formative assessment practice.

Revision is central to this process. Initial drafts often show incomplete constraints, vague language, or weak purpose. Through critique and revision, students test whether a peer can understand the task and whether the problem invites meaningful reasoning. Meta-analytic evidence suggests that problem-posing interventions are stronger when they include explicit opportunities to reflect, revise, and connect problem design to learning goals (Zhang et al., 2024). The present study therefore treated substantive revision as a process indicator rather than a peripheral activity.

2.2 Dialogic Pedagogy, Participation, and Student Agency

Dialogic pedagogy treats learning as the expansion of a shared space where participants question, revoice, compare, and refine ideas (Alexander, 2018; Wegerif, 2007). In a dialogic classroom, talk is not merely participation. It becomes epistemic work because students make reasons public, respond to alternatives, and co-construct more robust meanings. This is particularly important in problem posing, where a proposed task often becomes meaningful only after peers attempt to interpret and challenge it.

Research over several decades connects dialogic interaction with stronger reasoning when talk includes uptake of peer ideas, justification, elaboration, and equitable participation (Howe & Abedin, 2013; Howe et al., 2019; Mercer & Littleton, 2007). These features align with problem posing because students must test the sufficiency of information and the clarity of task demands. Bakker et al. (2015) also show that dialogic scaffolding can support disciplinary reasoning when students are guided to connect informal talk with more formal conceptual language.

However, dialogue does not emerge automatically. Students may dominate, defer, or remain at the level of quick opinion unless classroom routines provide norms and talk moves. Asterhan et al. (2020) emphasise that productive dialogue depends on how disagreement, uptake, and responsibility are organised. For the present study, this meant that AI support had to do more than generate suggestions. It had to prompt the social conditions through which students could justify, critique, and revise problems together.

2.3 Generative AI, Agentic AI, and Human Control

AI in education has evolved from intelligent tutoring systems, analytics, and adaptive platforms towards conversational and generative systems that participate in longer sequences of learning activity (Hwang et al., 2020; Poquet & de Laat, 2021; Roll & Wylie, 2016; Schiff, 2021; Wang et al., 2024; Zawacki-Richter et al., 2019). These systems can provide immediate feedback and reduce delays in formative support, but their educational value depends on design. When students use AI to obtain complete answers, the technology can reduce cognitive engagement and create overreliance (Kasneci et al., 2023; Zhai et al., 2024).

Recent literature increasingly distinguishes between reactive generative AI, intelligent tutoring systems, and agentic AI. Free-form generative AI can produce flexible responses but offers little built-in alignment with classroom norms. Intelligent tutoring systems can provide structured support, but they often focus on predetermined knowledge paths. Agentic AI introduces coordinated role-based functions, memory, planning, and multi-step support, which may be useful for complex classroom routines if governance and human control are explicit (Bozkurt et al., 2026; Kostopoulos et al., 2025).

The present study responds to calls for pedagogy-first integration across educational levels. Tang et al. (2024), for example, conceptualise generative AI in education as a dialogic agent that can help students engage critically with multiple perspectives rather than accept AI-generated content as final authority. Chiu (2024), Crompton and Burke (2023), and An et al. (2025) similarly emphasise that AI integration requires clear institutional purposes, ethical boundaries, and educator capacity. Reviews of AI dialogue systems in language education likewise suggest that conversational support should be judged by interactional competence, socio-affective learning, and learner responsibility rather than response fluency alone (Jeon, 2025; Zhai & Wibowo, 2023; Zhou et al., 2026). This study operationalised that principle through an answer-constrained workspace that redirected answer-seeking into planning, critique, and peer dialogue.

Table 1: Comparison of AI support models relevant to classroom problem posing

Model	Typical instructional role	Risk for problem posing	Design response in this study
Free-form generative AI	Produces open-ended text in response to broad student prompts.	Can supply finished problems, replace authorship, or normalise answer substitution.	Restricted direct answer generation and redirected students to planning and critique questions.
Intelligent tutoring system	Guides learners through predetermined content or solution pathways.	May narrow inquiry to fixed steps and treat success as correctness rather than problem quality.	Used open planning and dialogic prompts while keeping final judgement with students and the instructor.
Answer-constrained Agentic AI workspace	Coordinates planning, critique, and dialogue-coaching functions across an inquiry cycle.	May still create dependency if instructor oversight and task boundaries are weak.	Separated agent roles, retained revision histories, and required peer justification before final submission.

Note. The comparison synthesises design distinctions discussed in recent AI in education literature and explains the functional choice made in the intervention.

2.4 Conceptual Framework and Intervention Model

The conceptual framework combined three traditions. First, Freirean problem-posing pedagogy emphasised learner voice, social meaning, and inquiry into lived or familiar situations (Freire, 2000). Second, structured problem-posing research foregrounded invention, reformulation, coherence, and response-worthiness as dimensions of quality (Cai et al., 2024; Silver, 2013). Third, dialogic pedagogy treated learning as collaborative meaning-making in which ideas become stronger through questioning, justification, and revision (Alexander, 2018; Mercer & Littleton, 2007).

The workspace was designed according to four principles: ask before telling, preserve student authorship, make revisions visible, and keep the instructor in charge of final pedagogical judgement. The expected mechanism of change was that consistent planning, and critique prompts would increase substantive revision, while dialogue-coaching prompts would improve uptake, justification, and participation equity. Figure 1 summarises the intervention cycle.

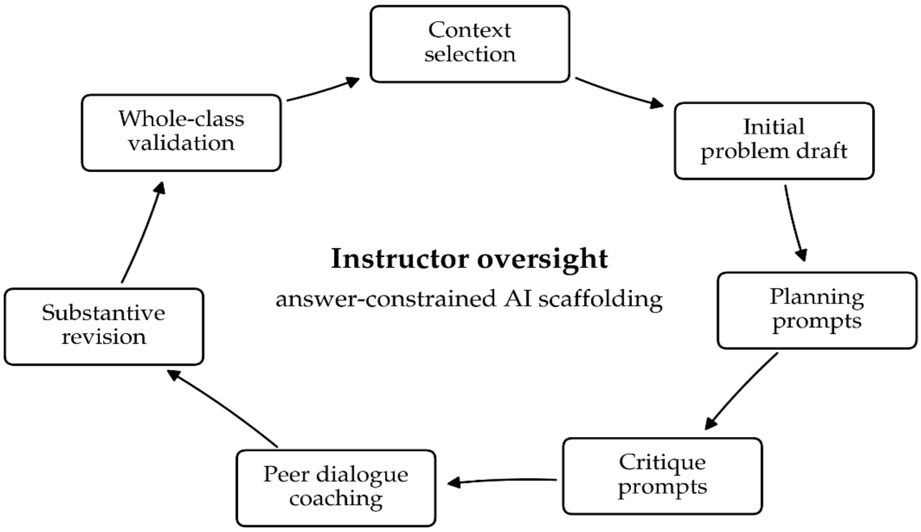

Figure 1: Agentic dialogic problem-posing cycle used in the intervention

Students began from a familiar or curriculum-relevant context, drafted an initial problem, consulted the planning agent, tested the draft through critique prompts and peer dialogue, revised the problem, and presented selected outputs for whole-class validation. The dialogue coach did not evaluate correctness; it suggested talk moves such as revoicing, asking for evidence, inviting another voice, and checking whether a proposed problem allowed meaningful interpretation.

3. Method

This section explains the mixed-methods design, participants, intervention procedures, instruments, validity evidence, ethics, and analytic procedures. It also clarifies the instructor’s orchestration role and the safeguards used to prevent AI from functioning as a solution engine.

3.1 Research Design and Context

The study used an explanatory sequential mixed-methods quasi-experimental design (Creswell & Plano Clark, 2018). We collected quantitative data first to estimate group differences in problem-posing quality, dialogic engagement, and revision cycles. Qualitative data were then used to explain how students and the instructor experienced the intervention and how the observed outcomes emerged in course practice.

The study was implemented in four intact course sections at a public university. The sections served students from several undergraduate programmes. To preserve site anonymity and foreground design transferability, institutional descriptors are reported only in broad terms. Both conditions completed the same course-aligned inquiry unit over eight weeks. The unit required students to interpret familiar or disciplinary scenarios, identify relevant information, compare alternatives, and construct problems that peers could discuss and respond to. The same course instructor facilitated all four sections, reducing instructor-level variation while preserving the routine structure of course delivery.

3.2 Participants and Baseline Equivalence

Participants were 96 students enrolled in four intact undergraduate course sections facilitated by the same instructor. We assigned two sections to the agentic AI condition (n = 48) and two to the comparison condition (n = 48). The sample represented several fields of study. Available academic records showed similar prior course achievement across groups, and pretest scores indicated comparable starting levels on the outcome measures. Table 2 presents the participant profile and baseline equivalence.

Table 2: Participant profile and baseline equivalence

Variable	Agentic AI (n = 48)	Comparison (n = 48)	Statistic	p
Female, n (%)	31 (64.6)	23 (47.9)	$\chi^2 = 2.07$	.150
Prior course grade, M ± SD	85.23 ± 3.39	84.35 ± 4.15	t = 1.13	.262
Problem-posing quality pretest, M ± SD	10.85 ± 1.77	10.66 ± 1.56	t = 0.55	.585
Dialogic engagement pretest, M ± SD	2.73 ± 0.34	2.70 ± 0.37	t = 0.39	.697

Note. AI = artificial intelligence; M = mean; SD = standard deviation. All baseline comparisons were not statistically significant at p <.05.

3.3 Intervention, Instructor Orchestration, and AI Workspace

Each week comprised two 90-minute class meetings. Students first produced an individual draft of a problem, then worked in pairs or triads to test and revise their drafts, and finally shared selected outputs during a whole class debrief. Tasks were grounded in familiar or disciplinary contexts such as campus decision-making, household resource use, everyday purchasing, community planning, health information, media claims, and environmental choices. This flexible design kept the intervention cross-disciplinary rather than bound to a single field.

Before implementation, the research team and course instructor aligned the weekly routines, prompt language, data privacy procedures, and debrief expectations. The AI workspace did not reduce the instructor's role. The instructor introduced each task, monitored group work, redirected inappropriate AI use, validated final outputs, and led whole-class discussions. This orchestration was important because no AI response was treated as authoritative without student justification and instructor review.

The agentic workspace was implemented as an instructor-managed generative AI interface based on a GPT-class large language model. The study evaluated the pedagogical configuration rather than the comparative performance of a single commercial model. For transparency, the workspace is described by its functional configuration: fixed role prompts, separate planning and critique functions, dialogue-coaching prompts, answer-blocking response rules, and a visible revision history. Students did not use an unrestricted public chatbot.

In the agentic AI condition, students used three coordinated components. The planning agent asked students to specify the central issue, relevant information, constraints, and intended response. The critique agent checked coherence, sufficiency of information, plausibility, and clarity of wording. The dialogue coach suggested peer talk moves, including challenging assumptions, requesting evidence, revoicing a partner's idea, or inviting a quieter member to contribute. The interface maintained draft histories so students could examine how their problems changed.

In the comparison condition, students completed the same sequence without AI mediation. They used a printed prompt set developed with the instructor to mirror the planning, critique, and dialogue functions of the experimental condition. Learning objectives, task contexts, group routines, and public debrief were kept parallel across conditions. This design allowed the study to estimate the value added by agentic mediation beyond using a structured dialogic routine alone.

To protect student authorship, the workspace was configured so that attempts to solicit direct answers or finished tasks returned redirective prompts such as "What information is still missing?" or "How could your group restate the problem more clearly?" rather than completed responses. These prompts were original to the intervention design and are summarised in Appendix 1.

3.4 Instruments, Validity, and Reliability

Problem-posing quality was measured using the problem-posing quality rubric, an analytic rubric with four dimensions: contextual authenticity, disciplinary coherence, cognitive demand, and openness for inquiry or response. Each dimension was scored from 1 to 5, yielding a total score from 4 to 20. Pretest and posttest tasks were parallel performance tasks aligned with the course learning outcomes. Two trained raters independently scored all responses. Interrater agreement for the total score was strong, with an intraclass correlation coefficient (ICC) of.88.

Dialogic engagement was measured using the dialogic engagement observation scale, which rated four indicators: uptake of peer ideas, use of justification, generative questioning, and equitable contribution. Each indicator was scored from 1 to 5 and averaged to produce a composite score. Ratings were based on transcribed 12-minute collaborative segments sampled at pretest and posttest. The 12-minute duration was used consistently across groups to capture a complete planning, peer testing, and revision episode while keeping transcription and rating demands manageable. Interrater agreement for the composite score was strong (ICC =.86).

A revision-cycle count was derived from draft histories and artefact comparisons. A substantive revision was counted when a student or group changed a key relation or assumption, added or removed a constraint, clarified the problem goal, or altered the question demand in a meaningful way. Surface revisions, including spelling correction, punctuation, isolated word substitution, or cosmetic rephrasing that did not alter the task's reasoning demand, were not counted.

Content adequacy was supported through expert review of the rubrics, alignment checking with task objectives, and rater calibration using sample responses before independent scoring. Confirmatory factor analysis was not conducted because the instruments were analytic performance-rating tools with four theoretically defined dimensions and because the sample was not designed for latent measurement modelling. This limitation is acknowledged in section 8, and future multi-site studies should validate the scales using larger item-level datasets.

3.5 Data Sources, Ethics, and Data Protection

Quantitative data included pretest and posttest problem-posing quality scores, dialogic engagement ratings, prior course achievement, and revision-cycle counts. Qualitative sources included four student focus groups involving 24 participants, selected draft histories, final problem artefacts, and the instructor's post-intervention interview. Written informed consent was obtained from each participant before data collection, and participation was voluntary and did not affect course grades. We removed names from transcripts and artefacts, and excerpts were reported without institutional location or identifying details. Raw transcripts were not publicly deposited because the consent agreement did not authorise unrestricted release of interaction data.

3.6 Data Analysis

Independent-samples tests examined baseline equivalence. Intervention effects on posttest problem-posing quality and dialogic engagement were estimated through analyses of covariance, controlling for the relevant pretest score and prior course grade. Revision-cycle counts were compared using an independent-samples t-test. Effect sizes are reported as partial eta squared for covariance models and Cohen's d for the revision comparison. Residual plots and variance checks indicated that model assumptions were acceptable.

Qualitative data were coded through iterative comparison of focus group transcripts, draft histories, artefacts, and the instructor interview. Initial codes captured ideation, critique, revision, participation, answer-seeking, and instructor orchestration. These codes were then clustered into themes that explained how the workspace shaped task interpretation and group interaction. Integration occurred through a joint reading of the statistical outcomes and qualitative themes to identify meta-inferences across data sources.

4. Results

This section presents the quantitative and qualitative findings. It begins with baseline equivalence, reports effects on problem-posing quality and dialogic engagement, examines revision cycles as process evidence, and then integrates qualitative explanations from students, artefacts, and the instructor interview.

4.1 Baseline Equivalence

As shown in Table 2, the two groups were similar at baseline. No statistically significant differences were observed in gender distribution, prior course grade, problem-posing quality pretest scores, or dialogic engagement pretest scores. All p values exceeded.05, supporting the interpretation that posttest differences were associated with the intervention rather than initial group imbalance.

4.2 Effects on Problem-Posing Quality and Dialogic Engagement

Table 3: Outcome means and posttest comparisons

Outcome	AI pre	Comp. pre	AI post	Comp. post	Test/effect
Problem-posing quality	10.85 ± 1.77	10.66 ± 1.56	13.26 ± 1.52	12.13 ± 1.59	Adj. means 13.21 vs. 12.17; $F(1, 92) = 13.19$; $p < .001$; partial eta squared =.13
Dialogic engagement	2.73 ± 0.34	2.70 ± 0.37	3.11 ± 0.28	2.84 ± 0.33	Adj. means 3.10 vs. 2.85; $F(1, 92) = 18.18$; $p < .001$; partial eta squared =.17
Revision cycles	NA	NA	5.60 ± 1.07	3.91 ± 0.78	$t(94) = 8.88$; $p < .001$; $d = 1.81$

Note. AI = artificial intelligence; NA = not applicable because revision cycles were counted during the intervention period.

Students in the agentic AI condition outperformed the comparison group on both primary outcomes. For problem-posing quality, the agentic AI group improved from a pretest mean of 10.85 (SD = 1.77) to a posttest mean of 13.26 (SD = 1.52), while the comparison group improved from 10.66 (SD = 1.56) to 12.13 (SD = 1.59). After controlling for pretest performance and prior course grade, the posttest advantage remained significant, $F(1, 92) = 13.19$, $p < .001$, partial eta squared =.13. The adjusted means were 13.21 and 12.17, respectively.

A similar pattern appeared for dialogic engagement. The agentic AI group increased from 2.73 (SD = 0.34) to 3.11 (SD = 0.28), while the comparison group increased from 2.70 (SD = 0.37) to 2.84 (SD = 0.33). The covariance model indicated a significant intervention effect, F (1, 92) = 18.18, p <.001, partial eta squared =.17. Figure 2 presents the adjusted posttest means and observed revision-cycle means.

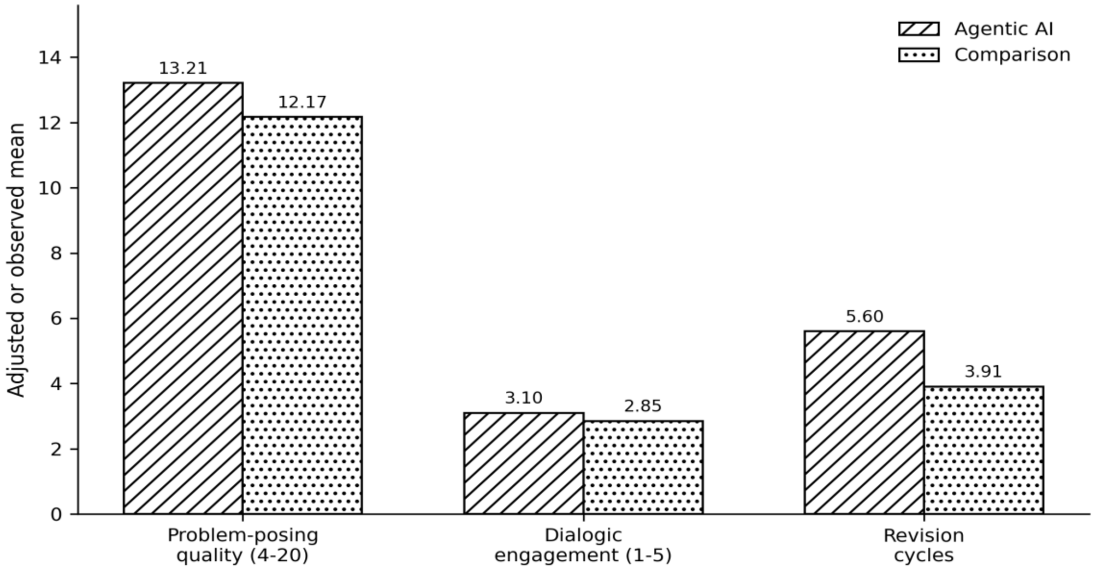

Figure 2: Adjusted posttest outcomes and revision-cycle means by condition

4.3 Revision Cycles as Process Evidence

Process data were consistent with the outcome patterns. Students in the agentic AI group completed more substantive revisions (M = 5.60, SD = 1.07) than those in the comparison group (M = 3.91, SD = 0.78), t(94) = 8.88, p <.001, d = 1.81. This difference suggests that the intervention did not merely accelerate drafting; it extended students' willingness to revisit and refine problem structure before submission.

Across artefact histories, substantive revisions typically involved adding missing conditions, clarifying the intended respondent, narrowing an overly broad prompt, or revising the question demand to invite explanation rather than recall. Higher revision counts were associated with stronger posttest problem-posing quality, suggesting that the intervention worked in part by lengthening the cycle of critique and repair. In practical classroom terms, the workspace made revision more visible, more frequent, and more consequential.

4.4 Qualitative Explanations of the Intervention Effect

The qualitative analysis clarified why the agentic AI group produced stronger outcomes. Three themes recurred across focus groups, draft histories, final artefacts, and the instructor interview: authorship through prompted planning, critique as a normal part of revision, and dialogue coaching as support for participation equity. Table 4 summarises the integrated themes.

Table 4: Integrated qualitative themes explaining the intervention effect

Theme	Integrated finding	Representative excerpt
Authorship through prompted planning	Planning prompts helped students specify key concepts, constraints, and intended question demand before peer testing. Artefact histories in the agentic AI group showed more explicit conditions and fewer vague endings.	"It did not make the problem for us. It kept asking what the key idea should be, so we had to decide." (Student, Focus Group 1)
Critique as a normal part of revision	Critique prompts exposed missing information, implausible assumptions, and wording that weakened interpretability. Many agentic AI group drafts moved through several meaningful revisions before submission.	"We thought the question was already okay, but the critique asked if another person could respond from the information we gave. That made us change the conditions." (Student, Focus Group 3)
Dialogue coaching and participation equity	The dialogue coach prompted revoicing, requests for evidence, and invitations for quieter students to speak. The instructor reported more consistent turn-taking but still intervened when students tried to shortcut reasoning.	"The useful part was consistency. Every group got nudged to justify and invite another voice, but I still had to step in when they tried to fish for an answer." (Instructor interview)

Note. Excerpts were lightly edited for anonymity and readability without altering meaning.

First, the planning agent widened the possibility space without displacing authorship. Students described the prompts as helping them decide what the central issue, constraints, and context should mean rather than giving them a finished problem. This helps explain the improvement in problem-posing quality: the prompts pushed students to decide what kind of intellectual work their problem should require.

Second, the critique agent normalised revision by making flaws in wording or coherence visible early. Students did not describe critique as correction imposed by the tool. They described it as a check that helped them see whether another student could interpret and respond to the problem from the information provided.

Third, the dialogue coach redistributed talk, especially in mixed-ability groups. Students reported that the prompts helped them ask one another for evidence, restate ideas more clearly, and invite fewer vocal members into the conversation. The instructor's account showed that the workspace was most useful when it supported, rather than replaced, instructor orchestration.

5. Discussion

The findings indicate that answer-constrained agentic AI can improve dialogic problem posing when it is designed as a bounded scaffold. This section interprets the results in relation to problem-posing theory, dialogic pedagogy, AI design, and classroom implementation.

5.1 Why Design Improved Problem Quality

The first contribution of the intervention was to externalise the hidden architecture of problem posing. By repeatedly asking students to identify the central issue, specify constraints, and justify the intended response, the planning agent shifted attention from surface context toward task structure. This finding aligns with problem-posing scholarship that treats reformulation and constraint management as central to deeper learner thinking (Brown & Walter, 2005; Cai et al., 2024; Silver & Cai, 1996).

A second contribution was the normalisation of revision as disciplined authorship. The comparison group also improved, which indicates that the dialogic routine itself had value. However, the larger gains in the agentic AI group suggest that consistent critique prompts increased opportunities for students to repair ambiguity, insufficiency, and weak task demand. This pattern supports recent evidence that problem-posing interventions are strongest when revision is explicit rather than incidental (Zhang et al., 2024).

5.2 Contribution to Dialogic Pedagogy and Participation Equity

The dialogic engagement gains show that AI support can influence the social conditions of reasoning, not only individual feedback. Dialogue research has long emphasised uptake, justification, and cumulative talk as mechanisms for learning (Alexander, 2018; Howe et al., 2019; Mercer & Littleton, 2007). In the present study, the dialogue coach appeared to help groups sustain these moves more consistently by giving students language for inviting evidence, revoicing, and challenging assumptions.

This result extends earlier work on conversational and peer-dialogue agents (Howard et al., 2017; Tegos et al., 2014) by placing AI support inside a whole-class pedagogical routine. Students did not simply interact with a tool. They used tool prompts to organise how they interacted with peers. The social mechanism matters because the goal of dialogic problem posing is not polished AI-assisted wording but stronger shared reasoning.

5.3 AI Autonomy, Human Control, and the Answer-Constrained Guardrail

The study also clarifies the difference between free-form generative AI and answer-constrained agentic scaffolding. Free-form tools can provide fast and fluent responses, but their flexibility can make them difficult to align with classroom authorship norms (Kasneci et al., 2023; Zhai et al., 2024). Recent work on AI agents warns that increasing system autonomy can create new assessment and verification problems if institutions prioritise output over evidence of human learning (Bozkurt et al., 2026). The present design responded by limiting autonomy at the point where student reasoning had to remain visible.

This distinction addresses the concern that AI innovation may become novelty driven. The intervention did not ask whether AI could generate good problems. It asked whether a bounded AI arrangement could prompt students to generate, test, and revise better problems. This interpretation is consistent with Tang et al. (2024), whose proof-of-concept study on education frames generative AI as a support for co-construction rather than a definitive knowledge provider. Technical limitations were still visible. The AI critique was sometimes generic or overly cautious, and the instructor had to redirect students who attempted to fish for finished answers. These constraints reinforce the need for instructor validation and transparent logs. The agentic workspace was valuable not because it was autonomous, but because its autonomy was deliberately narrowed to planning, critique, and dialogue support.

5.4 Implications for Practice, Policy, and Curriculum Design

For instructors, the findings suggest that AI integration should begin with narrow, visible, and pedagogically accountable roles. Planning prompts, sufficiency checks, and dialogue prompts can be introduced before more complex AI functions are considered. Professional development should therefore address orchestration, not only tool operation: instructors need strategies for sequencing individual drafting, peer testing, AI consultation, and public validation.

For institutional leaders and policymakers, the findings support governance that distinguishes between answer-generation tools and process-oriented scaffolds. Recent reviews of AI agents and AI-supported programming education identify overreliance, implementation barriers, and academic integrity as persistent concerns (Elnaffar et al., 2026; Zhai et al., 2024). Institutions should therefore require clear prompt boundaries, data protection procedures, instructor-accessible logs, and direct guidance for students who use AI-supported learning environments.

For curriculum design, the study suggests that problem posing can serve as a cross-disciplinary AI-resilient task when assessment values process evidence. Draft histories, peer justifications, and revision explanations show forms of thinking that final answers alone conceal. This may be especially useful in language, social science, science, and interdisciplinary inquiry, where students must interpret contexts, evaluate evidence, and make their assumptions visible. Future adaptations for students with learning disabilities may include simplified prompt menus, speech-to-text options, peer role cards, and sentence starters that preserve intellectual demand while reducing access barriers.

6. Conclusion

This study examined whether answer-constrained agentic generative AI can strengthen dialogic problem posing without displacing learner agency. Across an eight-week quasi-experimental intervention in four intact course sections, students who used an instructor-managed AI workspace for planning, critique, and dialogue coaching produced higher-quality problems, engaged more dialogically with peers, and completed more substantive revisions than students who used parallel printed prompts. The mixed-methods findings indicate that the intervention worked through three connected mechanisms: it helped learners clarify the structure and purpose of a problem before finalising it, normalised critique and revision as forms of disciplined authorship, and promoted more equitable participation through revoicing, evidence requests, and invitations to quieter group members. The study does not suggest that AI should replace instructor expertise, peer struggle, or student responsibility. Its contribution is more specific: generative AI can support humanising inquiry when it is bounded, answer-constrained, transparent, and embedded in instructor-led routines. The main limitation is the single-institution, single-instructor design, so transferability requires testing across disciplines, course levels, and institutional contexts. For curriculum design, the findings support AI-assisted tasks that assess process evidence, including drafts, justifications, and revisions, rather than only polished final products. Such designs may help institutions use AI to strengthen dialogue and authorship instead of outsourcing learning.

7. Recommendations and Future Studies

Instructors who wish to adopt similar designs should begin with limited and transparent AI roles: clarifying the central issue, checking information sufficiency, prompting justification, and supporting equitable talk. AI should not be used to generate finished student tasks during problem posing. Students should be required to explain why they accepted, modified, or rejected AI prompts.

Professional development should include four modules: designing answer-constrained prompts, facilitating peer dialogue with AI support, interpreting revision histories as formative evidence, and responding to answer-seeking behaviour. Institutions should also establish data privacy rules, direct student communication procedures, and instructor-accessible audit logs before deploying AI-supported workspaces. Where infrastructure or privacy risks are high, local or institutionally managed AI tools should be explored before public chatbot use. The approach may also be applied to veterinary and related professional curricula by integrating AI-supported dialogic scaffolds into problem-posing and inquiry-based learning.

Future studies should test the model across different disciplines, course levels, and institution types, including settings with limited digital infrastructure. Longer interventions are needed to examine whether dialogic gains persist after novelty effects diminish. Larger studies should also validate the rubrics through confirmatory factor analysis or structural equation modelling, compare the separate contribution of the planning, critique, and dialogue-coaching components, and examine whether the intervention benefits lower-performing, quieter, or differently abled learners in distinctive ways.

8. Limitations of the Study

This study has limitations. It was conducted in one institution, across four intact course sections, with a single instructor and an eight-week duration. Although the common instructor and parallel tasks strengthened internal comparison, the design cannot eliminate all section-level influences and cannot support strong causal claims beyond the studied context.

The outcome measures centred on classroom problem posing and peer dialogue rather than long-term achievement. The study also focused on a bounded AI configuration that redirected answer-seeking; other AI configurations could produce different results, including fewer desirable ones. Confirmatory factor analysis was not conducted for the performance rubrics, so construct validity rests on expert review, theoretical alignment, and interrater reliability. Finally, the qualitative component explained likely mechanisms of change but did not code every interaction across every lesson.

Conflict of Interest
The authors declare that they have no competing interests related to this study.

9. Data Availability Statement
De-identified analytic summaries, scoring rubrics, and sample prompts are available from the corresponding author upon reasonable request. Raw student transcripts and unredacted artefacts are not publicly available because the consent agreement did not permit unrestricted release of interaction data.

10. Author Contributions
All authors contributed to conceptualisation, study design, interpretation of findings, manuscript revision, and approval of the final version. Agnes M. Taveros led the investigation, literature search, participant coordination, data curation, and preparation of the initial draft. Joedel Peñaranda supervised the study, led formal analysis and project administration, and critically revised the manuscript. Marianne Agnes T. Mendoza contributed to instrument validation, literature integration, and manuscript refinement.

11. Acknowledgements
The authors thank the participating students, the course instructor, and the academic administrators who supported the study. The authors used ChatGPT (GPT-5.5; OpenAI) to assist with spelling correction, grammatical editing, and language refinement to improve the clarity and readability of the manuscript. The authors alone made all substantive intellectual content, scientific judgements, and editorial decisions. The authors critically reviewed and verified all revisions, accept full responsibility for the content of the manuscript, and approve its final version for submission and publication.

12. References
Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology, 15*(3), Article ep429. https://doi.org/10.30935/cedtech/13152

Alexander, R. (2018). Developing dialogic teaching: Genesis, process, trial. *Research Papers in Education, 33*(5), 561–598. https://doi.org/10.1080/02671522.2018.1481140

An, Y., Yu, J. H., & James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education, 22*, Article 10. https://doi.org/10.1186/s41239-025-00507-3

Asterhan, C. S. C., Howe, C., Lefstein, A., Matusov, E., & Reznitskaya, A. (2020). Controversies and consensus in research on dialogic teaching and learning. *Dialogic Pedagogy, 8*, S1–S16. https://doi.org/10.5195/dpj.2020.312

Bakker, A., Smit, J., & Wegerif, R. (2015). Scaffolding and dialogic teaching in mathematics education: Introduction and review. *ZDM Mathematics Education, 47*(7), 1047–1065. https://doi.org/10.1007/s11858-015-0738-8

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education, 86*(2), 369–385. https://doi.org/10.1007/s10734-022-00937-2

Bozkurt, A., Crompton, H., & Fell Kurban, C. (2026). The devil is in the det[ai]ls: AI agents, ghost students, and the crisis of verified presence in an agentic AI world. *Open Praxis, 18*(1), 1–12. https://doi.org/10.55982/openpraxis.18.1.1145

Brown, S. I., & Walter, M. I. (2005). *The art of problem posing* (3rd ed.). Lawrence Erlbaum Associates. https://eric.ed.gov/?id=ED493812

Cai, J., Koichu, B., Rott, B., & Jiang, C. (2024). Advances in research on mathematical problem posing: Focus on task variables. *Journal of Mathematical Behavior, 76,* Article 101186. https://doi.org/10.1016/j.jmathb.2024.101186

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence, 6,* Article 100197. https://doi.org/10.1016/j.caeai.2023.100197

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20,* Article 22. https://doi.org/10.1186/s41239-023-00392-8

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D.,... Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elnaffar, S., Rashidi, F., & Abualkishik, A. Z. (2026). Teaching with AI: A systematic review of chatbots, generative tools, and tutoring systems in programming education. *International Journal of Learning, Teaching and Educational Research, 25*(1). https://doi.org/10.26803/ijlter.25.1.1

Freire, P. (2000). *Pedagogy of the oppressed* (30th anniversary ed.). Continuum.

Howard, C., Jordan, P., Di Eugenio, B., & Katz, S. (2017). Shifting the load: A peer dialogue agent that encourages its human collaborator to contribute more to problem solving. *International Journal of Artificial Intelligence in Education, 27*(1), 101–129. https://doi.org/10.1007/s40593-015-0071-y

Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education, 43*(3), 325–356. https://doi.org/10.1080/0305764X.2013.786024

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher-student dialogue during classroom teaching: Does it really impact on student outcomes? *Journal of the Learning Sciences, 28*(4–5), 462–512. https://doi.org/10.1080/10508406.2019.1573730

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, Article 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jeon, E.-Y. (2025). Artificial intelligence in ESL/EFL education: Evidence from recent reviews (2024–2025). *International Journal of Learning, Teaching and Educational Research, 24*(10), 509–526. https://doi.org/10.26803/ijlter.24.10.24

Jere, S. (2026). Can machines think? Assessing the accuracy of GenAI chatbots in a physics university entrance exam. *International Journal of Learning, Teaching and Educational Research, 25*(1), 332–352. https://doi.org/10.26803/ijlter.25.1.17

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T.,... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kostopoulos, G., Gkamas, V., Rigou, M., & Kotsiantis, S. (2025). Agentic AI in education: State of the art and future directions. *IEEE Access, 13*, 177467–177491. https://doi.org/10.1109/ACCESS.2025.3620473

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

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Siemens, G., & Lodge, J. M. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence, 3*, Article 100056. https://doi.org/10.1016/j.caeai.2022.100056

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* Routledge. https://www.routledge.com/Dialogue-and-the-Development-of-Childrens-Thinking-A-Sociocultural-Approach/Mercer-Littleton/p/book/9780415404792

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Poquet, O., & de Laat, M. (2021). Developing capabilities: Lifelong learning in the age of AI. *British Journal of Educational Technology, 52*(4), 1695–1708. https://doi.org/10.1111/bjet.13123

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Schiff, D. (2021). Out of the laboratory and into the classroom: The future of artificial intelligence in education. *AI & Society, 36*(1), 331–348. https://doi.org/10.1007/s00146-020-01033-8

Silver, E. A. (1994). On mathematical problem posing. *For the Learning of Mathematics, 14*(1), 19–28. https://flm-journal.org/Articles/2A5D152778141F58C1966ED8673C15.pdf

Silver, E. A. (2013). Problem-posing research in mathematics education: Looking back, looking around, and looking ahead. *Educational Studies in Mathematics, 83*(1), 157–162. https://doi.org/10.1007/s10649-013-9477-3

Silver, E. A., & Cai, J. (1996). An analysis of arithmetic problem posing by middle school students. *Journal for Research in Mathematics Education, 27*(5), 521–539. https://doi.org/10.2307/749846

Singer, F. M., Ellerton, N., & Cai, J. (2013). Problem-posing research in mathematics education: New questions and directions. *Educational Studies in Mathematics, 83*(1), 1–7. https://doi.org/10.1007/s10649-013-9478-2

Stoyanova, E., & Ellerton, N. F. (1996). A framework for research into students' problem posing in school mathematics. In P. C. Clarkson (Ed.), *Technology in mathematics* *education* (pp. 518–525). Mathematics Education Research Group of Australasia. https://merga.net.au/publications/annual-conference-proceedings/1996-conference-proceedings/

Tang, K. S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024). A dialogic approach to transform teaching, learning & assessment with generative AI in secondary education: A proof of concept. *Pedagogies: An International Journal, 19*(3), 493–503. https://doi.org/10.1080/1554480X.2024.2379774

Tegos, S., Demetriadis, S., & Tsiatsos, T. (2014). A configurable conversational agent to trigger students' productive dialogue: A pilot study in the CALL domain. *International Journal of Artificial Intelligence in Education, 24*(1), 62–91. https://doi.org/10.1007/s40593-013-0007-3

UNESCO. (2021). *AI and education: Guidance for policymakers*. United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54675/PCSP7350

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications, 252,* Article 124167. https://doi.org/10.1016/j.eswa.2024.124167

Wegerif, R. (2007). *Dialogic education and technology: Expanding the space of learning*. Springer. https://doi.org/10.1007/978-0-387-71142-3

Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence, 4,* Article 100134. https://doi.org/10.1016/j.caeai.2023.100134

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11, Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, C., Zhou, Y., Wijaya, T. T., Chen, J., & Ning, Y. (2024). Effects of problem posing instructional interventions on student learning outcomes: A three-level meta-analysis. *Thinking Skills and Creativity, 53*, Article 101587. https://doi.org/10.1016/j.tsc.2024.101587

Zhang, H., & Cai, J. (2021). Teaching mathematics through problem posing: Insights from an analysis of teaching cases. *ZDM Mathematics Education, 53*(4), 961–973. https://doi.org/10.1007/s11858-021-01260-3

Zhou, B., Lim, S. P., Kabir, N., Asna, A., & Lu, T. N. (2026). From automation to socio-affective learning: A bibliometric analysis of artificial intelligence in English as a foreign language education. *Problems of Education in the 21st Century, 84*(1), 229–247. https://doi.org/10.33225/pec/26.84.229

Appendix 1

Sample Prompts Used in the Agentic Workspace

The prompts below summarise the roles assigned to each component of the workspace. They are included to support transparency and replication of the intervention design.

Table A1: Sample prompts used to operationalize the agentic workspace

Workspace component	Pedagogical function	Sample prompts used during the intervention
Planning agent	Clarifies the conceptual structure of the proposed problem before peer testing begins.	What is the central issue or unknown? Which ideas, concepts, or information sources matter most? What information does a responder need but not yet know?
Critique agent	Checks coherence, sufficiency of information, and interpretive clarity.	Can another student respond to this from the information given? Are any details implausible, inconsistent, or irrelevant? Is the wording precise enough to allow one clear interpretation?
Dialogue coach	Prompts productive peer talk and participation equity while the group evaluates a draft.	Can someone restate the idea in a different way? What evidence supports this condition? Which group member has not spoken yet? What assumption should the group test?
Answer-constrained guardrail	Redirects students away from outsourcing the task to the system.	What can your group decide first? What information is still missing? How could you restate the problem more clearly instead of asking for a complete response?

Appendix 2

Scoring Dimensions and Observation Indicators

Appendix 2 provides compact summaries of the analytic categories used to score problem-posing quality and dialogic engagement.

Table B1: Problem-posing quality rubric summary

Dimension	High-performance description
Contextual authenticity	The problem uses a recognisable, purposeful context with plausible details for the intended learner or audience.
Disciplinary coherence	The information is sufficient and consistent, the focal issue is clear, and the relations or assumptions support a meaningful task.
Cognitive demand	The problem invites reasoning, comparison, interpretation, or explanation rather than routine substitution.
Openness for inquiry or response	The task allows meaningful extension, alternative pathways, or follow-up questioning without becoming vague.

Table B2: Dialogic engagement observation scale summary

Indicator	Indicator of stronger dialogic engagement
Uptake of peer ideas	Students explicitly build on, refine, or challenge something previously said by a peer.
Use of justification	Students explain why a claim, relation, assumption, or revision is reasonable within the task.
Generative questioning	Students ask questions that open the task for clarification, extension, or testing rather than seeking a ready-made response.
Equitable contribution	Participation is distributed more evenly, and quieter members are invited into the discussion.