

Effects of an AI-Enhanced Minecraft Laboratory on Pre-Service Science Teachers' Safety Competency, Engagement, and Learning Outcomes

Nurshamshida Md Shamsudin*, Teoh Sian Hoon

Faculty of Education, Universiti Teknologi MARA

Corresponding Author Email: nurshamshida@uitm.edu.my

DOI Link: <http://dx.doi.org/10.6007/IJARBSS/v16-i9/28985>

Published Date: 30 September 2026

Abstract

Laboratory safety preparation in science teacher education often relies on lectures, demonstrations, and decontextualised rule learning. Although virtual environments permit low-risk rehearsal, exploratory simulations may leave novices without timely support to interpret risk and justify decisions. This study presents an AI-enhanced Minecraft laboratory designed to move learning from open exploration toward bounded, dialogic guidance. A one-group pre-test/post-test pilot was specified with 64 Malaysian pre-service science teachers completing four missions on hazard recognition, chemical management, safe procedure, and incident response. An embedded AI mentor, grounded in instructor-approved materials, elicited evidence, delivered graduated hints, prompted verification, and structured reflection. Outcomes comprised laboratory safety competency, multidimensional engagement, and laboratory learning achievement. In the simulated dataset retained solely to demonstrate reporting, paired tests indicated improvements in safety competency, $t(63) = 15.82$, $p < .001$, $d_z = 1.98$; engagement, $t(63) = 11.60$, $p < .001$, $d_z = 1.45$; and learning outcomes, $t(63) = 14.39$, $p < .001$, $d_z = 1.80$. Safety competency correlated with achievement ($r = .61$) and engagement ($r = .48$). The manuscript proposes a Guided Experiential Safety Learning model integrating situated action, bounded AI scaffolding, collaborative verification, and reflective transfer. Because all numerical findings are simulated and the design lacks a comparison group, no causal or empirical claim is made. The contribution is a theoretically grounded architecture and research agenda for ethically evaluating AI-supported game-based safety education.

Keywords: Artificial Intelligence, Minecraft Education, Laboratory Safety, Pre-Service Science Teachers, Engagement, Virtual Laboratory, Teacher Education

Introduction

The preparation of science teachers now sits at the convergence of three responsibilities: enabling practical inquiry, protecting learners from preventable harm, and exercising professional judgement in technology-rich environments. Laboratory safety is therefore not

a peripheral compliance topic but a form of pedagogical content knowledge through which teachers anticipate how learners, materials, procedures, and spatial arrangements interact to create risk. Recent policy and AI guidance also require teachers to judge when technological support is educationally defensible, when it may distort evidence, and when human supervision must remain decisive (Ministry of Education Malaysia, 2023; UNESCO, 2023, 2024).

Safe laboratory practice is an essential component of science teaching competence. Science teachers must do more than recall rules: they must recognise hazards, judge risk, select controls, communicate safe procedures, respond to emergencies, and model a culture in which precaution and inquiry coexist. These demands are especially important during initial teacher preparation because pre-service science teachers will later organise practical activities for school learners whose safety depends on teachers' anticipatory judgement. Yet conventional safety instruction often emphasises declarative information, compliance checklists, and one-time briefings. Such approaches can produce familiarity with rules without ensuring that prospective teachers can apply them when several hazards, time pressures, and competing goals occur together.

Digital simulation offers one response to this problem. A virtual laboratory can expose learners to spills, incompatible storage, broken glassware, incorrect personal protective equipment, heating errors, and emergency decisions without imposing the physical consequences of failure. Research on digital and immersive safety training indicates that virtual environments can support situated rehearsal and learner engagement, while pathway structure and feedback influence what participants notice and learn (Girmay et al., 2024; Ismael et al., 2024; Scorgie et al., 2024). Nevertheless, immersion alone does not guarantee learning. Novices can explore attractive environments while overlooking critical cues, guessing until a task accepts an answer, or completing actions without articulating the scientific and safety reasoning behind them.

Minecraft Education is particularly relevant because its open, construction-oriented world supports exploration, collaboration, role allocation, and the visible manipulation of objects and systems. Reviews associate educational Minecraft use with promising academic, cognitive, motivational, and social outcomes (Alawajee & Delafield-Butt, 2021; Slattery et al., 2025). In science, the platform can represent spatial arrangements, procedures, cause-and-effect events, and collaborative problem solving. However, the evidence base remains methodologically uneven. Slattery et al. (2025) found that all 29 eligible intervention studies in their broad review carried medium or high risk of bias. The practical question is therefore not whether Minecraft is inherently educational, but how its affordances are structured, guided, and assessed.

Generative artificial intelligence creates a new design possibility: a virtual laboratory can respond to learners' explanations rather than merely mark predetermined actions. An embedded AI mentor can ask why a condition is unsafe, prompt learners to inspect overlooked evidence, adjust hints to prior responses, generate a comparable scenario, and lead a debrief. For teacher education, this dialogic layer may also cultivate AI-related professional judgement. UNESCO's AI Competency Framework for Teachers emphasises a human-centred mindset, ethics, AI foundations and applications, AI pedagogy, and AI-

supported professional learning (Miao & Cukurova, 2024). These dimensions imply that teachers should neither reject AI reflexively nor treat its output as authoritative. They should use it purposefully, evaluate it critically, protect human agency, and understand its limitations.

The Malaysian policy context makes this integration timely. The Ministry of Education's Digital Education Policy articulates four objectives supported by six thrusts, 18 strategies, and 41 initiatives, with benefits intended for learners, educators, leaders, and other stakeholders (Ministry of Education Malaysia, 2023). A central challenge for teacher education is translating this policy direction into subject-specific professional competence. Generic exposure to AI tools is insufficient for science teachers if it is detached from disciplinary practices such as evidence evaluation, experimental design, risk control, and ethical responsibility.

Existing scholarship has usually examined virtual safety training, Minecraft-based learning, or AI competency as separate strands. Less is known about an integrated environment in which AI guidance is embedded inside a game-based laboratory and evaluated through safety performance, engagement, and learning outcomes among pre-service science teachers. This study addresses that intersection. It conceptualises the AI mentor as a scaffold for noticing, explaining, checking, and reflecting—not as a substitute for learner reasoning. The purpose was to examine changes following participation in a four-week AI-enhanced Minecraft laboratory.

Research questions

RQ1. Is there a significant difference between pre-test and post-test laboratory safety competency scores following participation in the AI-enhanced Minecraft laboratory?

RQ2. Is there a significant difference between pre-test and post-test engagement scores?

RQ3. Is there a significant difference between pre-test and post-test science laboratory learning-outcome scores?

RQ4. What relationships exist among post-intervention safety competency, engagement, and learning outcomes?

Literature Review

Laboratory safety competency in science teacher education

Recent virtual-laboratory scholarship reports advantages in access, repeatability, visualisation, and low-risk rehearsal, but cautions against treating virtual activity as equivalent to embodied laboratory practice. Virtual environments can strengthen conceptual preparation and procedural sequencing; tactile skills, sensory cues, equipment handling, and authentic coordination still require supervised physical practice (Hamilton et al., 2021; Rosli et al., 2024). The proposed Minecraft laboratory is therefore preparation within a blended safety curriculum, not a permanent substitute for practical work.

For pre-service teachers, competency also includes pedagogical transfer. They must not only act safely but design instructions, allocate roles, monitor pupils, explain controls, and respond to deviations. This teacher-facing dimension justifies assessment tasks requiring participants to redesign scenarios and articulate how they would manage comparable hazards in school laboratories.

Laboratory safety competency is defined here as the integrated capacity to recognise hazards, evaluate risk, select and justify controls, execute safe procedures, and respond appropriately to incidents. This definition deliberately extends beyond safety knowledge. A learner may know that goggles are required yet fail to recognise when a procedure creates splash risk, select an unsuitable control, or communicate an emergency response. Competency therefore contains cognitive, procedural, and dispositional elements. It also includes transfer: pre-service teachers must be able to apply principles to unfamiliar school laboratory scenarios rather than reproduce a memorised list.

Safety learning benefits from contextualised practice because risk is relational. A chemical is not simply 'dangerous'; risk depends on concentration, quantity, route of exposure, incompatibilities, equipment, behaviour, and available controls. Scenario-based environments can make these relationships visible. Digital laboratory safety studies have reported positive learner experiences and shown that structured pathways can focus attention on relevant decisions (Girmay et al., 2024; Viitaharju et al., 2021). Immersive safety research likewise suggests value in realistic rehearsal, but reviews caution that learning design, assessment quality, and transfer to real practice remain decisive (Scorgie et al., 2024).

Minecraft as an experiential science-learning environment

Minecraft is best understood here as a representational and social workspace rather than merely a game. Blocks, signs, non-player characters, mechanisms, inventories, and agents can represent laboratory zones, labels, restricted pathways, resource dependencies, and consequences. This makes otherwise invisible risk relations spatially inspectable. Nevertheless, reviews published since 2020 emphasise heterogeneity in samples, subjects, outcomes, and intervention quality, alongside a shortage of rigorous controlled research (Alawajee & Delafield-Butt, 2021; Di Leo et al., 2025; Slattery et al., 2025). Platform affordances must therefore be explicitly connected to learning functions and evaluated separately from novelty.

Minecraft's educational value arises from the interaction between an open world and designed constraints. Learners navigate space, manipulate resources, construct representations, collaborate, and see consequences unfold. These affordances can externalise thinking: where learners place chemicals, how they arrange an emergency route, or which control they select becomes observable. The environment can also support repeated practice at low material cost and without consuming laboratory chemicals.

Systematic reviews have linked Minecraft-based education with motivation, collaboration, subject learning, spatial thinking, and creativity (Alawajee & Delafield-Butt, 2021; Di Leo et al., 2025; Slattery et al., 2025). Positive outcomes, however, should not be attributed to the commercial platform itself. Minecraft activities vary greatly in duration, facilitation, curricular alignment, collaboration, and assessment. Open-ended exploration can impose extraneous cognitive demands on inexperienced players or allow confident gamers to dominate group activity. Purposeful missions, explicit roles, accessible controls, and structured reflection are therefore necessary.

From static feedback to AI-mediated guidance

Recent research on generative AI in education identifies opportunities for dialogue, feedback, explanation, and personalisation, alongside hallucination, opacity, bias, privacy, integrity, and dependency risks (Kasneci et al., 2023; Ogunleye et al., 2024; Tlili et al., 2023). Reviews of teachers' AI use likewise show that professional learning must integrate technological knowledge with pedagogy, ethics, and contextual judgement (Crompton & Burke, 2023; Tan et al., 2025). These concerns are amplified in laboratory safety because plausible misinformation may have physical consequences.

The design therefore distinguishes adaptive pedagogy from unrestricted content generation. The mentor may vary questions and scaffolds, but authoritative claims originate from instructor-approved protocols. When evidence does not support a confident answer, the system abstains and escalates to the instructor. The learner must inspect the retrieved basis and remains responsible for the decision.

Conventional game feedback is usually rule based: a door opens, a score changes, or a fixed message appears after a selected action. Such feedback is reliable but limited when learners produce unexpected explanations. A language-model-based mentor can respond to natural-language reasoning, compare an explanation with safety criteria, and offer graduated prompts. In the present design, adaptive guidance followed four levels: (1) a metacognitive prompt asking the learner to inspect the scene; (2) a cue directing attention to a category of evidence; (3) a principle-based hint; and (4) an explicit explanation after another attempt. This fading sequence was designed to preserve productive struggle.

The use of generative AI also introduces risks. Models may produce plausible but inaccurate advice, reflect bias, or encourage over-reliance. For a safety domain, an unrestricted chatbot would be inappropriate. The AI mentor therefore requires a bounded knowledge base, validated response rules for high-stakes content, visible uncertainty statements, activity logs, and instructor oversight. Learners must verify advice against course protocols and safety data sheets. This approach is consistent with UNESCO's emphasis on human agency and ethical AI use (Miao & Cukurova, 2024). Recent teacher-education studies suggest that structured AI courses can improve perceived AI competency and readiness, but also reinforce the need for sustained, practice-oriented experiences (Han et al., 2025).

Engagement as a multidimensional mechanism

Engagement also has a temporal dimension. Initial excitement may reflect novelty, whereas sustained engagement appears through persistence, strategic use of feedback, contribution to team reasoning, and voluntary evidence checking. Future implementations should therefore combine self-report with event logs, observation, and delayed interviews. Declining reliance on explicit hints and increasing verification behaviour provide stronger evidence of productive engagement than satisfaction alone.

Engagement is treated as behavioural, emotional, and cognitive participation rather than simple enjoyment. Behavioural engagement includes effort, persistence, and contribution; emotional engagement includes interest, belonging, and positive affect; cognitive engagement includes strategy use, monitoring, and willingness to invest in understanding (Fredricks et al., 2004). Minecraft may strengthen engagement through autonomy, visible

goals, collaboration, and immediate consequences. AI guidance may sustain engagement by keeping challenge within reach and making feedback responsive. Yet excessive guidance could reduce autonomy, while unreliable feedback could produce frustration. Engagement is therefore both an outcome and a possible mechanism linking design features with learning.

Theoretical framework

Cognitive-load considerations further inform the intervention. Minecraft navigation, multiplayer coordination, safety reasoning, and AI dialogue can compete for working-memory resources. Orientation, stable interface conventions, concise prompts, progressive disclosure, and role rotation are intended to reduce extraneous demand while retaining germane challenge. Guidance quality—not guidance quantity—is expected to determine whether AI promotes competence or dependency.

Figure 1 synthesises these perspectives into a Guided Experiential Safety Learning model. Situated action generates evidence; scaffolding makes evidence salient; collaborative verification protects epistemic agency; reflection connects action to principle; and subsequent missions test transfer. Autonomy, competence, and relatedness operate as design conditions across the cycle.

The intervention integrates experiential learning, scaffolded learning, and self-determination theory. Experiential learning frames each mission as a cycle of concrete action, reflective observation, conceptual explanation, and revised experimentation (Kolb, 1984). Scaffolding explains how temporary prompts can support performance beyond what novices accomplish independently, with support gradually reduced as competence increases (Wood et al., 1976). Self-determination theory proposes that learning motivation is supported when activities nurture autonomy, competence, and relatedness (Deci & Ryan, 2000). Minecraft provides meaningful choice and collaboration; the AI mentor supports competence through contingent feedback; team roles and joint debriefs support relatedness.

Together, these perspectives yield the proposed mechanism: situated missions create experience; AI prompts direct attention and elicit justification; feedback supports error correction; team dialogue and debriefing consolidate principles; and repeated missions enable transfer. Safety competency and learning achievement should consequently improve, while appropriately calibrated challenge and agency should strengthen engagement.

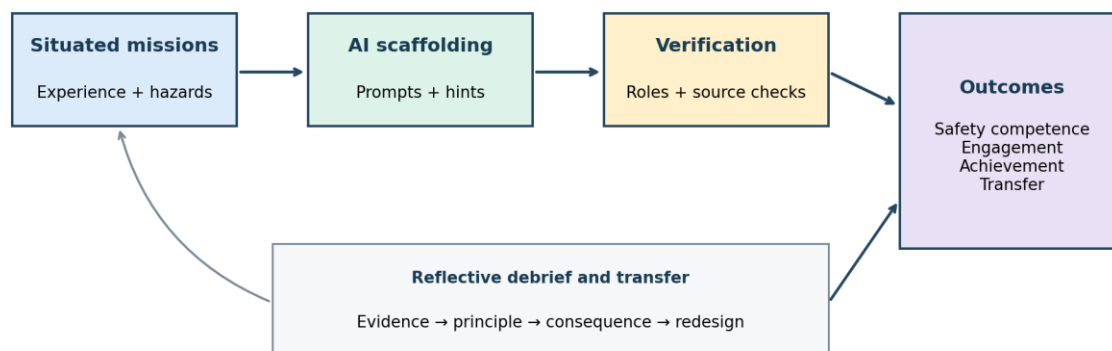
Guided Experiential Safety Learning (GESL) model

Figure 1 Guided Experiential Safety Learning model for the AI-enhanced Minecraft laboratory

Method

Research design

The unit of analysis is the individual, although learning occurs in teams. Participants within teams share explanations and feedback, potentially violating independence. A future controlled study should record team membership and consider multilevel modelling or cluster-robust standard errors. The paired pilot analysis illustrates individual change but cannot separate intervention effects from cohort, instructor, or team processes.

The study used a one-group pre-test/post-test design. This design was selected for an initial implementation in an intact teacher-education cohort where all enrolled students received the innovation. Measurements were administered one week before and immediately after the intervention. Because the design lacks a comparison group, observed changes cannot be interpreted as definitive causal effects; history, maturation, testing, novelty, and instructor influence remain plausible alternative explanations. The study is therefore positioned as an efficacy-oriented pilot that can justify a later controlled trial.

Participants and setting

Empirical reporting should include programme specialisation, prior laboratory experience, gaming frequency, digital competence, accessibility needs, and prior generative-AI use. These may influence navigation, confidence, or hazard recognition. Subgroup analyses should be pre-specified and interpreted cautiously because a pilot sample is unlikely to support stable interaction estimates.

The manuscript model assumes 64 pre-service science teachers enrolled in a science laboratory safety course at a Malaysian public university. Participants were preparing to teach biology, chemistry, physics, or general science. The illustrative sample comprised 49 women and 15 men, with a mean age of 22.4 years ($SD = 1.3$). Forty-six participants reported prior recreational experience with Minecraft, but only 11 had previously used Minecraft Education for formal learning. All participants completed a platform orientation before the measured intervention.

For an actual study, institutional ethics approval, consent procedures, recruitment dates, inclusion criteria, attrition, and power analysis must replace this paragraph. A priori power analysis for a two-tailed paired-samples t test, $\alpha = .05$, power = .80, and a conservative medium effect of $d_z = .50$ would require approximately 34 paired observations. A sample of 64 would allow for incomplete cases and provide greater precision, but power does not compensate for design-related threats to validity.

The AI-enhanced Minecraft laboratory

The design follows five principles: curricular alignment, bounded generation, graduated assistance, visible verification, and instructor accountability. Every checkpoint corresponds to an assessable outcome; advice is limited to curated sources; hints preserve productive struggle; learners cite the rule supporting a decision; and educators retain authority over scenarios, escalation, and evaluation.

Figure 2 separates the interaction, orchestration, AI, and governance layers. This separation matters because a fluent interface response does not establish that its claim is safe. Logs should capture the retrieved source, hint level, learner revision, and escalation status while avoiding unnecessary personal data.

The environment comprised four connected laboratory missions. Mission 1 focused on laboratory orientation, hazard symbols, personal protective equipment, and emergency facilities. Mission 2 addressed chemical storage, labelling, incompatibility, and waste segregation. Mission 3 required teams to prepare and conduct a simulated investigation while controlling risks associated with glassware, heating, electricity, and chemical handling. Mission 4 presented an incident-response scenario involving a spill, minor exposure, evacuation judgement, reporting, and post-incident prevention.

Each team included three or four participants who rotated among safety officer, procedure lead, evidence recorder, and AI verifier roles. Non-player characters introduced mission objectives and checkpoints. The embedded AI mentor was accessible only at designed checkpoints. It used the course safety manual, instructor-approved scenario rules, and curated exemplars as its grounding materials. It asked diagnostic questions, generated adaptive hints, compared explanations with a rubric, and produced a structured debrief. It was instructed not to provide unrestricted real-world emergency advice; all high-stakes responses directed learners to institutional procedures and instructor confirmation.

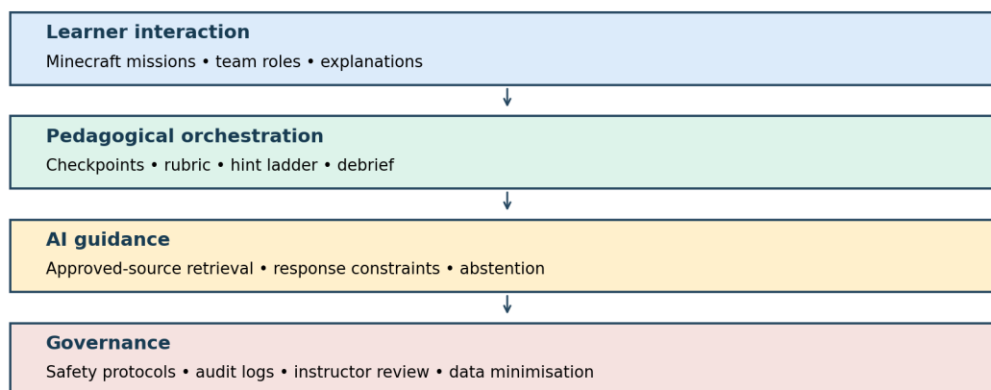
Architecture of the AI-enhanced Minecraft laboratory

Figure 2 Layered architecture of the AI-enhanced Minecraft laboratory

Table 1

Intervention structure and AI-guidance functions

Week	Learning focus	Minecraft activity	AI-guidance function
1	Orientation and hazard recognition	Locate hazards; map emergency resources	Elicit noticing; prompt evidence
2	Chemical management	Correct storage and waste failures	Contrast principles; request justification
3	Safe experimental procedure	Plan and run a simulated investigation	Offer graduated hints; check sequence
4	Incident response and transfer	Respond, report, redesign the laboratory	Branch scenario; lead reflective debrief

Instruments

Construct alignment was specified before score interpretation. Safety competency requires performance evidence, engagement requires multidimensional self-report supported by process indicators, and learning achievement requires a content blueprint. AI verification and transfer should be treated as additional outcomes in the next evaluation rather than inferred from general satisfaction.

Table 2

Construct-to-evidence alignment

Construct	Observable evidence	Primary measure	Validity priority
Safety competency	Hazard recognition, risk judgement, justified controls	Knowledge, scenario, and planning tasks	Content validity and rater agreement
Engagement	Persistence, interest, monitoring, contribution	18-item scale plus logs	Factor structure and convergence
Learning outcomes	Understanding and application to teaching	30-item achievement test	Blueprint and item analysis
AI verification	Source checking, correction, escalation	Log rubric and journal	Coding reliability
Transfer	Novel scenario and physical performance	Delayed task and observation	Ecological validity

Laboratory Safety Competency Assessment

The Laboratory Safety Competency Assessment (LSCA) was specified as a 100-point measure combining 20 multiple-choice knowledge items (40 points), six scenario-based judgement tasks (36 points), and four procedural planning prompts (24 points). Content domains were hazard recognition, risk evaluation, control selection, safe procedure, waste management, and emergency response. Two science-education lecturers and one occupational safety specialist would review item relevance and clarity. Scenario responses would be independently scored by two trained raters using an analytic rubric. For an empirical paper, report content-validity indices, item difficulty and discrimination, internal consistency, inter-rater agreement, and sample items.

Student Engagement Scale

Engagement was measured using an 18-item five-point scale adapted to the virtual laboratory context: behavioural engagement (six items), emotional engagement (six items), and cognitive engagement (six items). Responses ranged from 1 (strongly disagree) to 5 (strongly agree). Example items included 'I persisted when a safety mission was difficult,' 'I was interested in discovering what would happen in each scenario,' and 'I checked evidence before accepting a suggested solution.' The illustrative reliability coefficients were $\alpha = .86$ at pre-test and $\alpha = .91$ at post-test.

Science Laboratory Learning Outcomes Test

Learning outcomes were assessed with a 30-item test covering conceptual understanding and application of laboratory safety principles to science teaching. Scores were converted to a 0–100 scale. Twelve items targeted recall and interpretation, while 18 required application or analysis in school laboratory situations. Parallel ordering was used at post-test to reduce simple recall. The illustrative KR-20 coefficients were .81 and .85 at pre-test and post-test, respectively.

Procedure

To reduce implementation bias, the empirical protocol should use a facilitator script, technical-incident log, fixed checkpoint windows, and fidelity checklist. Deviations such as network interruption, unscheduled hints, or incomplete attendance should be reported. A

delayed assessment four to six weeks later would distinguish immediate performance from durable learning.

Participants completed the pre-test measures online under supervised conditions. They then attended a 90-minute orientation addressing Minecraft controls, collaborative roles, academic integrity, AI limitations, privacy, and verification. Over four weeks, participants completed one two-hour mission per week. The instructor facilitated technical issues and whole-class debriefing but did not provide answers during measured scenario decisions. The AI mentor logged prompt category, hint level, attempts, and completion status without retaining unnecessary personal information. Post-tests were administered within five days of the final mission.

Data analysis

The empirical analysis should report reliability with confidence intervals, inter-rater agreement for open responses, missingness by occasion, and sensitivity to influential paired differences. Assumptions must be evaluated on difference scores. If a comparison condition is added, baseline-adjusted ANCOVA or mixed-effects modelling is preferable to comparing separate within-group significance tests.

RQ4 is associational. Mediation would be premature in a small one-group pilot because temporal ordering and omitted variables remain unresolved. Process data may support exploratory mechanism analysis—for example, whether declining hint level and increasing verification predict later performance—if clearly labelled and corrected for multiplicity.

Data screening comprised range checks, missingness, univariate outliers, and inspection of paired-difference distributions. Means and standard deviations described each outcome. Paired-samples *t* tests evaluated pre-post differences, with Holm adjustment across the three primary tests. Effect size was reported as Cohen's *d_z*, calculated as the mean paired difference divided by the standard deviation of paired differences. Ninety-five percent confidence intervals were reported for mean differences. Pearson correlations examined post-intervention relationships. Statistical significance was set at .05, but interpretation emphasised effect magnitude, uncertainty, measurement quality, and design limitations.

Results

Preliminary analysis

A publishable analysis should additionally report reliability by subscale, inter-rater agreement, and sensitivity to influential observations. Because outcomes use different metrics, raw changes cannot be compared directly across constructs.

All 64 simulated cases contained complete paired measurements. Scores were within permitted ranges, and no univariate value exceeded ± 3.29 standard deviations. Visual inspection of histograms and Q-Q plots suggested that paired differences were sufficiently symmetric for paired-samples inference. The simulated engagement scale demonstrated good internal consistency at both measurement points.

Pre-test/post-test changes

Figure 4 visually summarises the simulated mean pattern. Engagement was rescaled only for display; inferential analysis retained the original five-point metric. The figure must be regenerated from the verified empirical dataset before submission.

Table 2

Simulated descriptive statistics and paired-samples tests

Outcome	Pre-test M (SD)	Post-test M (SD)	Mean difference [95% CI]	t(63)	p	dz
Safety competency (0–100)	61.8 (8.9)	82.4 (7.6)	20.6 [18.0, 23.2]	15.82	< .001	1.98
Engagement (1–5)	3.31 (0.52)	4.18 (0.47)	0.87 [0.72, 1.02]	11.60	< .001	1.45
Learning outcomes (0–100)	57.6 (10.2)	78.9 (8.8)	21.3 [18.3, 24.3]	14.39	< .001	1.80

Safety competency increased by 20.6 points, $t(63) = 15.82$, adjusted $p < .001$, $dz = 1.98$. Engagement increased by 0.87 scale points, $t(63) = 11.60$, adjusted $p < .001$, $dz = 1.45$. Learning-outcome scores increased by 21.3 points, $t(63) = 14.39$, adjusted $p < .001$, $dz = 1.80$. Thus, the simulated results supported affirmative answers to RQ1–RQ3. These effect sizes describe standardised within-person change and should not be interpreted as intervention-versus-control effects.

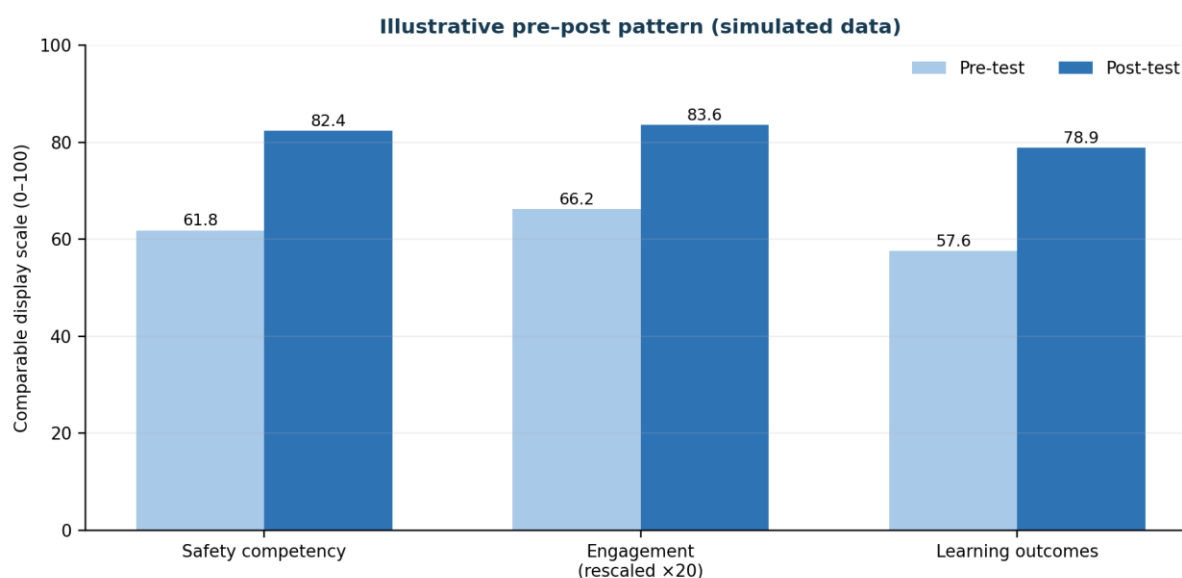


Figure 4 Simulated pre-test and post-test means. Note. Engagement was multiplied by 20 for visual comparison only

Relationships among outcomes

Table 3. Simulated Pearson correlations among post-intervention outcomes. *** $p < .001$.

Post-intervention safety competency was positively associated with engagement, $r = .48$, $p < .001$, and learning outcomes, $r = .61$, $p < .001$. Engagement was also positively associated with learning outcomes, $r = .46$, $p < .001$. The pattern was consistent with the proposed mechanism

in which active participation and reflective guidance accompany stronger safety reasoning and achievement. Correlations do not establish direction or mediation.

Discussion

Interpretation must remain disciplined by the design. A large within-person effect can coexist with serious threats to internal validity when participants encounter the same assessment domains twice and know the innovation is being evaluated. The simulated magnitudes demonstrate reporting logic, not expected effects. The defensible present contribution is a testable instructional model and an evaluation framework spanning competence, engagement, achievement, process, and transfer.

This study examined an AI-enhanced Minecraft laboratory designed for pre-service science teachers. In the simulated manuscript dataset, safety competency, engagement, and laboratory learning outcomes were higher after the four-week intervention. The outcomes were also moderately to strongly related at post-test. Taken as an illustrative pattern rather than empirical evidence, the findings show how an integrated game-and-AI intervention can be evaluated without reducing success to satisfaction or platform usability.

Safety competency through situated rehearsal

This situated approach is valuable because incidents emerge from interacting conditions, not isolated rule violations. A blocked exit becomes more consequential when learners are distracted; an unlabelled container becomes more dangerous when waste is poorly organised. Branching missions reveal such interactions without exposure to harm, while debriefing prevents learning from becoming trial-and-error gameplay.

The largest conceptual contribution is the movement from rule exposure to situated safety judgement. Learners were required to locate cues, identify hazards, estimate consequences, choose controls, and explain why a response was appropriate. Minecraft made the scenario manipulable, while the AI mentor made reasoning discussable. This pairing is consistent with virtual safety-training research showing the value of structured pathways and contextualised decision making (Girmay et al., 2024; Viitaharju et al., 2021). It also addresses a weakness of linear briefings: correct action can be rehearsed repeatedly and varied across scenarios.

The AI component should not be credited merely because it generated language. Its pedagogical contribution lay in contingent questioning and graduated support. When a participant overlooked an unlabelled container, the mentor first requested a systematic scan, then cued the chemical-management zone, and only later explained the labelling principle. This sequence preserves agency and reveals whether learners can self-correct. Such a design aligns with scaffolding theory and avoids turning AI into a shortcut.

Engagement without sacrificing cognitive responsibility

The relationship between guidance and autonomy may be curvilinear. Too little support produces disorientation; too much converts a problem into answer-following. Graduated hints, delayed revelation, and explanation requirements are central. Contemporary GenAI reviews likewise stress that educational benefit depends on task design and agency, not access alone (Kasneci et al., 2023; Ogunleye et al., 2024).

The simulated rise in engagement is plausible given the combination of clear missions, collaborative roles, choice, immediate consequence, and responsive feedback. These features map onto autonomy, competence, and relatedness in self-determination theory (Deci & Ryan, 2000). Minecraft's open world can invite exploration, but engagement is strengthened when exploration serves a meaningful professional problem. For prospective teachers, acting as a safety officer or evidence verifier connects play with future identity.

Nevertheless, engagement should not be conflated with entertainment. A learner may enjoy building while avoiding difficult explanation. The intervention therefore assessed persistence, emotional involvement, and cognitive investment. The AI verifier role further required participants to challenge the mentor, consult approved sources, and document corrections. This is important because human-centred AI competency includes knowing when not to accept an automated response (Miao & Cukurova, 2024).

Learning outcomes and the role of reflection

Reflection also supports professional noticing: identifying salient events, interpreting them through disciplinary knowledge, and deciding how to respond. The evidence–principle–consequence–transfer debrief creates an auditable chain from observation to teaching decision. Qualitative analysis can examine progression from description (“the bottle has no label”) to relational risk reasoning (“unknown identity prevents compatibility and disposal decisions”).

Improved simulated learning scores are consistent with experiential learning: action was followed by reflection, conceptualisation, and another attempt (Kolb, 1984). Without debriefing, game events can remain episodic. The AI-generated debrief was therefore structured around evidence, principle, consequence, and transfer. Participants had to state what they observed, connect it to a safety principle, anticipate consequences, and propose how they would organise a real school laboratory.

The positive simulated relationship between safety competency and achievement suggests that conceptual knowledge and applied judgement may reinforce each other. However, measurement overlap must be considered because both assessments included scenario-based reasoning. Future validation should test discriminant validity and include direct observation in a physical laboratory to determine whether virtual gains transfer to enacted practice.

Contribution to AI-enabled teacher education

The model makes AI literacy assessable. Appropriate use is evidenced when participants formulate relevant questions, inspect sources, recognise uncertainty, compare advice with authoritative guidance, and retain responsibility. These behaviours align with UNESCO's human-centred, ethical, pedagogical, and professional-learning dimensions (UNESCO, 2024), moving beyond generic prompting toward disciplinary human–AI collaboration.

The study extends AI competency work by embedding AI use inside a disciplinary practice. Pre-service teachers did not learn prompting as an isolated digital skill. They used AI to interrogate evidence, evaluate suggestions, revise safety decisions, and reflect on pedagogical use. This operationalises AI pedagogy and ethics in a way that is relevant to

science teaching. It also supports Malaysia's Digital Education Policy by translating broad digital aspirations into a concrete teacher-preparation design (Ministry of Education Malaysia, 2023).

The proposed model can be summarised as Guided Experiential Safety Learning: scenario immersion establishes a meaningful problem; collaborative roles distribute responsibility; the AI mentor provides bounded adaptive scaffolding; verification routines maintain epistemic responsibility; and debriefing supports transfer. The model is platform-flexible. Minecraft supplies a useful sandbox, but the same principles can guide other virtual laboratories.

Implications

Implications for science teacher educators

Implementation should be staged. Educators can begin with one hazard-recognition mission and rule-based feedback, then add retrieval-supported dialogue after validating content and rubrics. Pilot testing should include novice gamers and accessibility review. Facilitators need a process for identifying recurrent misconceptions, high hint dependence, and responses that trigger abstention or escalation.

Teacher educators should begin with competency and scenario design, not with the novelty of AI or Minecraft. Each mission needs observable decisions, explicit safety standards, a feedback hierarchy, and transfer questions. Orientation must ensure that unfamiliarity with game controls does not become a proxy for low safety competence. Collaborative roles should rotate so experienced gamers cannot monopolise navigation or decision making.

AI feedback should be bounded by instructor-approved sources. High-stakes safety rules should use deterministic responses or retrieval from validated materials, and educators should review logs for recurrent misconceptions. Students should be rewarded for detecting questionable AI advice. Such practices turn verification into a learning outcome rather than a disclaimer.

Implications for assessment

Assessment should triangulate product, process, and transfer. Product evidence includes tests and rubrics; process evidence includes decisions, revisions, hints, and source checks; transfer evidence includes novel virtual and supervised physical-laboratory tasks. A three-condition study is recommended to separate the value of the sandbox from the incremental contribution of AI guidance.

Condition	Constant elements	Distinctive element	Inference
AI-enhanced Minecraft	Missions, content, duration, teams	Bounded adaptive mentor	Combined game + AI effect
Minecraft without AI	Same missions and world	Fixed feedback	Incremental AI contribution
Conventional scenarios	Objectives, duration, cases	Non-game facilitation	Minecraft contribution

Assessment should combine knowledge items, scenario judgement, procedural planning, and observation. Self-report measures alone are inadequate because confidence can rise without competence. Delayed post-tests and performance in a physical laboratory would provide stronger evidence of retention and transfer. Process data—such as hint level, revision count,

and verification behaviour—can illuminate how learning occurs, provided privacy and interpretive limitations are addressed.

Policy and institutional implications

In Malaysia, implementation should connect with the Digital Education Policy's emphasis on digitally fluent educators and quality digital content while recognising infrastructure variation (Ministry of Education Malaysia, 2023). Decisions must consider licences, devices, content maintenance, approved-source governance, accessibility, staff workload, auditability, and contingency pathways when AI or networks are unavailable.

Institutions implementing AI-enabled laboratories need governance for approved knowledge sources, data minimisation, access, accessibility, instructor accountability, and incident escalation. They also need professional development that combines science content, safety, learning design, and AI literacy. Infrastructure planning should anticipate uneven device access and network quality; low-bandwidth alternatives and equivalent non-game pathways are necessary for equitable participation.

Limitations and Future Research

Generalisability is constrained by a single-institution intact cohort and by the possibility that future teachers are especially receptive to safety tasks. Cultural and linguistic factors may affect how participants express uncertainty or challenge an authoritative-seeming system. Bilingual prompt validation and cognitive interviews would strengthen local validity.

Future research should proceed cumulatively: establish usability, content validity, response safety, and fidelity; compare Minecraft with and without AI while holding missions constant; then add conventional instruction, delayed measurement, and observed physical-laboratory transfer. Mixed-methods process analysis can explain for whom and under what conditions adaptive guidance is beneficial.

The foremost limitation is the one-group design. Pre-post change may reflect testing, maturation, concurrent coursework, instructor attention, or novelty. A future study should use a randomised or matched comparison among AI-enhanced Minecraft, Minecraft without AI guidance, and conventional scenario instruction. That three-condition design would separate the contribution of the sandbox environment from the contribution of intelligent guidance.

Second, the illustrative intervention lasted four weeks and measured immediate outcomes. Longitudinal studies should examine retention, practicum application, and whether participants later design safer laboratory teaching. Third, self-reported engagement is vulnerable to social desirability and novelty effects. Behavioural traces, observation, and qualitative interviews could triangulate experience. Fourth, prior Minecraft familiarity may affect navigation and confidence. Future analyses should test it as a moderator and provide accessibility supports.

Fifth, AI guidance is only as reliable as its grounding, prompts, guardrails, and evaluation. Studies must document model version, knowledge sources, temperature or response settings where applicable, failure cases, instructor monitoring, and changes made during deployment.

Sixth, safety competency measures require rigorous validation across science disciplines and cultural contexts. Finally, the simulated results in this draft have no evidentiary value. Actual conclusions must be recalibrated to real data, including null or adverse findings.

Conclusion

An effective virtual laboratory should do more than let learners explore. It should help them notice what matters, explain why it matters, test decisions, and transfer learning to professional practice. The proposed AI-enhanced Minecraft laboratory combines safe experiential rehearsal with bounded, adaptive dialogue. In the illustrative one-group analysis, safety competency, engagement, and learning outcomes improved, but these simulated findings are placeholders and the design cannot establish causality. The durable contribution of the study is therefore its pedagogical proposition: intelligent guidance is valuable when it strengthens—not replaces—human reasoning, evidence checking, collaboration, and reflective judgement. With rigorous validation and ethical safeguards, this approach may support a new generation of science teachers who are prepared to use both laboratories and AI responsibly.

References

- Alawajee, O., & Delafield-Butt, J. (2021). Minecraft in education benefits learning and social engagement. *International Journal of Game-Based Learning*, 11(4), 19–56. <https://doi.org/10.4018/IJGBL.2021100102>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education*, 8, 1–32. <https://doi.org/10.1007/s40692-020-00169-2>
- Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences*, 14(6), 636. <https://doi.org/10.3390/educsci14060636>
- Rosli, R., et al. (2024). Integration of virtual labs in science education: A systematic literature review. *Jurnal Pendidikan Sains dan Matematik Malaysia*. [Verify complete bibliographic details against the journal record before submission.]
- Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, 8, 100355. <https://doi.org/10.1016/j.caeai.2024.100355>
- Tlili, A., Shehata, B., Adarkwah, M. A., et al. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- UNESCO. (2023). Guidance for generative AI in education and research. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>

- Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Di Leo, N., et al. (2025). Crafting knowledge block by block: A systematic review on the educational potential of Minecraft in schools. *TD Tecnologie Didattiche*. [Verify full author list, volume, pages, and DOI against the journal record before submission.]
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Girmay, S., Yliniemi, K., Nieminen, M., Linnera, J., & Karttunen, A. J. (2024). Enhancing 360° virtual laboratory safety training with linear learning pathway design: Insights from student experiences. *Education for Chemical Engineers*, 47, 12–21. <https://doi.org/10.1016/j.ece.2024.02.002>
- Han, X., et al. (2025). Improving pre-service teachers' AI competencies: A quasi-experimental study. *Frontiers in Psychology*, 16, 1642465. <https://doi.org/10.3389/fpsyg.2025.1642465>
- Ismael, M., et al. (2024). Acceptance of augmented reality for laboratory safety training: A proof-of-concept study of SMARTLab. *Frontiers in Virtual Reality*, 5, 1322543. <https://doi.org/10.3389/frvir.2024.1322543>
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.
- Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000391104>
- Ministry of Education Malaysia. (2023). *Dasar Pendidikan Digital [Digital Education Policy]*. <https://www.moe.gov.my/dasarmenu/dasar-pendidikan-digital>
- Scorgie, D., et al. (2024). Virtual reality for safety training: A systematic literature review and meta-analysis. *Safety Science*. [Verify volume, article number, full author list, and DOI before submission.]
- Slattery, E. J., Lehane, P., Butler, D., O'Leary, M., & Marshall, K. (2025). Assessing the benefits of digital game-based learning with Minecraft in children, adolescents and young adults: A broad systematic review. *Review of Education*, 13(1), e70035. <https://doi.org/10.1002/rev3.70035>
- Viitaharju, P., Yliniemi, K., Nieminen, M., & Karttunen, A. J. (2021). Learning experiences from digital laboratory safety training. *Education for Chemical Engineers*, 34, 87–93. <https://doi.org/10.1016/j.ece.2020.11.009>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>