Journal of Education for Sustainability and Transformation
Volume 1, Number 4, 2026.
E-ISSN: 3125-0130
https://journal.serral.org/jestra

Simulation Supported Problem Based Learning and Critical Thinking: A Systematic Review of Comparative Evidence and Causal Attribution

Evarius Heru Pambudi, I Made Astra, Esmar Budi, Firmanul Catur Wibowo
Universitas Negeri Jakarta, Indonesia

Article Info
Keywords:

Causal attribution
Critical thinking
Interactive simulation
Problem-based learning
Systematic review

*** Corresponding author:***
evarius_1310825012@mhs.unj.ac.id
Evarius Heru Pambudi Yuwono

Abstract

ABSTRACT
This systematic review examined evidence on problem-based learning (PBL) supported by interactive simulation for critical thinking. A Scopus search on 1 October 2026 identified 134 records. Metadata filters for year, document type, language, and open-access status reduced this set to 52 records. The reconstructed audit excluded 33 records at title/abstract screening and retained 19 reports for narrative synthesis. Eight reports primarily examined PBL combined with simulation, while six directly compared simulation-supported PBL with an alternative instructional condition. Four comparisons assessed critical thinking through performance based tests or essays, whereas two relied on self-report questionnaires. Findings generally favored simulation-supported PBL, but the strength of evidence varied across study designs and outcome measures. Nonrandom allocation, incomplete reporting, bundled intervention components, and reliance on perceived rather than demonstrated critical-thinking ability limited causal interpretation. Studies combining PBL, simulation, and other instructional supports could not isolate the contribution of simulation. The evidence did not establish an additional causal benefit of scaffolding when simulation was already embedded in PBL. These findings warrant caution; a pooled estimate was not appropriate. Future studies should compare PBL alone, PBL with simulation, and PBL with simulation plus scaffolding under matched conditions, use validated performance-based measures, and report implementation fidelity in detail.

Received: 4 September 2026; **Received in revised form** 18 September 2026**; Accepted:** 25 September 2026

To cite this article: Pambudi, E. H., Astra, I. M., Budi, E., Wibowo, F. (2026). Simulation Supported Problem Based Learning and Critical Thinking: A Systematic Review of Comparative Evidence and Causal Attribution. *Journal of Education for Sustainability and Transformation, 1*(4), 36-48.

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License ©2026 by author/s

INTRODUCTION

Critical thinking requires learners to interpret information, analyze evidence, evaluate explanations, draw inferences, explain conclusions, and regulate their reasoning. In science learning, these practices connect observations to explanatory models and require learners to judge whether evidence warrants a claim. The widely used Delphi framework identifies interpretation, analysis, evaluation, inference, explanation, and self regulation as interrelated dimensions of critical thinking (Facione, 1990). Recent syntheses report that instruction can support critical thinking, while also showing variation across interventions, measures, and learning settings (Liu & Pasztor, 2022; Yu & Zin, 2023).

Problem-based learning (PBL) is relevant because it begins with a problem that requires learners to identify what they know, determine information needs, investigate possible explanations, and justify a solution. These activities can make reasoning visible and provide opportunities to compare evidence rather than recall content alone. A meta analysis reported positive average effects of PBL interventions on critical thinking in higher education, but study

and instructional heterogeneity cautions against assuming that every PBL design yields the same outcome (Hmelo-Silver, 2004; Liu & Pasztor, 2022). A later review similarly documented varied adaptations of PBL intended to support critical thinking (Yu & Zin, 2023).

Interactive simulations may give learners a controlled environment for manipulating variables, observing dynamic representations, repeating trials, and comparing results. Reviews of physical and virtual experimentation report benefits for some science-learning outcomes, particularly conceptual knowledge, but those outcomes are not interchangeable with directly measured critical-thinking performance (Worner et al., 2022; Flegr et al., 2023; Muilwijk & Lazonder, 2023). The presence of a simulation therefore does not, by itself, establish that learners become better at evaluating evidence, making inferences, or explaining conclusions. The assessment instrument and the construct it measures must remain explicit.

PBL and simulations can also place demands on learners planning, monitoring, and evidence evaluation. Scaffolding such as guiding questions, conceptual prompts, worked examples, or metacognitive cues may help learners manage these demands, but its effects depend on the type and timing of support. Reviews of teacher-implemented scaffolding identify diverse practices and note that adaptation to learners needs is not always reported clearly (Dominguez & Svihla, 2023). Without a matched comparison, it is difficult to separate a scaffold's contribution from the PBL structure, simulation, teacher guidance, or additional instructional time (Hmelo-Silver et al., 2007; Dominguez & Svihla, 2023).

Existing reviews address related but distinct questions: PBL effects on critical thinking, adaptations of PBL, and learning outcomes from physical and virtual experiments (Liu & Pásztor, 2022; Yu & Zin, 2023; Worner et al., 2022; Muilwijk & Lazonder, 2023). A focused synthesis is needed to distinguish controlled evidence for simulation-supported PBL from within group change, self reported perceptions, and associative findings. Positive results from these designs do not carry equivalent evidential weight.

The novelty of this review lies in its assessment of the strength and limits of comparative evidence, rather than in a general claim that problem-based learning (PBL) or simulation improves critical thinking. Earlier syntheses have examined the effects of PBL on critical thinking, critical-thinking-oriented adaptations of PBL, or learning from physical and virtual experiments; these reviews address related but distinct outcomes and do not, by themselves, establish the independent contribution of simulation within PBL (Liu & Pasztor, 2022; Yu & Zin, 2023; Worner et al., 2022; Muilwijk & Lazonder, 2023). This review brings these strands together by distinguishing direct comparisons from single group pretest posttest and associative studies, separating performance-based critical-thinking measures from self reported perceptions, and examining whether study designs can distinguish the contributions of PBL, simulation, and scaffolding. Its contribution is a design-sensitive narrative synthesis of 19 reports, with focused analysis of six direct comparisons, that specifies what the evidence supports, what remains uncertain, and which component-separating comparisons future studies should test. The review does not estimate a pooled causal effect because the primary studies differ substantially in their populations, interventions, measures, and designs.

Building on this distinction, the review evaluates comparative evidence and causal attribution rather than proposing a phase-by-phase model of e-scaffolding. It addresses three questions: (RQ1) What designs, comparison conditions, and critical-thinking measures characterize empirical studies of PBL with simulation or closely related scaffolding support? (RQ2) What is the direction and consistency of critical-thinking findings in direct comparisons, considering design and risk of bias? (RQ3) To what extent do existing designs permit attribution to PBL, simulation, or scaffolding, and which research designs could better separate these contributions?

METHOD

The review used a systematic search and narrative synthesis. Reporting was informed by PRISMA 2020 and PRISMA-S, which emphasize transparent reporting of selection procedures and search strategies (Page et al., 2021; Rethlefsen et al., 2021). Eligible records were English language journal articles or conference papers published from 2020 through 2026, marked as open access in Scopus, and reporting original empirical data in formal or professional education.

Eligible evidence comprised (a) PBL combined with interactive, computer based, virtual, or laboratory simulation; (b) PBL with explicit instructional or conceptual scaffolding; or (c) empirical reports that clarified implementation mechanisms relevant to these combinations. Direct critical-thinking performance measures, self report measures, and qualitative or process measures were retained but analyzed as different evidence types. Studies in the scaffolding/process and contextual strata were not treated as direct estimates of the effect of simulation-supported PBL.

Scopus was searched on 1 October 2026. The search string was:

TITLE-ABS-KEY (("problem-based learning" OR "problem based learning" OR PBL) AND ("critical thinking" OR "critical thinking skill*" OR "critical thinking abil*" OR "critical thinking disposition*") AND (scaffold* OR "instructional scaffolding" OR "learning scaffolding" OR "conceptual scaffolding" OR "metacognitive scaffolding" OR "procedural scaffolding" OR "interactive simulation*" OR "computer simulation*" OR "computer-based simulation*" OR "virtual simulation*" OR "virtual laborator*" OR "virtual experiment*" OR "interactive virtual laborator*" OR PhET))

The export contained 134 records. Sequential metadata filters excluded 41 records outside 2020–2026, 17 records with ineligible document types, one non-English record, and 23 without an open-access designation, leaving 52 records. The record-level audit identified no duplicate DOI or Scopus EID in the exported fields. Of the 52 records, 33 were excluded at title/abstract screening, eight were secondary or non empirical publications and 25 did not match the eligible intervention/outcome combination. Nineteen reports were retrieved and retained for narrative synthesis; no further full-text exclusions were recorded in the reconstructed audit (Figure 1).

The selection log was reconstructed from the filtered Scopus export, abstracts, metadata, and accessible full text reports. The corresponding author reviewed the reconstructed 52 record decision log for this revision. The materials did not document independent duplicate screening, the number of original screeners, a formal adjudication log, or an inter rater statistic. The review therefore does not claim independent double screening. This limitation is reported rather than replaced with an undocumented procedure.

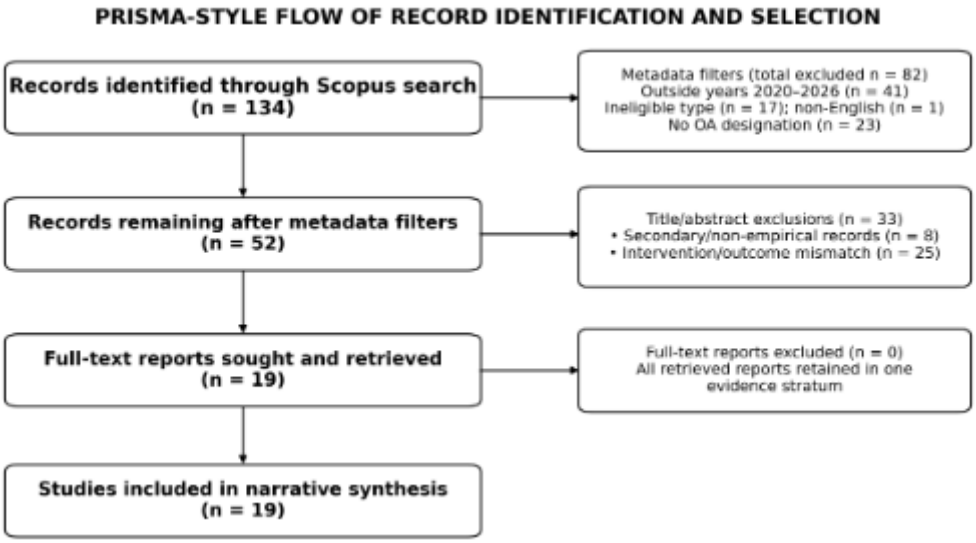


Figure 1. PRISMA style flow of record identification and selection

The extraction matrix recorded study context, sample, design, intervention package, comparator, critical-thinking instrument, reported statistics, direction of findings, and limitations relevant to causal interpretation. Studies were grouped by their dominant analytical contribution: PBL plus simulation (eight), PBL processes or scaffolding (eight), and contextual implementation (three). These strata describe the main use made of each report in this review; they do not imply that instructional components are mutually exclusive.

Risk of bias was assessed narratively by design specific domains rather than through one numerical scale. For randomized studies, the appraisal considered allocation, baseline balance, attrition, deviations from intervention, outcome measurement, and reporting. For quasi experimental studies, it considered comparator suitability, baseline equivalence, confounding, intervention fidelity, co interventions, outcome validity, missing data, and statistical reporting, informed by the revised JBI tool for quasi-experimental studies (Barker et al., 2024). For one-group pre/post designs, the lack of a counterfactual, history, maturation, testing effects, attrition, and measurement quality were central. Qualitative studies were considered for sampling, data richness, methodological congruity, transparent analysis, reflexivity, participant voice, and triangulation. For path analysis and structural-equation studies, assessment focused on construct measurement, self-report, confounding, model specification, temporal order, and the distinction between association and causation.

The appraisal was completed as a single audit pass from the available records, with AI assisted support for organization and preliminary classification. No independent second appraisal, disagreement resolution procedure, or inter rater statistic was documented. Judgements are therefore provisional narrative assessments, not a validated common rating across designs. The author team should verify each judgement against the primary report before submission.

A meta analysis was not conducted because populations, disciplines, intervention packages, comparators, instruments, and effect metrics differed. N gain, raw posttest scores, questionnaire totals, and qualitative outcomes could not be pooled as a common effect. The effectiveness synthesis therefore prioritizes six direct comparisons of simulation supported PBL with alternative instruction; other studies inform the evidence profile, possible mechanisms, and implementation context.

RESULTS AND DISCUSSION

The 19 retained reports spanned secondary, higher, and professional education, including physics, mathematics, teacher education, nursing, forensic science, urban studies, vocational education, computer science, design education, and postgraduate immunology. Ten studies were conducted in Indonesia; the remaining contexts included South Korea, Singapore, Spain, Switzerland, Sri Lanka, China, the United Kingdom, and Nigeria. This diversity provides breadth, but also limits direct comparison because instructional objectives and critical-thinking instruments were not standardized**.**

Eight studies primarily examined PBL with simulation: Ramadan et al. (2020), Simanjuntak et al. (2020), Serungke et al. (2020), Parno et al. (2021), Lee and Son (2021), Suhirman and Prayogi (2023), Lailiyah et al. (2026), and Xue et al. (2026). Tools included PhET, Adobe Flash, virtual laboratories, virtual simulation media, and mobile-supported simulation. Eight reports focused primarily on PBL processes or scaffolding, and three described contextual implementation. The profile and the principal claim each design can support are summarized in Table 1.

The 19 reports fell into three analytical groups: eight reports (42.1%) primarily examined PBL combined with simulation, eight (42.1%) focused on PBL processes or scaffolding, and three (15.8%) primarily described implementation context. Ten of the 19 reports were conducted in Indonesia; the remaining contexts included South Korea, Singapore, Spain, Switzerland, Sri Lanka, China, the United Kingdom, and Nigeria. The six direct comparisons involved samples

ranging from 50 to 105 participants. Two used randomized designs, while the other four used nonrandomized or less clearly reported comparison designs. Four comparisons assessed critical thinking with performance-based tests or essays; two used self-report questionnaires. The remaining 13 reports mainly contributed single-group pre/post, qualitative, process, or associative evidence and therefore cannot be interpreted as equivalent controlled estimates.

Table 1. Characteristics of the 19 included reports and limits of inference

Study	Context / sample	Design / intervention	Critical-thinking outcome and key limitation
Ramadan et al. (2020)	Grade 10 physics; Indonesia; N=60	PBL + PhET e-handout/LMS; two implementation classes	Performance CT test; N-gain 0.35 vs 0.29 across implementation classes; not an untreated control
Simanjuntak et al. (2020)	Grade 11 physics; Indonesia; N=84	Quasi-experimental control group; PBL + Adobe Flash vs traditional teaching	Performance CT test/essay; posttest 74.45 vs 53.95; N-gain 58% vs 28%; non random allocation
Serungke et al. (2020)	Grade 10 physics; Indonesia; N=60	Controlled comparison; PBL + virtual laboratory vs conventional teaching	CT test; N-gain 69.54 vs 40.44; reporting detail limits appraisal
Sukatiman et al. (2020)	Vocational Geomatics; Indonesia; N=8	Qualitative case; scaffolding-supported PBL + ICT	Process/HOTS evidence on analysis and evaluation; no controlled CT estimate
Parno et al. (2021)	Grade 11 physics; Indonesia; N=78	Three-group static comparison; PBL-STEM + virtual simulation vs PBL vs conventional instruction	CT test; N-gain 0.75 / 0.49 / 0.32; STEM is a co-intervention
Cahyono et al. (2021)	Preservice teachers; Indonesia; N=28	One-group pre/post; PBL + argumentation scaffolding	Essay CT score about 26.71 to 61.03; N-gain 0.61; minor internal reporting inconsistency
Lee & Son (2021)	Nursing; South Korea; N=105	Controlled pre/post; simulation-PBL vs demonstration	Self-report CT questionnaire; larger increase in simulation-PBL group ($p < 0.001$)
Kummer et al. (2022)	Forensic education; Switzerland	Implementation study; 360-degree crime-scene simulation and communication app	Interaction and satisfaction/process evidence; no direct CT performance estimate
Chua (2023)	Preservice teachers; Singapore; N=1,041	Path analysis of PBL processes, scaffolding, and learning strategies	MSLQ self-report subscale; associative evidence, not a causal CT estimate

Suhirman & Prayogi (2023)	Upper-secondary students; Indonesia; N=88	Randomized pre/post control; PBL + mobile PhET vs expository instruction	Performance CT test; N-gain 0.67 vs 0.18; significant between-group difference reported ($p < 0.05$)
Serrano-Estrada et al. (2023)	Urban studies; Spain; N=59	Problem-based activity with digital/QGIS scaffolding and geolocated data	Questionnaires and student artifacts; interpretive/critical skills, without a strong controlled causal estimate
Macdiarmid et al. (2024)	Graduate-entry nursing; United Kingdom; student interviews (n=5) plus staff reflective data	Educational design research; unfolding case studies supported by multiple resources including simulation labs	Qualitative perspectives on critical thinking/clinical reasoning; contextual rather than comparative evidence
Lee (2025)	Design education; South Korea; two scaffolding groups	15-week PBL/design-thinking course; hard vs soft scaffolding with generative AI	Focus groups/surveys; no standardized direct CT effect estimate
Omeh et al. (2025)	Computer-science education; Nigeria; N=62	Nonrandomized pre/post; AI-assisted PBL vs PBL with instructional videos	Seven item self report CT scale; no significant between group difference reported
Wagino et al. (2025)	Automotive vocational education; Indonesia; N=80	CB-SEM questionnaire study; PBL/e-learning with interactive tools and virtual simulations	Self reported CT and engagement; no objective CT counterfactual estimate
Abeysuriya et al. (2026)	Postgraduate immunology; Sri Lanka; N=30 (29 feedback forms)	PBL with knowledge scaffolding; post-intervention feedback	Self reported problem solving, confidence, and acceptability; no control or objective CT test
Lailiyah et al. (2026)	Mathematics, Grades 5 and 9; Indonesia; N=126	PBL + PhET; mixed quasi-experimental and associative analyses	CT/problem solving tests; positive associations on selected indicators varied by school level; attribution remains unclear
Muhali & Asy'ari (2026)	Preservice science teachers; Indonesia; N=125	One-group pre/post across four intact classes; conceptual PBL + scaffolding	Eight item essay CT test; within-group $p < .001$; moderate N-gain (0.44 – 0.66); no independent control

Xue et al. (2026)	Trauma-orthopedic medical education; China; N=50	Assessor-blinded randomized comparison; PBL + virtual simulation vs lecture	Seven-dimension self-report CT questionnaire; mean difference 43.84, 95% CI (34.70, 52.98); small single-site sample

All six direct comparisons reported outcomes favoring the simulation-supported or integrated PBL condition, but the direction of findings should not be confused with equivalent evidential strength. Simanjuntak et al. (2020) reported posttest scores of 74.45 versus 53.95 and N-gain values of 58% versus 28% for PBL with Adobe Flash and traditional teaching, respectively. Serungke et al. (2020) reported N-gain values of 69.54 versus 40.44 for PBL with a virtual laboratory and conventional teaching. Parno et al. (2021) reported N-gain values of 0.75 for PBL-STEM with virtual simulation, 0.49 for PBL, and 0.32 for conventional instruction; the bundled STEM component prevents isolation of the simulation effect. Suhirman and Prayogi (2023) reported N-gain values of 0.67 versus 0.18 for mobile-PhET-supported PBL and expository instruction, with a significant between-group difference ($p < .05$). Lee and Son (2021) found a greater increase in self-reported critical thinking for simulation-PBL than for demonstration ($p < .001$). Xue et al. (2026) reported a mean difference of 43.84 on a self-report critical-thinking questionnaire (95% CI [34.70, 52.98]) in favor of PBL with virtual simulation. Thus, four comparisons used performance-based measures and two used self-report measures; only two studies used randomized designs, and their outcomes differed in type. Because scores, N-gain values, questionnaires, and comparators were not sufficiently comparable, these results support a generally favorable direction but not a single pooled effect estimate.

Table 2. Direct comparisons of simulation-supported PBL

Study	Comparison	Outcome / reported result	Direction	Main limitation
Simanjuntak et al. (2020)	PBL + Adobe Flash vs traditional teaching	Performance CT test/essay; posttest 74.45 vs 53.95; N-gain 58% vs 28%	Favors PBL simulation	Quasi experimental; allocation not random and baseline equivalence incompletely established.
Serungke et al. (2020)	PBL + virtual laboratory vs conventional teaching	Performance CT test; N-gain 69.54 vs 40.44	Favors PBL simulation	Controlled comparison, but incomplete reporting limits risk of bias appraisal.

Parno et al. (2021)	PBL-STEM + virtual simulation vs PBL vs conventional teaching	Performance CT test; N-gain 0.75 / 0.49 / 0.32	Highest gain in integrated condition	PBL, STEM, and simulation were bundled; the incremental effect of simulation cannot be isolated.
Lee & Son (2021)	Simulation-PBL vs demonstration	Self-report CT questionnaire; larger increase for simulation PBL ($p < 0.001$)	Favors simulation-PBL on perceived CT	Self reported outcome; nursing specific context and perception/performance distinction.
Suhirman & Prayogi (2023)	PBL + mobile PhET vs expository instruction	Performance CT test; N-gain 0.67 vs 0.18; significant difference reported ($p < 0.05$)	Favors PBL simulation	Randomized comparison is a strength; fidelity and transfer to other contexts remain uncertain.
Xue et al. (2026)	PBL + virtual simulation vs lecture	Self report CT questionnaire; mean difference 43.84, 95% CI (34.70, 52.98)	Favors integrated condition on questionnaire score	Assessor-blinded randomized design; small single-site sample and self report outcome.

The six direct comparisons do not provide a uniform level of causal confidence. Randomized comparisons by Suhirman and Prayogi (2023) and Xue et al. (2026) offer stronger protection against baseline confounding than nonrandomized comparisons, but their evidence still differs because the former used a performance test while the latter used a self-report questionnaire. Simanjuntak et al. (2020) and Serungke et al. (2020) included alternative instructional conditions, but nonrandom allocation or incomplete reporting leaves some concerns. Parno et al. (2021) is informative about an integrated PBL-STEM-simulation package, not the isolated effect of simulation. Ramadan et al. (2020) reported different gains in two implementation classes that both received the intervention and therefore does not provide a conventional untreated comparison.

Measurement changes the meaning of a positive finding. Performance tests and essays can indicate how learners reason under defined conditions, provided that instruments and scoring rubrics are valid and reported. Self-report questionnaires measure perceived critical-thinking capability or learning experience and should not be treated as interchangeable with performance. Accordingly, the findings from Lee and Son (2021) and Xue et al. (2026) support a claim about perceived CT scores, not a direct claim that performance improved by the same amount. The

distinction also applies to studies using learning-strategy subscales, satisfaction, or feedback forms.

Outside the six direct comparisons, one-group pre/post studies, qualitative case studies, path-analysis models, and implementation surveys contribute evidence about feasibility or plausible learning processes. Cahyono et al. (2021) and Muhali and Asy'ari (2026) reported gains after PBL with scaffolding but lacked independent untreated controls. Chua (2023) identified associations between PBL process variables and self-reported learning strategies, not causal improvement in performance-based CT. Abeysuriya et al. (2026), Kummer et al. (2022), Macdiarmid et al. (2024), Serrano-Estrada et al. (2023), and Wagino et al. (2025) clarify learner experience and implementation context but cannot substitute for controlled effectiveness evidence. Omeh et al. (2025) found no statistically significant between-group difference in self-reported CT between AI-assisted PBL and PBL supported by instructional videos; this finding cautions against assuming that adding a technology necessarily improves outcomes.

The available designs permit only limited attribution to individual components. Across the 19 reports, six directly compared simulation-supported PBL with another instructional condition, but these comparisons evaluated intervention packages rather than a complete set of matched contrasts separating PBL, simulation, and scaffolding. Parno et al. (2021) combined PBL, STEM, and virtual simulation, so its higher N-gain cannot be attributed to simulation alone. Ramadan et al. (2020) compared two classes that both received the intervention and therefore did not provide an untreated counterfactual. Lailiyah et al. (2026) combined quasi-experimental and associative analyses, which do not yield one unambiguous isolated treatment effect. Cahyono et al. (2021) and Muhali and Asy'ari (2026) reported pre/post gains after scaffolded PBL without an independent untreated control, while Chua (2023) reported associations between PBL-process variables and self-reported learning strategies. Accordingly, the corpus does not establish the additional causal contribution of scaffolding when simulation is already embedded in PBL; this remains an evidence gap rather than evidence that scaffolding has no effect.

The evidence likewise does not establish an additional causal benefit of explicit scaffolding once simulation is already embedded in PBL. Scaffolding may help learners select relevant information, connect observations to claims, explain evidence, and monitor reasoning. These are plausible instructional mechanisms, but the included corpus does not contain a sufficient set of matched comparisons that isolates the added contribution of scaffolding in the presence of simulation (Hmelo-Silver et al., 2007; Dominguez & Svihla, 2023). The appropriate conclusion is therefore a research need, not a claim of no effect.

Across the process and contextual studies, the recurring candidate factors include the authenticity and tractability of problems, alignment between inquiry tasks and the target reasoning skill, teacher facilitation, structured but non-substitutive guidance, opportunities for collaboration and reflection, and digital usability. These factors are recurring explanations or hypotheses rather than verified causal moderators. Simulation may enable learners to manipulate variables and repeat trials, but its value depends on how observations are incorporated into evidence-based explanations (de Jong, 2006; Rutten et al., 2012). Guidance must also be matched to learners' needs instead of added without regard to task demands (Lazonder & Harmsen, 2016).

To address the attribution problem identified in RQ3, future studies should compare three conditions: (1) PBL alone, (2) PBL plus interactive simulation, and (3) PBL plus simulation plus explicit scaffolding. Researchers should keep content, teaching time, teacher contact, group structure, and assessment comparable across conditions. The first contrast would estimate the incremental contribution of simulation within PBL, whereas the second would estimate the additional contribution of scaffolding. A factorial design could estimate main effects and interactions when sample size and implementation permit. Studies should report allocation, baseline equivalence, attrition, missing-data handling, implementation fidelity, co-interventions, effect estimates with confidence intervals, and delayed or transfer outcomes. Measures should

distinguish performance-based critical thinking from perceived capability and should use transparent rubrics aligned with defined dimensions (Facione, 1990).

The search was limited to Scopus, English-language open-access articles and conference papers, and publications from 2020 through 2026. Relevant studies indexed elsewhere, published in other languages, or not designated open access may have been missed. The included studies differed in educational level, discipline, intervention package, comparator, instrument, and design, so meta-analysis was not appropriate. The record-level screening log was reconstructed from the available export and reports, and the source materials did not document independent duplicate screening or formal adjudication. Risk-of-bias judgements are narrative and design-specific rather than a single validated score. The findings should therefore be read as a critical narrative synthesis, not as a pooled estimate of effect.

CONCLUSION

The reconstructed review retained 19 empirical reports from 52 filtered Scopus records. Eight studies primarily examined PBL with simulation, and six provided direct comparisons with alternative instruction. Their findings generally favored simulation-supported PBL, but causal certainty varied because of allocation methods, bundled components, incomplete reporting, and differences between performance tests and self-report outcomes.

The included evidence does not establish an added causal benefit of explicit scaffolding when simulation is already integrated into PBL. Process and implementation studies identify plausible mechanisms, but cannot replace controlled component comparisons. Future studies should directly compare PBL alone, PBL with simulation, and PBL with simulation plus scaffolding under matched conditions, using validated performance-based measures, transparent implementation reporting, and delayed or transfer assessment.

ACKNOWLEDGMENT

No non-author contribution or external support requiring acknowledgment was documented in the materials used to prepare this manuscript.

AUTHOR CONTRIBUTION STATEMENT

EHP contributed to conceptualization, methodology, investigation, data curation, formal analysis, visualization, and writing original draft preparation. IMA, EB, and FCW contributed to supervision, validation, and writing, review and editing. All authors reviewed and approved the final manuscript.

AI DISCLOSURE STATEMENT

OpenAI ChatGPT was used during manuscript preparation to support English language translation, organization of sections and tables, copyediting, formatting, and preliminary organization of screening records. The authors remain responsible for reviewing and editing the complete manuscript and for verifying the search and screening decisions, extracted data, risk-of-bias judgments, references, and interpretations against the original reports.

REFERENCES

Abeysuriya, V., Gamage, J., Jayawardana, W. A. D., & Abeysuriya, V. (2026). Enhancing adult learning through problem-based learning and knowledge scaffolding: A case study from postgraduate immunology education. *BMC Medical Education*, *26*, Article 1059. https://doi.org/10.1186/s12909-026-09488-x

Barker, T. H., Habibi, N., Aromataris, E., Stone, J. C., Leonardi-Bee, J., Sears, K., Hasanoff, S., Klugar, M., Tufanaru, C., Moola, S., & Munn, Z. (2024). The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. *JBI Evidence Synthesis*, *22*(3), 378–388. https://doi.org/10.11124/JBIES-23-00268

Cahyono, B., Kartono, K., Waluya, B., Mulyono, M., & Setyawati, R. D. (2021). Problem-based learning supported by arguments scaffolding that affect critical thinking teacher candidates. *Cypriot Journal of Educational Sciences*, *16*(6), 2956–2969. https://doi.org/10.18844/cjes.v16i6.6480

Chua, B. L. (2023). Path analysis: The predictive relationships of problem-based learning processes on preservice teachers' learning strategies. *Interdisciplinary Journal of Problem-Based Learning*, *17*(2). https://doi.org/10.14434/ijpbl.v17i2.37322

de Jong, T. (2006). Computer simulations: Technological advances in inquiry learning. *Science*, *312*(5773), 532–533. https://doi.org/10.1126/science.1127750

Dominguez, S., & Svihla, V. (2023). A review of teacher implemented scaffolding in K-12. *Social Sciences & Humanities Open*, *8*(1), Article 100613. https://doi.org/10.1016/j.ssaho.2023.100613

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*. California Academic Press.

Flegr, S., Kuhn, J., & Scheiter, K. (2023). When the whole is greater than the sum of its parts: Combining real and virtual experiments in science education. *Computers & Education*, *197*, Article 104745. https://doi.org/10.1016/j.compedu.2023.104745

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*, 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Kummer, N., Delémont, O., Voisard, R., & Weyermann, C. (2022). The potential of digital technologies in problem-based forensic learning activities. *Science & Justice*, *62*(6), 740–748. https://doi.org/10.1016/j.scijus.2022.04.005

Lailiyah, S., Kusaeri, K., Asmiyah, S., Naryaningsih, P. D., Dina, A. M., & Irmanila, E. (2026). Combined PBL and PhET simulation for critical thinking in mathematical reasoning and problem-solving skills. *Journal of Education and Learning*, *20*(2), 1138–1152. https://doi.org/10.11591/edulearn.v20i2.21849

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, *86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Lee, H. (2025). Exploratory study on scaffolding strategies for autonomous generative AI design education: A case of learning by doing-based PBL design thinking education. *Archives of Design Research*, *38*(1), 199–220. https://doi.org/10.15187/adr.2025.02.38.1.199

Lee, J., & Son, H. K. (2021). Comparison of learning transfer using simulation problem-based learning and demonstration: An application of Papanicolaou smear nursing education. *International Journal of Environmental Research and Public Health*, *18*(4), Article 1765. https://doi.org/10.3390/ijerph18041765

Liu, Y., & Pasztor, A. (2022). Effects of problem-based learning instructional intervention on critical thinking in higher education: A meta-analysis. *Thinking Skills and Creativity*, *45*, Article 101069. https://doi.org/10.1016/j.tsc.2022.101069

Macdiarmid, R., Merrick, E., & Winnington, R. (2024). Using unfolding case studies to develop critical thinking for Graduate Entry Nursing students: An educational design research study. *BMC Nursing*, *23*, Article 399. https://doi.org/10.1186/s12912-024-02076-8

Muhali, M., & Asy'ari, M. (2026). Conceptual problem-based learning model: Promising intervention to enhance prospective science teachers' critical thinking skills. *Social Sciences & Humanities Open*, *13*, Article 102509. https://doi.org/10.1016/j.ssaho.2026.102509

Muilwijk, S. E., & Lazonder, A. W. (2023). Learning from physical and virtual investigation: A meta-analysis of conceptual knowledge acquisition. *Frontiers in Education*, *8*, Article 1163024. https://doi.org/10.3389/feduc.2023.1163024

Omeh, C. B., Ayanwale, M. A., Mnguni, L. E., & Olelewe, C. J. (2025). Fostering programming skill and critical thinking through AI-assisted PBL integration. *Journal of New Approaches in Educational Research*, *14*, Article 22. https://doi.org/10.1007/s44322-025-00041-0

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Parno, P., Fauziyah, S., Pramono, N. A., Anggraini, R. T., Hidayat, A., Supriana, E., & Ali, M. (2021). The increase of students' critical thinking abilities on optical instrument topic through PBL-STEM with virtual simulation media. *Journal of Physics: Conference Series*, *1918*(5), Article 052067. https://doi.org/10.1088/1742-6596/1918/5/052067

Ramadan, E. M., Jumadi, & Astuti, D. P. (2020). Application of e-handout based on PhET simulation to improve critical thinking skills and learning independence of high school students. *Journal of Physics: Conference Series*, *1440*, Article 012025. https://doi.org/10.1088/1742-6596/1440/1/012025

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, *10*, Article 39. https://doi.org/10.1186/s13643-020-01542-z

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, *58*(1), 136–153. https://doi.org/10.1016/j.compedu.2011.07.017

Serrano-Estrada, L., Martin, T. J., & Martí, P. (2023). Understanding city dynamics: Using geolocated social media in a problem-based activity as an investigative tool to enhance student learning. *Journal of Geography in Higher Education*, *47*(1), 106–127. https://doi.org/10.1080/03098265.2021.2004390

Serungke, M., Muhibbuddin, & Suhrawardi. (2020). Implementation of problem-based learning (PBL) with virtual laboratory to improve students' critical thinking and achievement. *Journal of Physics: Conference Series*, *1460*, Article 012134. https://doi.org/10.1088/1742-6596/1460/1/012134

Simanjuntak, M., Purba, C., Panggabean, B., Mustafa, A., & Hutahaean, J. (2020). The effect of problem based learning (PBL) towards critical thinking used computer simulation. In *Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar (MSCEIS 2019)*. European Alliance for Innovation. https://doi.org/10.4108/eai.12-10-2019.2296467

Suhirman, S., & Prayogi, S. (2023). Problem-based learning utilizing assistive virtual simulation in mobile application to improve students' critical thinking skills. *International Journal of Education and Practice*, *11*(3), 351–364. https://doi.org/10.18488/61.v11i3.3380

Sukatiman, M., Akhyar, M., Siswandari, S., & Roemintoyo. (2020). Enhancing higher-order thinking skills in vocational education through scaffolding-problem based learning. *Open Engineering*, *10*(1), 612–619. https://doi.org/10.1515/eng-2020-0070

Wagino, W., Maksum, H., Purwanto, W., Ambiyar, A., Samala, A. D., Criollo-C., S., & Koto, R. D. (2025). Assessing student expectations, perceived quality, value, and satisfaction through

PBL and e-learning integration in vocational education. *Journal of Educators Online*, *22*(4). https://doi.org/10.9743/JEO.2025.22.4.15

Worner, S., Kuhn, J., & Scheiter, K. (2022). The best of two worlds: A systematic review on combining real and virtual experiments in science education. *Review of Educational Research*, *92*(6), 911–952. https://doi.org/10.3102/00346543221079417

Xue, G., Zhang, X., Ji, S., & Qi, J. (2026). Integrating problem-based learning with virtual simulation in trauma orthopedic education: Effects on learning outcomes and student satisfaction. *BMC Medical Education*, *26*, Article 1240. https://doi.org/10.1186/s12909-026-09558-0

Yu, L., & Zin, Z. M. (2023). The critical thinking-oriented adaptations of problem-based learning models: A systematic review. *Frontiers in Education*, *8*, Article 1139987. https://doi.org/10.3389/feduc.2023.1139987