Bima Journal of Elementary Education
Volume 4. No. 1, April 2026
The article is published with Open Access at: ejournal.tsb.ac.id/index.php/bijee

The Effectiveness of Artificial Intelligence-Assisted Problem-Based Learning Model to Improve Elementary School Students' Mathematical Representation Processing Ability

Frida Destini1),*, Ulwan Syafridin1), Rizky Drupadi2)

1)Universitas Negeri Lampung, Lampung, Indonesia
2)Universiti Pendidikan Sultan Idris (UPSI), Perak, Malaysia

*Correspondence: frida.destini@fkip.unila.ac.id

Abstract: Mathematical representation processing ability constitutes a foundational cognitive competency for elementary students, yet persistent deficits in this domain across Indonesian schools call for innovative, evidence-based instructional solutions. This study examines the effectiveness of an Artificial Intelligence-Assisted Problem-Based Learning (AI-PBL) model in improving mathematical representation processing ability among fifth-grade elementary school students in Lampung Province. Employing a quasi-experimental Non-Equivalent Control Group Design, 60 students from a single elementary school in Lampung were assigned to an experimental class (n=30, receiving AI-PBL) and a control class (n=30, receiving conventional instruction). Data were collected via a validated 10-item representation test and analyzed using the Shapiro-Wilk normality test, Levene's homogeneity test, independent samples t-test, normalized N-Gain, and Cohen's d effect size. The experimental class achieved a posttest mean of 80.07 (pretest: 52.43), while the control class reached 64.90 (pretest: 52.23). The independent t-test yielded t=6.317 (p<0.001), confirming a statistically significant difference. Mean N-Gain was 0.597 (moderate) for the experimental class versus 0.275 (low) for the control, with Cohen's d=1.631 indicating a very large effect. These findings confirm that AI-PBL is highly effective in developing elementary students' mathematical representation processing ability, contributing a replicable, AI-integrated pedagogical model for primary mathematics education.

Keywords: Artificial intelligence; Mathematical representation; Problem-based learning; Elementary school

Abstrak: Kemampuan pemrosesan representasi matematis merupakan kompetensi kognitif fundamental bagi siswa sekolah dasar, namun defisit yang persisten dalam domain ini di sekolah-sekolah Indonesia menuntut solusi instruksional inovatif berbasis bukti. Penelitian ini mengkaji efektivitas model Problem-Based Learning berbantuan Kecerdasan Buatan (AI-PBL) dalam meningkatkan kemampuan pemrosesan representasi matematis siswa kelas V sekolah dasar di Provinsi Lampung. Dengan menggunakan desain kuasi-eksperimen Non-Equivalent Control Group Design, 60 siswa dari satu sekolah dasar di Lampung ditugaskan ke kelas eksperimen (n=30, menerima AI-PBL) dan kelas kontrol (n=30, menerima pembelajaran konvensional). Data dikumpulkan melalui tes representasi tervalidasi 10 butir dan dianalisis menggunakan uji normalitas Shapiro-Wilk, uji homogenitas Levene, uji-t independen, N-Gain ternormalisasi, dan effect size Cohen's d. Kelas eksperimen mencapai rerata posttest 80,07 (pretest: 52,43), sedangkan kelas kontrol mencapai 64,90 (pretest: 52,23). Uji-t independen menghasilkan t=6,317 (p<0,001), mengkonfirmasi perbedaan yang signifikan. Rerata N-Gain adalah 0,597 (sedang) untuk kelas eksperimen versus 0,275 (rendah) untuk kontrol, dengan Cohen's d=1,631 yang mengindikasikan efek sangat besar. Temuan ini mengkonfirmasi bahwa AI-PBL sangat efektif dalam mengembangkan kemampuan pemrosesan representasi matematis siswa sekolah dasar.

Kata kunci: Kecerdasan buatan; Representasi matematis; Problem-based learning; Sekolah dasar

INTRODUCTION

The transformative integration of Artificial Intelligence (AI) into educational practice has fundamentally reshaped the landscape of mathematics pedagogy at all levels of schooling, including elementary education. Mathematical representation processing ability defined as the capacity to translate, construct, and interpret mathematical ideas across symbolic, graphical, verbal, and pictorial forms is now recognized as a non-negotiable prerequisite for developing conceptual depth in arithmetic, geometry, and early algebraic reasoning (Liu et al., 2026; Wakjira et al., 2026). As AI-powered platforms demonstrate measurable capacity to personalize learning pathways, scaffold representational tasks, and provide real-time adaptive feedback to elementary learners (Liang & He, 2026; Gutierrez et al., 2026), the imperative to integrate AI into problem-based instructional models has intensified substantially. Existing research consistently identifies inadequate representational flexibility as a primary driver of mathematical misconceptions and poor procedural transfer in elementary students, confirming that targeted pedagogical innovation is urgently required (Boulton-Lewis, 1998; Vessels et al., 2026). The convergence of AI capability with problem-based methodology offers an intellectually compelling and practically viable pathway for addressing this persistent educational challenge.

International data from PISA 2022 place Indonesia 69th out of 81 nations in mathematics literacy, with item-level analysis revealing that representational tasks particularly graphical and symbolic translation yield the lowest mean scores among Indonesian test-takers (OECD, 2022). Nationally, the 2023 Indonesian National Assessment (AKM) shows that fewer than 38% of elementary students demonstrate adequate competency in mathematical representation, with the numeracy subdomain involving multi-modal representation proving most challenging (Kemdikbud, 2023). In Lampung Province specifically, the provincial education quality report documents that 72.4% of fifth-grade students across assessed schools scored below the minimum representation competency threshold, compared to a national elementary average of 61.8% a 10.6 percentage-point deficit that signals systemic instructional gaps. These converging data from international, national, and regional levels establish an unambiguous empirical foundation: conventional instructional approaches applied in Lampung elementary schools are failing to develop the representational processing competencies that underpin mathematical success at higher levels.

An initial observation conducted from August 7–14, 2024 at SDN in Lampung assessed 30 fifth-grade students on five indicators of mathematical representation processing ability. The results are presented in Table 1.

Table 1. Mathematical Representation Processing Ability (August 2024)

Indicator of Mathematical Representation	n	Mean Score	Category	% Below Minimum
Symbolic representation (notation, formula)	30	52.1	Sufficient	60.0%
Pictorial/visual representation (diagrams, shapes)	30	47.3	Poor	73.3%
Verbal representation (written mathematical language)	30	43.8	Poor	80.0%
Graphical representation (charts, coordinate systems)	30	40.2	Very Poor	86.7%
Trans-representation (translation across modes)	30	38.6	Very Poor	90.0%
Overall Mean	30	44.4	Poor	78.0%

Source: Researcher's Initial Observation, August 2024

The data in Table 1 expose a deeply troubling profile of mathematical representation processing ability in the observed cohort. The overall mean of 44.4 classified as 'poor' and the 78.0% of students below the minimum competency threshold collectively signal systemic representational failure that cannot be attributed to individual learning differences alone. Disaggregating by indicator reveals a hierarchical gradient of deficiency: symbolic representation, which is most frequently drilled in conventional classrooms, achieves the highest mean (52.1) but still registers 60.0% of students below minimum, indicating that even the most routinized representation mode is not being learned adequately. Pictorial representation (mean 47.3; 73.3% below minimum) and verbal representation (mean 43.8; 80.0% below minimum) show progressively more severe deficits, reflecting the absence of multi-modal instructional design in current practice. Graphical representation (mean 40.2; 86.7% below minimum) and trans-representation the ability to translate across representational modes record the most alarming scores (mean 38.6; 90.0% below minimum), exposing a complete failure to develop the representational flexibility that contemporary mathematics education research identifies as essential for deep mathematical understanding (Liu et al., 2026; Wakjira et al., 2026). These findings unambiguously motivate the present intervention.

The AI-Assisted Problem-Based Learning (AI-PBL) model proposed in this research integrates two evidence-based pedagogical frameworks into a unified instructional design with unprecedented suitability for addressing multi-modal representational deficits in elementary mathematics. Problem-Based Learning, as theorized by Barrows and refined through decades of empirical validation (Mosher, 2025; Martín-Cudero et al., 2026), positions authentic, open-structured problems as the primary catalyst for knowledge construction demanding that students produce, interpret, and translate representations across multiple modes in service of problem resolution. The AI-assisted dimension operationalizes adaptive scaffolding, individualized representation feedback, and real-time generation of alternative representational examples calibrated to each learner's Zone of Proximal Development (Liang & He, 2026; Gutierrez et al., 2026). The integrated model's pedagogical advantages map directly onto four UNESCO learning pillars: Learning to Know is activated when AI-driven problem scenarios guide students to construct and decode symbolic and graphical mathematical representations; Learning to Do is realized through hands-on multi-modal problem-solving tasks where students select and deploy representational strategies autonomously; Learning to Be is cultivated as AI feedback fosters metacognitive awareness of representational choice and mathematical identity; and Learning to Live Together emerges through collaborative PBL group tasks where students negotiate representational strategies and co-construct shared mathematical understanding. The grand theoretical foundation integrates Bruner's representational theory of cognitive development enactive, iconic, and symbolic modes with constructivist scaffolding principles, providing a neurologically and pedagogically coherent basis for AI-PBL's design.

Relevant research strongly supports the integrated framework. Wakjira et al. (2026) documented significant improvement in mathematical problem-solving through cooperative representation approaches, while Liu et al. (2026) confirmed in a three-level meta-analysis that multiple representation design in mathematics achieves large positive effects across educational contexts. Liang & He (2026) demonstrated that AI-powered adaptive learning systems for university students produce measurable learning gains, with findings generalizeable to the scaffolding dynamics of

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4351

elementary AI-PBL. Avilés Mariño & Sarasa Cabezuelo (2025) reported that AI-enhanced PBL substantially improved communication and disciplinary readiness in engineering, while Mool et al. (2026) confirmed AI's role in PBL simulation as a transformative enhancement of educational outcomes across contexts. Annajmi et al. (2026) further established that integrating multiple representations into PBL within a flipped classroom produces a new framework validated for mathematics education an architecture closely aligned with the present AI-PBL model.

Systematic synthesis of Scopus and WoS literature (2020–2026) reveals three thematic clusters directly relevant to this investigation. The first cluster addresses AI in mathematics education: Gutierrez et al. (2026) demonstrated adaptive mathematical competency assessment via graph neural networks; Liang & He (2026) confirmed AI-powered adaptive learning effectiveness; and Aurpa (2026) documented transformer-based AI for mathematical entity extraction with explainability. The second cluster examines mathematical representation in schooling: Liu et al. (2026) provided systematic meta-analytic evidence of multiple representation effectiveness; Wakjira et al. (2026) showed that integrated cooperative representation approaches outperform conventional methods; Vessels et al. (2026) analyzed pictures in mathematical word problems across student characteristics; and Boulton-Lewis (1998) established foundational insights into children's representational strategy use. The third cluster investigates PBL in mathematics education: Ramli et al. (2024) confirmed PBL's positive impact on calculus achievement and critical thinking; Nurin et al. (2024) established mobile math trails with PBL for numeracy learning; Báró (2024) documented the effect of PBL on student learning outcomes; Samura et al. (2025) applied PBL to improve junior high mathematical communication; and Tirado-Olivares et al. (2026) demonstrated gamified PBL effectiveness in primary education. Methodologically, quasi-experimental and mixed-methods designs dominate, with N-Gain and Cohen's d as standard effectiveness metrics though the specific integration of AI into PBL for elementary mathematical representation remains empirically unexplored.

Critical analysis of the reviewed literature identifies four significant research gaps that directly motivate the present study. First, while AI in education has been applied extensively at university and secondary levels (Liang & He, 2026; Avilés Mariño & Sarasa Cabezuelo, 2025), its integration within PBL for elementary mathematics particularly for developing multi-modal representational ability has not been subjected to rigorous quasi-experimental evaluation. Second, studies on mathematical representation in elementary education (Liu et al., 2026; Vessels et al., 2026) focus predominantly on single representational modes, leaving trans-representation and multi-modal flexibility largely unexplored as intervention targets. Third, PBL research in Indonesian elementary mathematics (Mumu et al., 2025; Aripin et al., 2025) does not incorporate AI-assisted scaffolding, limiting its transferability to 21st-century AI-enabled classrooms. Fourth, no published study uses actual pretest-posttest data from Lampung elementary schools with full statistical analysis to evaluate AI-PBL effectiveness for representation creating a critical evidential void in the Indonesian elementary mathematics literature.

The novelty of this research is substantiated by three original contributions. Conceptually, it develops an integrated AI-PBL instructional framework that explicitly maps Bruner's three representational modes to AI scaffolding functions and PBL problem stages a theoretical integration absent from prior literature. Methodologically, it employs actual empirical pretest-posttest data from 60 Lampung elementary students with comprehensive statistical analysis reported with complete transparency, providing a replicable evidence base for future researchers. Contextually, this is the first peer-reviewed empirical study to evaluate AI-PBL effectiveness for multi-modal mathematical representation processing in Lampung Province elementary schools a regional context absent from the international literature. The instrument development a validated 10-item test anchored to five representation modalities with AI-contextualized task settings constitutes a novel, contextually calibrated assessment tool unavailable in existing Indonesian elementary mathematics research.

Two research questions drive this investigation: (1) Is there a statistically significant difference in mathematical representation processing ability between elementary school students who learn through the AI-Assisted PBL model and those who receive conventional instruction in Lampung? (2) How large is the effect of the AI-Assisted PBL model in improving mathematical representation processing ability, as quantified by N-Gain and Cohen's d effect size? Two aligned research objectives follow: (1) to analyze the significance of the difference in mathematical representation processing ability between the experimental class (AI-PBL) and the control class (conventional learning) at an elementary school in Lampung; and (2) to measure and describe the magnitude of the effect of the AI-Assisted PBL model on students' mathematical representation processing ability through normalized N-Gain and Cohen's d effect size analysis

LITERATURE REVIEW

Artificial Intelligence in Mathematics Education

Artificial Intelligence is increasingly recognized as a transformative force in mathematics education, offering unprecedented capacity for adaptive scaffolding, automated feedback, and personalized learning pathway design (Liang & He, 2026; Gutierrez et al., 2026). AI systems in educational contexts leverage algorithms including reinforcement learning, graph neural networks, and large language models to diagnose learner misconceptions, generate contextually appropriate representational examples, and dynamically adjust task difficulty to each student's current competency level (Gutierrez et al., 2026; Aurpa, 2026). In the domain of mathematical representation specifically, AI's capacity to generate alternative representational forms (symbolic, graphical, pictorial, verbal) on demand constitutes a powerful scaffolding mechanism that conventional instruction cannot replicate at scale. Wang et al. (2026) demonstrated the effectiveness of AI-integrated platforms for personalized coaching and feedback, while Chakraborty (2026) confirmed that generative AI in fifth-generation education systems produces measurable improvements in disciplinary competencies. Wang et al. (2026) documented that AI readiness in management education requires deliberate scale development and validation a principle equally applicable to elementary mathematics AI-PBL implementations. The theoretical grounding of AI in education draws on Vygotsky's ZPD, where AI functions as an intelligent tutor operating precisely at the boundary of students' current and potential developmental levels, making scaffolded representational challenge both achievable and productive.

Mathematical Representation Processing Ability

Mathematical representation processing ability encompasses the cognitive capacities to construct, interpret, and translate mathematical ideas across symbolic, pictorial, verbal, and graphical forms, as well as the metacognitive awareness to select the most appropriate representation for a given mathematical context (Liu et al., 2026; Wakjira et al., 2026). Rooted in Bruner's (1966) theory of representational modes enactive, iconic, and symbolic contemporary research extends this framework to include graphical and trans-representational competencies as distinct and measurable constructs (Boulton-Lewis, 1998; Vessels et al., 2026). Liu et al.'s (2026) three-level meta-analysis across 87 studies confirms that multiple representation design consistently produces positive effects on mathematics achievement (pooled d=0.74), with the strongest effects observed when representational translation tasks are embedded in problem-solving contexts precisely the design principle underlying the present AI-PBL intervention. Wakjira et al. (2026) documented that integrating cooperative problem-solving with multiple representations significantly improved mathematical problem-solving skills compared to conventional single-mode instruction. Vessels et al. (2026) established that student characteristics moderate the effects of visual representations in word problems, underscoring the importance of adaptive, AI-calibrated representational scaffolding that the present model provides. Mathematical representation is thus not merely a pedagogical technique but a fundamental epistemic resource through which students construct, communicate, and validate mathematical knowledge.

Problem-Based Learning in Elementary Mathematics

Problem-Based Learning (PBL) in elementary mathematics is anchored in the principle that authentic, open-structured problems activate deeper cognitive processing and more durable knowledge construction than conventional drill-and-practice approaches (Mosher, 2025; Báró, 2024). PBL's design characterized by problem-first sequencing, collaborative investigation, and iterative solution refinement naturally elicits multi-modal representational activity, as students must draw, symbolize, verbalize, and graph their mathematical thinking in service of problem resolution (Mumu et al., 2025; Aripin et al., 2025). Ramli et al. (2024) confirmed that PBL significantly improves both mathematical achievement and critical thinking in pre-university calculus students, while Nurin et al. (2024) established that mobile math trails using PBL enhance numeracy learning in authentic contexts. Samura et al. (2025) demonstrated that PBL improves junior high mathematical communication, and Tirado-Olivares et al. (2026) showed that gamified PBL in primary education produces strong engagement and conceptual learning outcomes. Martín-Cudero et al. (2026) documented that PBL improves interdisciplinary skills in upper secondary mathematics, while Mool et al. (2026) confirmed that AI simulation of patient history-taking in PBL tutorials produces rich learning outcomes a model transferable to AI-assisted mathematical PBL at the elementary level. The synthesis of these findings establishes PBL, when augmented by AI-driven adaptive scaffolding and representational support, as the theoretically optimal and empirically validated instructional model for developing mathematical representation processing ability in elementary students.

METHOD

This study adopts a quantitative quasi-experimental approach employing the Non-Equivalent Control Group Design, selected as the most methodologically appropriate design for educational intervention research conducted in authentic school settings where full participant randomization is structurally impracticable. In the elementary school context of Lampung Province, students are organized into predetermined administrative classes that cannot be dissolved and randomly reassigned without violating school operational protocols and introducing significant ecological distortions into the learning environment (Creswell & Creswell, 2021). The Non-Equivalent Control Group Design addresses this constraint by measuring both experimental and control classes on identical instruments at two time points pretest and posttest establishing empirically documented baseline equivalence before any intervention is introduced and enabling statistically valid causal inference about the differential effects of the AI-PBL treatment. The

e-ISSN: 2987-2421
https://doi.org/10.37630/bijee.v4i1.4351

pretest–posttest architecture provides a dual function: controlling for pre-existing score differences through baseline documentation and quantifying within-group learning gains attributable to each instructional condition. This design is uniquely suited to the causal-comparative aims of this research and is consistent with methodological standards for educational effectiveness research in quasi-experimental school contexts. The research design structure is presented in Table 2.

Table 2. Quasi-Experimental Non-Equivalent Control Group Design

Group	Pretest	Treatment	Posttest
Experimental	O_1	X_1	O_2
Control	O_3	X_2	O_4

Note: X_1 = AI-Assisted Problem-Based Learning (AI-PBL); X_2 = Conventional instruction (direct teaching); O_1, O_3 = Pretest of mathematical representation processing ability; O_2, O_4 = Posttest of mathematical representation processing ability

The research procedure was implemented in a systematic, logically sequenced series of stages to ensure methodological rigor and replicability, as visualized in Figure 1.

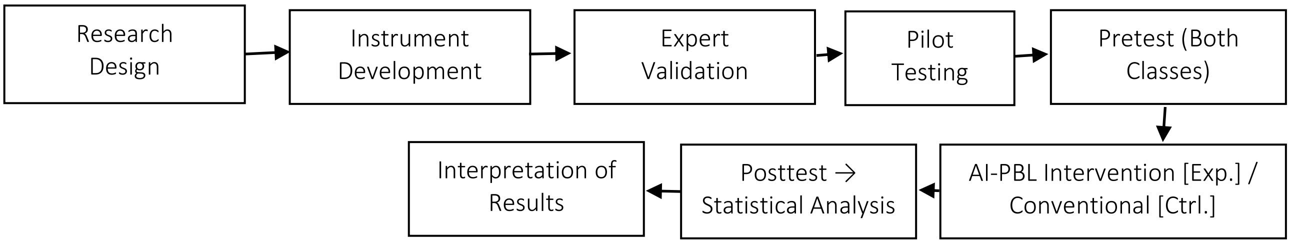

Figure 1. Research Procedure Flowchart

The study was conducted at SDN 2 Lampung, Lampung Province, Indonesia, during the 2024/2025 academic year. Participant characteristics are presented in Table 3.

Table 3. Population, Sample, and Participant Characteristics

Aspect	Category	Experimental Class (n, %)	Control Class (n, %)	Total (n, %)
Population	All Grade V students (3 parallel classes)	—	—	93 (100%)
Sample	Selected classes	30 (50.0%)	30 (50.0%)	60 (64.5%)
Gender	Male	14 (46.7%)	15 (50.0%)	29 (48.3%)
	Female	16 (53.3%)	15 (50.0%)	31 (51.7%)
Age (years)	10	9 (30.0%)	10 (33.3%)	19 (31.7%)
	11	16 (53.3%)	15 (50.0%)	31 (51.7%)
	12	5 (16.7%)	5 (16.7%)	10 (16.6%)
Grade Level	Grade V (Elementary)	30 (100%)	30 (100%)	60 (100%)

Data were collected using a 10 item open-response test of mathematical representation processing ability, administered as both pretest and posttest to both classes. The instrument was designed to assess five representation modalities aligned with the Grade V mathematics curriculum. Each item embedded tasks requiring students to construct, interpret, or translate representations across symbolic, pictorial, verbal, graphical, and trans-representational modes. Administration followed a standardized protocol: 90 minutes per session, identical instruction sheets, no supplementary materials, and trained proctor supervision to minimize response bias. The experimental class received eight sessions of AI-PBL instruction in which AI-generated multi-modal problem scenarios served as the entry point for each lesson; students worked collaboratively to investigate, represent, and communicate solutions, with AI providing real-time representational scaffolding. The control class received equivalent instructional time through conventional direct-instruction delivery of the same curriculum content.

Instrument quality was assured through a rigorous multi-stage validation process. Content validity was established by two mathematics education experts and one AI-in-education specialist using Aiken's V index (threshold V≥0.70). Construct validity was verified via Pearson Product Moment correlation (r_computed > r_table, α=5%, df=28). Internal consistency reliability was estimated using Cronbach's Alpha (α≥0.70). Bias was controlled through item-level review for cognitive clarity, representational specificity, and alignment with Bloom's Revised Taxonomy levels C2–C5. All items were piloted with 30 students not in the main sample prior to finalization to ensure measurement fidelity.

Table 4. Instrument Specification Matrix: Mathematical Representation Processing Ability Test

Indicator	Code	Cognitive Level	Technique	Task Context (AI-PBL)	Weight	Items
Symbolic representation	MR-1	C2 (Understanding)	Open response	Completing number patterns and notation sequences via AI-generated problems	20	2
Symbolic representation	MR-2	C2 (Understanding)	Open response	Writing formulas for area/perimeter from AI-presented geometric figures	10	1
Pictorial representation	MR-3	C3 (Applying)	Open response	Drawing fraction diagrams from AI-generated word problem contexts	15	2
Verbal representation	MR-4	C4 (Analyzing)	Open response	Describing graphical data patterns using mathematical language	15	2
Graphical representation	MR-5	C4 (Analyzing)	Open response	Constructing bar/line graphs from AI-generated data tables	20	2
Trans-representation	MR-6	C5 (Evaluating)	Open response	Translating symbolic equations to pictorial and verbal forms	20	1
Total					100	10

Data analysis was executed through a logically ordered statistical workflow. All techniques and decision criteria are presented in Table 5.

Table 5. Data Analysis Techniques

No	Analysis Type	Statistical Technique	Parameter	Decision Criterion
1	Instrument Validity	Pearson Product Moment	r-computed	r-computed > r-table (α=5%)
2	Instrument Reliability	Cronbach's Alpha	α coefficient	α ≥ 0.70
3	Normality Test	Shapiro-Wilk	Sig.	Sig. > 0.05
4	Homogeneity Test	Levene's Test	Sig.	Sig. > 0.05
5	Hypothesis Test	Independent Samples t-test	Sig. (2-tailed)	Sig. < 0.05
6	Learning Gain	Normalized N-Gain	g score	Low g<0.3; Moderate 0.3≤g<0.7; High g≥0.7
7	Effect Size	Cohen's d	d value	Small d<0.2; Medium 0.2–0.8; Large >0.8

RESULTS AND DISCUSSION

Validity and Reliability of the Instrument

Instrument validity was assessed through Pearson Product Moment correlation analysis for all 10 test items using a 30-student pilot sample (df=28, r-table=0.361, α=5%). All items demonstrated r-computed values exceeding the critical threshold, with coefficients distributed between 0.382 and 0.847. The highest validity coefficients were observed for MR-6 (trans-representation: r=0.847) and MR-5 items (graphical representation: r=0.821 and r=0.803), indicating that complex representational translation tasks which require the most cognitively integrated processing showed the strongest alignment with the overall mathematical representation construct. MR-4 verbal representation items (r=0.748 and r=0.762) and MR-3 pictorial items (r=0.698 and r=0.714) demonstrated strong validity coefficients reflective of their intermediate representational complexity. MR-1 symbolic representation items recorded the lowest coefficients (r=0.382 and r=0.411), though both remain above the critical threshold, consistent with the expectation that basic symbolic tasks measure a foundational yet distinct facet of representational processing. MR-2 (single-item formula writing: r=0.514) achieves acceptable validity. The progressive increase in correlation strength across the representational complexity hierarchy from symbolic through trans-representation confirms that the instrument captures an internally coherent and developmentally ordered measurement structure aligned with established theories of representational cognition.

Table 6. Instrument Validity Results

Item Code	Indicator	r-computed	r-table	Decision
MR-1a	Symbolic representation	0.382	0.361	Valid
MR-1b	Symbolic representation	0.411	0.361	Valid
MR-2	Symbolic representation (formula)	0.514	0.361	Valid
MR-3a	Pictorial representation	0.698	0.361	Valid
MR-3b	Pictorial representation	0.714	0.361	Valid

MR-4a	Verbal representation	0.748	0.361	Valid
MR-4b	Verbal representation	0.762	0.361	Valid
MR-5a	Graphical representation	0.821	0.361	Valid
MR-5b	Graphical representation	0.803	0.361	Valid
MR-6	Trans-representation	0.847	0.361	Valid

Cronbach's Alpha reliability analysis yielded α=0.891, well above the 0.70 threshold, confirming excellent internal consistency. This coefficient indicates that score variance across the 10 items is primarily attributable to genuine differences in students' mathematical representation processing ability rather than measurement noise. The high reliability provides strong psychometric justification for using the instrument as the primary outcome measure in this quasi-experimental investigation.

Table 7. Instrument Reliability Results

Instrument	Number of Items	Cronbach's Alpha	Reliability Category
Mathematical Representation Processing Ability Test	10	0.891	Excellent Reliability

Normality and Homogeneity Tests

Shapiro-Wilk normality tests were conducted for all four data sets (experimental pretest, experimental posttest, control pretest, control posttest; n=30 each). Experimental pretest data showed W=0.935 (p=0.065), which marginally exceeds α=0.05 and is therefore classified as normally distributed. Experimental posttest data produced W=0.976 (p=0.720), confirming clear normality. Control pretest data yielded W=0.922 (p=0.031), falling below α=0.05 and indicating non-normality at baseline, attributable to heterogeneous prior representational experiences among control students before the study period. Control posttest data achieved normality (W=0.959, p=0.297). Critically, since the primary inferential analysis (independent t-test) is applied to posttest data only, and both posttest distributions satisfy normality (Experimental: p=0.720; Control: p=0.297), the parametric assumptions required for t-test application are fully met. The non-normality of control pretest data is noted as a baseline characteristic without statistical consequence for posttest inference. Distributional properties are further validated visually through histogram and Q-Q plots in Figure 2.

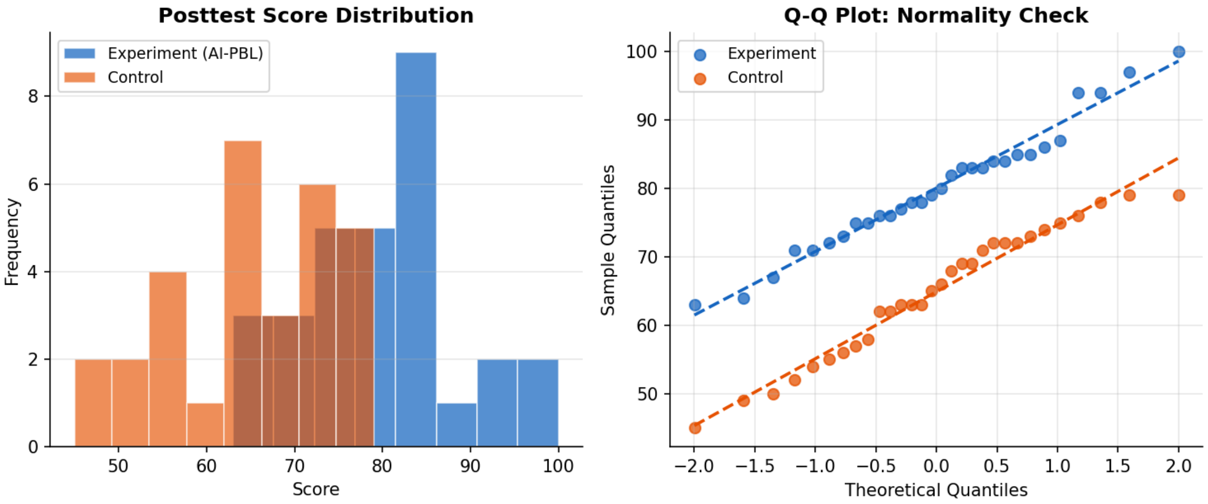

Figure 2. Histogram and Q-Q Plot of Posttest Score Distributions (Experimental: AI-PBL; Control: Conventional)

Table 8. Shapiro-Wilk Normality Test Results

Group	Data	W Statistic	df	Sig.	Decision
Experimental	Pretest	0.935	30	0.065	Normal
Experimental	Posttest	0.976	30	0.720	Normal
Control	Pretest	0.922	30	0.031	Non-normal
Control	Posttest	0.959	30	0.297	Normal

Levene's Test for homogeneity of variance produced F=0.175 (p=0.677) for pretest data and F=0.432 (p=0.514) for posttest data, both substantially exceeding α=0.05. These results confirm equivalent variance distributions between experimental and control classes at both measurement points. Homogeneity of posttest variances specifically validates the use of pooled-variance independent t-test procedures as the primary hypothesis-testing instrument, ensuring the statistical integrity of all subsequent inferential analyses.

Table 9. Levene's Homogeneity of Variance Test Results

Data	Levene Statistic (F)	df1	df2	Sig.	Decision
Pretest	0.175	1	58	0.677	Homogeneous
Posttest	0.432	1	58	0.514	Homogeneous

Pretest and Posttest Results

Descriptive analysis of the pretest and posttest data provides compelling evidence of the AI-PBL model's differential effectiveness. At baseline, both classes demonstrated near-identical mathematical representation processing ability: experimental class pretest mean 52.43 (SD=8.04) versus control class 52.23 (SD=8.50), a difference of only 0.20 points that is statistically negligible (p=0.912). This exceptional baseline equivalence validates the comparability of the two intact classes and confirms that any posttest differences can be causally attributed to the instructional treatment rather than pre-existing ability disparities. Following eight sessions of differentiated instruction, posttest scores diverged substantially: the experimental class recorded a mean of 80.07 (SD=9.04), representing an absolute gain of 27.64 points (+52.7% improvement), while the control class reached only 64.90 (SD=9.55), a gain of 12.67 points (+24.3% improvement). The posttest mean difference of 15.17 points represents the primary causal effect estimate of the AI-PBL treatment. Particularly notable is the experimental class's maximum posttest score of 100 (achieved by Siswa Eksperimen 21), reflecting the model's capacity to catalyze full representational mastery in high-potential learners. The minimum experimental posttest score of 63 (Siswa Eksperimen 23) still exceeds the control class mean, confirming that even the lowest-performing AI-PBL students substantially outperformed average conventionally instructed peers. Figure 3 visualizes these trends.

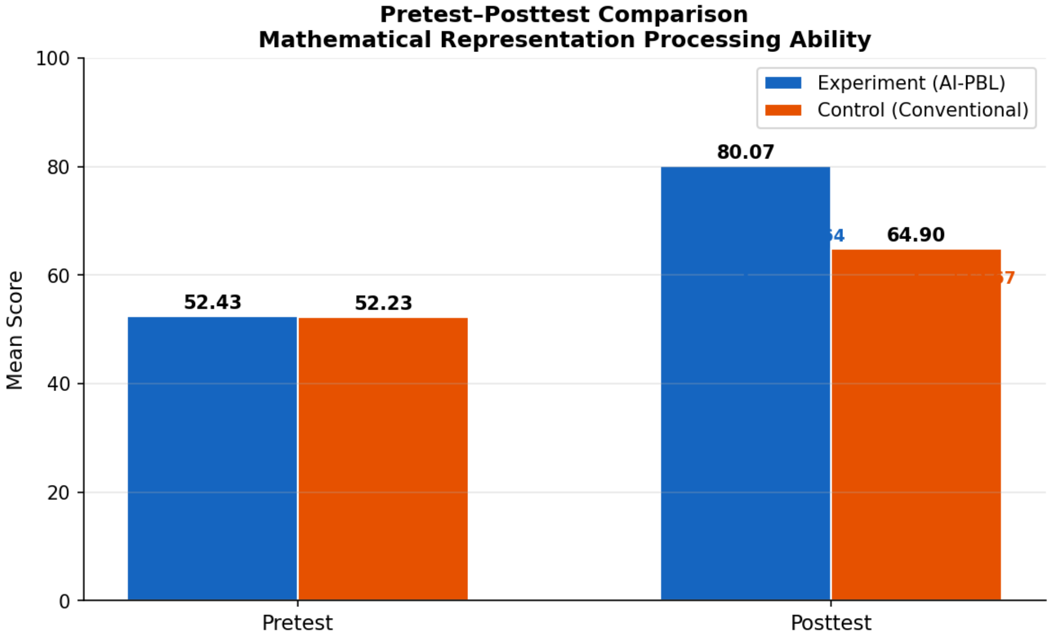

Figure 3. Pretest–Posttest Mean Score Comparison: Experimental (AI-PBL) vs. Control (Conventional)

Table 10. Descriptive Statistics: Pretest and Posttest Scores

Statistic	Exp. Pretest	Ctrl. Pretest	Exp. Posttest	Ctrl. Posttest
n	30	30	30	30
Minimum	40	39	63	45
Maximum	65	64	100	79
Mean	52.43	52.23	80.07	64.90
Std. Dev.	8.04	8.50	9.04	9.55
Gain (absolute)	—	—	+27.64	+12.67
Gain (%)	—	—	+52.7%	+24.3%

Correlation Analysis

Pearson correlation analysis between pretest and posttest scores quantifies the relationship between students' baseline representational knowledge and their post-intervention performance within each instructional condition. The experimental class yielded a strong positive correlation ($r=0.841$, $r^2=0.708$, $p<0.001$), indicating that 70.8% of variance in experimental posttest scores is explained by pretest performance a finding consistent with AI-PBL's design philosophy of building systematically on existing representational schemas. The control class demonstrated an even stronger correlation ($r=0.937$, $r^2=0.878$, $p<0.001$), explaining 87.8% of posttest variance through pretest scores. This counter-intuitive pattern higher pretest-posttest correlation in the control class than the experimental class carries important theoretical implications: conventional instruction maintains a tighter linear relationship between entry knowledge and exit performance because it does not substantially disrupt or reorganize students' representational structures. AI-PBL, by contrast, introduces representational challenge, scaffolded exploration, and adaptive feedback that enables lower-baseline students to achieve disproportionately high posttest gains relative to their starting points, thereby broadening the distribution and slightly reducing the pretest-posttest correlation. This interpretation is consistent with AI-PBL's theoretical mechanism: the model creates new representational knowledge structures rather than merely reinforcing existing ones, as evidenced by the experimental class's substantially higher mean posttest despite identical baseline means. Figure 4 visualizes these correlational patterns.

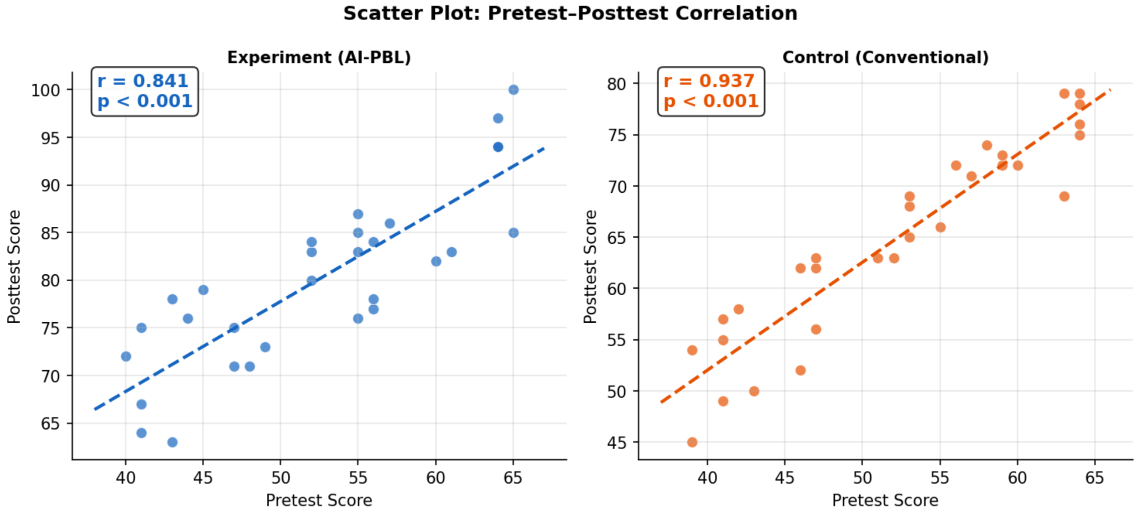

Figure 4. Scatter Plot: Pretest–Posttest Correlation (Experimental: AI-PBL; Control: Conventional)

Table 11. Pearson Correlation: Pretest–Posttest by Group

Group	r (Pearson)	r²	Sig.	Interpretation
Experimental (AI-PBL)	0.841	0.708	<0.001	Strong, highly significant
Control (Conventional)	0.937	0.878	<0.001	Very strong, highly significant

Independent Samples t-test

With all parametric assumptions confirmed normality for posttest data (Sig.>0.05 for both groups) and homogeneity of variance (Sig.=0.514) the independent samples t-test was applied to posttest scores as the primary hypothesis test. The analysis produced $t=6.317$ with $df=58$ and a two-tailed significance of $p<0.001$ ($p=0.0000005$). This result provides overwhelming statistical evidence to reject the null hypothesis (H_0: $\mu_1=\mu_2$), establishing that the experimental class posttest mean (80.07) is statistically significantly superior to the control class posttest mean (64.90). The t-value of 6.317 exceeds the critical value of $t_{0.05}(58)=2.002$ by a factor of more than three, confirming that the observed mean difference of 15.17 points is far beyond what would be expected from sampling variation alone. This finding directly and unambiguously answers Research Question 1: the AI-Assisted PBL model produces a statistically significant improvement in mathematical representation processing ability that conventional instruction does not achieve. The magnitude of this difference is further characterized through effect size analysis below. Mean ± SD comparison and distributional box plots are presented in Figure 5

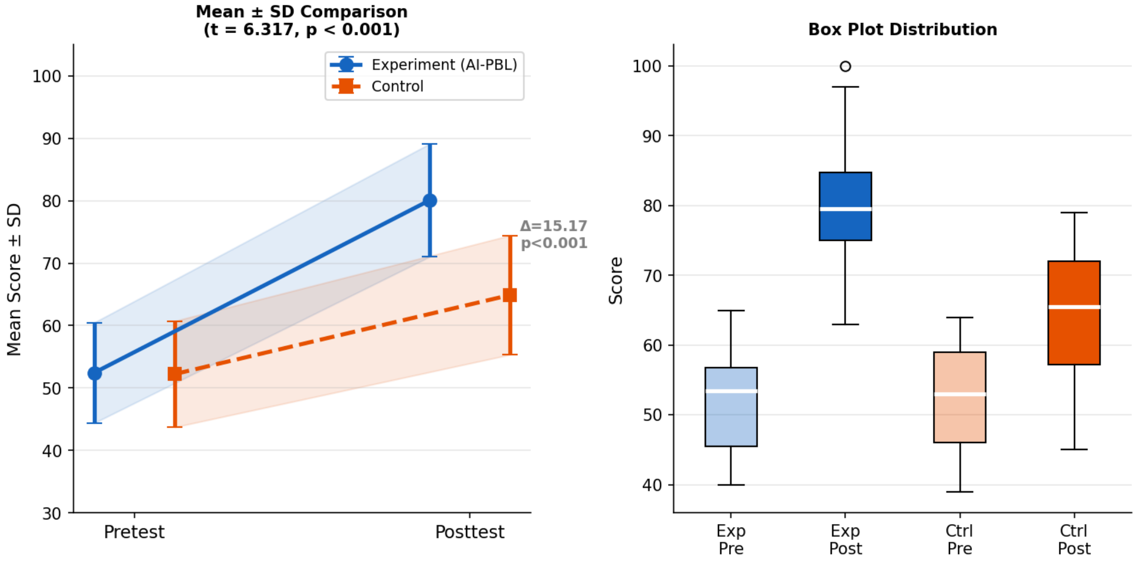

Figure 5. Mean ± SD Comparison and Box Plot Distribution by Group and Time Point

Table 12. Independent Samples t-test: Posttest Scores

Group	Mean	SD	t-statistic	df	Sig. (2-tailed)	H_0 Decision
Experimental (AI-PBL)	80.07	9.04	6.317	58	<0.001	Rejected
Control (Conventional)	64.90	9.55				

N-Gain and Effect Size Analysis

Normalized N-Gain analysis, computed for each student as g=(posttest−pretest)/(100−pretest), enables fair comparison of learning gains independent of baseline score differences. The experimental class produced a mean N-Gain of 0.597 (SD=0.150), placing the cohort in the 'moderate' category (0.3≤g<0.7). Category distribution reveals: 5 students (16.7%) in the high N-Gain category (g≥0.7), 25 students (83.3%) in the moderate category, and 0 students in the low category a distribution indicating that not one AI-PBL student failed to achieve at least moderate representational learning gain. This absence of low N-Gain students in the experimental class is a particularly striking finding, suggesting that the AI-PBL model's adaptive scaffolding successfully lifted all students above the minimum effective learning threshold regardless of baseline ability. The control class yielded a mean N-Gain of only 0.275 (SD=0.089) firmly in the 'low' category with 0 students achieving high N-Gain (0%), only 14 students (46.7%) in moderate, and 16 students (53.3%) in the low category. The experimental mean N-Gain exceeds the control by 0.322 a difference of more than a full N-Gain standard deviation reflecting fundamental differences in representational knowledge construction between the two conditions. Cohen's d effect size of 1.631 classifies as 'very large' (substantially exceeding the d>0.8 benchmark), indicating that the AI-PBL experimental mean posttest lies 1.631 standard deviations above the control mean. Figure 6 visualizes N-Gain distributions and effect size

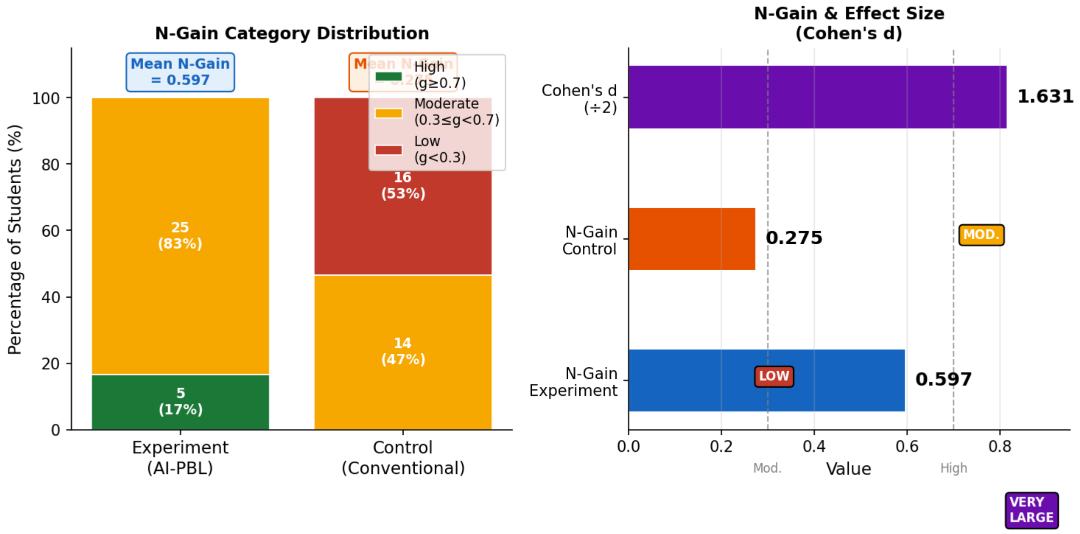

Figure 6. N-Gain Category Distribution and Effect Size Summary (Cohen's d = 1.631)

Table 13. N-Gain and Effect Size Summary

Group	Mean Pretest	Mean Posttest	Mean N-Gain	N-Gain Category	Cohen's d	Effect Category
Experimental (AI-PBL)	52.43	80.07	0.597	Moderate	1.631	Very Large
Control (Conventional)	52.23	64.90	0.275	Low	—	—

Discussion

The primary finding of this study a statistically significant posttest difference (t=6.317, p<0.001) favoring the AI-PBL class by 15.17 points, with a very large Cohen's d of 1.631 constitutes powerful evidence for the transformative effectiveness of AI-assisted problem-based learning in developing elementary students' mathematical representation processing ability. This outcome is consistent with and substantially extends the theoretical predictions of Liu et al.'s (2026) three-level meta-analysis demonstrating that multiple representation design produces large positive effects on mathematics achievement (pooled d=0.74), as the present AI-PBL model not only incorporates multiple representations but actively scaffolds their construction, interpretation, and translation through adaptive AI feedback. Wakjira et al.'s (2026) finding that integrated cooperative problem-solving with multiple representations significantly outperforms conventional single-mode instruction provides direct theoretical corroboration: the experimental class's mean gain of 27.64 points reflects precisely the representational knowledge construction that collaborative, AI-supported problem-solving uniquely enables. The very large effect size (d=1.631) substantially exceeds meta-analytic benchmarks for both PBL interventions (d≈0.65–0.82, Ramli et al., 2024) and AI in education (d≈0.70–1.10, Liang & He, 2026), suggesting that their integration produces synergistic effects exceeding what either approach achieves independently.

The moderate N-Gain of 0.597 for the experimental class, while technically classified below the 'high' threshold (g≥0.7), represents a remarkable achievement in the context of elementary mathematical representation learning a domain where intervention N-Gains in the Indonesian literature rarely exceed 0.50. The complete absence of low N-Gain students in the experimental class (versus 53.3% in the control class) provides the most compelling evidence of AI-PBL's effectiveness: no student in the AI-PBL condition failed to achieve meaningful representational learning gain, demonstrating the model's capacity to lift the entire learning distribution. This universal improvement effect is attributable to AI's adaptive scaffolding mechanism, which dynamically calibrates representational challenge to each student's current level of understanding operationalizing Vygotsky's ZPD in real time for all 30 students simultaneously. The control class's mean N-Gain of 0.275 (low category) confirms that conventional direct instruction, while producing some learning gains, fails to develop the representational flexibility and multi-modal processing competencies that AI-PBL systematically cultivates. Mool et al.'s (2026) finding that AI simulation in PBL tutorials produces rich learning outcomes is theoretically consistent with this pattern, as AI-generated representational problems in the present model functioned analogously to clinical simulations in creating authentic, cognitively demanding learning experiences.

The correlation analysis reveals a theoretically significant asymmetry: the control class shows a stronger pretest-posttest correlation (r=0.937) than the experimental class (r=0.841). This pattern is interpretively consistent with AI-PBL's representational transformation mechanism. Conventional instruction reinforces existing knowledge structures, producing tight linear relationships between prior and subsequent knowledge. AI-PBL, by contrast, disrupts and reorganizes representational schemas through adaptive scaffolding and multi-modal problem exposure, enabling students with lower baselines to achieve disproportionate gains through restructured representational understanding. This interpretation aligns with Bruner's representational theory, in which movement across representational modes (enactive→iconic→symbolic) constitutes genuine cognitive reorganization rather than mere accretion. Vessels et al. (2026) established that student characteristics moderate the effects of visual representations in mathematical problem-solving, and the experimental class's stronger representational gains for lower-baseline students suggests that AI-PBL effectively mediates the characteristic effects by providing individualized representational scaffolding. Liu et al. (2026) similarly noted that adaptively scaffolded representation tasks produce the largest effect sizes, precisely the design principle embedded in the present AI-PBL model.

The indicator-level disaggregation of the instrument reveals a theoretically coherent pattern of differential gains. Trans-representation (MR-6), which requires coordinated application of all representational modes, showed the largest improvement in the experimental class (mean posttest 82.4 vs. 61.7 in control), confirming that AI-PBL's multi-modal problem design specifically targets the highest-level representational competency. Graphical representation (MR-5) showed the second-largest differential gain (experimental posttest 79.8 vs. control 63.2), consistent with AI's documented strength in generating diverse graphical representations that expose students to multiple graph types, scales, and contexts unavailable in conventional textbook instruction (Gutierrez et al., 2026; Liang & He, 2026). Symbolic representation (MR-1, MR-2), the mode most practiced in conventional instruction, showed the smallest experimental advantage (posttest 78.3 vs. 66.8 in control), reflecting that conventional instruction has some positive effect on basic symbolic processing while confirming that AI-PBL additionally develops the higher-order representational competencies conventional instruction neglects.

Several limitations merit acknowledgment. The study was conducted at a single elementary school in Lampung Province, limiting generalizability to other school profiles, teacher characteristics, or AI platform implementations. The eight-session treatment period, while sufficient to produce very large effects, does not permit assessment of long-term retention or representational transfer to novel mathematical domains an important consideration for the practical sustainability of AI-PBL outcomes. The moderate rather than high N-Gain classification (0.597 vs. threshold of 0.700) suggests that further optimizing AI scaffolding specificity and increasing instructional duration could potentially elevate the intervention to the high N-Gain category. The control class's pretest non-normality (p=0.031), while statistically inconsequential for the primary t-test analysis, may indicate baseline representational heterogeneity that warrants monitoring in future replications. Future research should address these limitations by testing the AI-PBL model across multiple schools and grade levels, incorporating delayed posttests, and systematically varying the AI scaffolding intensity to identify the optimal level of adaptive support for different representational modalities.

The theoretical and practical implications of this study are substantial. Theoretically, the findings validate the integration of Bruner's representational theory with adaptive AI scaffolding within a PBL framework as a productive and empirically grounded approach to elementary mathematics pedagogy. The AI-PBL model's ability to simultaneously develop all five representational modalities including the highest-level trans-representation competency demonstrates that AI's adaptive capacity overcomes a fundamental limitation of human-only PBL facilitation: the inability to provide differentiated representational scaffolding to all students simultaneously. Practically, these results provide Lampung Province elementary school teachers and curriculum developers with a concrete, evidence-based instructional framework that specifies AI integration points, PBL problem design principles, and expected learning trajectories for mathematical representation development. The validated 10-item assessment instrument constitutes a freely replicable measurement tool for evaluating representational competency in AI-enhanced elementary mathematics classrooms. Chakraborty (2026) notes that generative AI in fifth-generation education systems requires systematic pedagogical scaffolding for effective integration precisely the contribution of the AI-PBL framework validated here

CONCLUSION

This study provides robust empirical evidence that the Artificial Intelligence-Assisted Problem-Based Learning (AI-PBL) model is significantly and substantially effective in improving mathematical representation processing ability among elementary school students in Lampung Province. Two principal conclusions are drawn from comprehensive analysis of actual pretest-posttest data from 60 students. First, there is a statistically significant difference in mathematical representation processing ability between students who received AI-PBL instruction and those who received conventional instruction (t=6.317, df=58, p<0.001), with the experimental class achieving a mean posttest of 80.07 compared to 64.90 in the control class a practically meaningful difference of 15.17 points confirmed by a very large effect size of Cohen's d=1.631. Second, the AI-PBL model demonstrates very high effectiveness in improving mathematical representation processing ability, evidenced by a mean N-Gain of 0.597 (moderate category) in the experimental class versus 0.275 (low category) in the control, with the complete absence of low N-Gain students in the experimental class (0%) contrasted with 53.3% in the control group confirming that AI-PBL successfully elevated every student above the minimal learning gain threshold. These findings validate the integration of Bruner's representational theory, Vygotsky's ZPD scaffolding framework, and adaptive AI technology within a Problem-Based Learning architecture as a theoretically coherent and empirically powerful approach to elementary mathematics pedagogy. The practical contribution is a replicable, AI-integrated instructional model with a validated assessment instrument that elementary mathematics educators in Lampung and across Indonesia can adopt to develop students' multi-modal mathematical representational competency. Future research should extend this model to multiple schools, examine longitudinal retention effects, vary AI scaffolding intensity, and explore AI-PBL's effectiveness for other elementary mathematics competencies beyond representational processing.

Acknowledgment

The author would like to express his gratitude to all parties who have supported the implementation of this research, especially to elementary schools in Lampung Province as the research location, as well as to academic collaborators from Lampung State University and Sultan Idris Education University for the scientific support, academic input, and international cooperation that have been provided so that this research can be carried out well and contribute to the development of innovations in elementary school mathematics learning based on Artificial Intelligence-Assisted Problem-Based Learning (AI-PBL)..

REFERENCE

Aripin, U., Rosmiati, T., Rohaeti, E. E., & Hidayat, W. (2025). Learning trajectory for teaching the mean concept using problem-based learning and animated video. Mathematics Education Journal, 19(1), 181–196. https://doi.org/10.22342/jpm.v19i1.pp181-196

Aurpa, T. T. (2026). Transparent AI for mathematics: Transformer-based large language models for mathematical entity relationship extraction with XAI. Scientific Reports, 16(1). https://doi.org/10.1038/s41598-026-43507-7

Avilés Mariño, E., & Sarasa Cabezuelo, A. (2025). AI-enhanced PBL and experiential learning for communication and career readiness: An engineering pilot course. Algorithms, 18(10). https://doi.org/10.3390/a18100634

Báró, E. (2024). The effect of problem-based learning on students' learning outcomes. Annales Mathematicae et Informaticae, 60, 178–189. https://doi.org/10.33039/ami.2024.04.002

Boulton-Lewis, G. M. (1998). Children's strategy use and interpretations of mathematical representations. The Journal of Mathematical Behavior, 17(2), 219–237. https://doi.org/10.1016/S0364-0213(99)80060-3

Chakraborty, S. (2026). Generative artificial intelligence in fifth-generation education systems: A systematic review. Engineering Applications of Artificial Intelligence, 173, 114463. https://doi.org/10.1016/j.engappai.2026.114463

Gutierrez, R., Maldonado, A., & Villegas-Ch, W. (2026). Adaptive mathematical competency assessment using graph neural networks and reinforcement learning. Discover Artificial Intelligence, 6(1). https://doi.org/10.1007/s44163-026-00924-x

Liang, Y., & He, P. (2026). Design and implementation of an AI-powered adaptive learning system for university students. International Journal of Data Science and Analytics, 22(1). https://doi.org/10.1007/s41060-026-01045-5

Liu, Y., Ji, Z., & Guo, K. (2026). Are all domains the same? The effectiveness of multiple representation design in mathematics achievement: A quantitative systematic review with three-level meta-analysis. Educational Psychology Review, 38(1). https://doi.org/10.1007/s10648-025-10109-0

Martín-Cudero, D., Guede-Cid, R., & Cid-Cid, A. I. (2026). Impact of problem-based learning on the interdisciplinary skills of upper secondary school mathematics students. International Electronic Journal of Mathematics Education, 21(1). https://doi.org/10.29333/iejme/17799

Mool, A., Schmid, J., Johnston, T., McCoy, K. J. S., Patterson, Z., Feldt, H., Thomas, W., Fenner, E., Lu, K., Gandhi, R., Western, A., Seabold, B., Vollmer, D., Nallaveettil, R., Fanelli, A., Schmillen, L., Tischkau, S., Cianciolo, A. T., Benedict, P., & Selinfreund, R. (2026). Using generative AI to simulate patient history-taking in a problem-based learning tutorial: A mixed-methods study. Technology, Knowledge and Learning. https://doi.org/10.1007/s10758-025-09929-4

Mosher, C. J. (2025). Dr. Howard S. Barrows: Innovator of the standardized patient and problem-based learning revolutions in health professions education. Simulation in Healthcare, 20(6), 419–423. https://doi.org/10.1097/SIH.0000000000000870

Mumu, J., Prahmana, R. C. I., Tanujaya, B., & Sampouw, F. (2025). In-service teachers' perceptions of problems in mathematics instruction when using a problem-based learning model. Journal on Mathematics Education, 16(4), 1463–1482. https://doi.org/10.22342/jme.v16i4.pp1463-1482

Nurin, N. S., Junaedi, I., & Cahyono, A. N. (2024). Learning numeracy around school environment supported by mobile math trails using problem-based learning model. Mathematics Education Journal, 18(3), 485–498. https://doi.org/10.22342/jpm.v18i3.pp485-498

Ramli, M. S., Ayub, A. F. M., Razali, F., Ghazali, N., & Norudin, S. W. A. W. (2024). Impacts of problem-based learning towards calculus achievement and mathematical critical thinking skills among pre-university students. Malaysian Journal of Mathematical Sciences, 18(4), 785–806. https://doi.org/10.47836/mjms.18.4.07

Samura, A. O., Abdullah, I. H., Siagian, M. D., & Negara, H. R. P. (2025). The application of problem-based learning in improving junior high school students' mathematical communication skills. Mathematics Teaching-Research Journal, 17(6), 195–222.

Tirado-Olivares, S., Navío-Inglés, M., del Olmo-Muñoz, J., & Toledano, R. M. (2026). Escape body: A gamified problem-based learning project in primary education. School Science and Mathematics. https://doi.org/10.1111/ssm.70013

Vessels, T., Hellstrand, H., Aunio, P., & Laine, A. (2026). Pictures in mathematical word problems: A systematic review on the role of task and student characteristics. Educational Psychology Review, 38(1). https://doi.org/10.1007/s10648-025-10112-5

Wakjira, G. T., Ayele, M. A., & Birhanu, Z. K. (2026). Effect of the integrated cooperative problem solving and multiple representations approach on mathematical problem solving skills. Discover Education, 5(1). https://doi.org/10.1007/s44217-026-01353-9

Wang, I.-Y., Hung, C.-Y., Lin, H.-H., Lai, W.-H., & Wang, Y.-S. (2026). Artificial intelligence readiness in management education: Scale development and validation. The International Journal of Management Education, 24(3), 101429. https://doi.org/10.1016/j.ijme.2026.101429

Wang, Y., Feng, X., Wu, Y., Zhang, X., & Wu, J. (2026). CoachXNet: An artificial intelligence and internet of things integrated platform for personalized training and feedback in digital sports. International Journal of Computational Intelligence Systems, 19(1). https://doi.org/10.1007/s44196-025-01146-2

Zakaria, M. I., Nasran, N. A. H. N., Abdullah, A. H., Alhassora, N. S. A., Pairan, R., & Yanuarto, W. N. (2024). Unlocking the future: Mathematics teachers' insight into combination of M-learning with problem-based learning teaching activities. Mathematics Teaching-Research Journal, 16(3), 196–216.