



OASIS: Jurnal Ilmiah Kajian Islam

Volume 11 Nomor 1 Tahun 2026

<https://www.syekh Nurjati.ac.id/jurnal/index.php/oasis>

ISSN 1979-1399| e-ISSN 2548-3889

Comparing Problem-Based Learning and Conventional Instruction on Aqidah Akhlak Learning Outcomes: A Quasi-Experimental Study at an Indonesian Senior Secondary School

Aldinda Fauziah

Institut Islam Mamba'ul 'Ulum Surakarta, Indonesia

fauziaaldinda@gmail.com

Nur Hidayah

Institut Islam Mamba'ul 'Ulum Surakarta, Indonesia

nurhidayahsyafii@gmail.com

Alwan Arhan

University of Sains and Technology, Yemen

arhanalwan06@gmail.com

Abstract

This study was motivated by the limited number of studies comparing the effectiveness of Problem-Based Learning (PBL) and conventional instruction in Aqidah Akhlak education at the senior secondary school level. The purpose of this study was to analyze and compare the Aqidah Akhlak learning outcomes of students taught using the PBL model and those taught through conventional instruction at SMA Muhammadiyah 3 Surakarta. The study employed a quantitative approach using a quasi-experimental Pretest–Posttest Control Group Design. The sample consisted of 68 students selected through purposive sampling, including 34 students in the experimental group and 34 students in the control group. Data were collected using a multiple-choice achievement test that met established validity and reliability criteria. Data analysis involved descriptive statistics, normality and homogeneity tests, Independent Samples t-tests, Paired Samples t-tests, and effect size analysis. The findings indicate that both instructional models improved students' learning outcomes. The mean score of the PBL group increased from 79.21 to 90.15, while that of the conventional instruction group increased from 74.88 to 89.26. However, the Independent Samples t-test revealed no statistically significant difference between the two groups ($p = 0.497 > 0.05$). This result was supported by a Cohen's d value of 0.20 (small effect) and an Eta Squared value of 0.007 (negligible effect), indicating that the influence of the instructional model on differences in learning outcomes was minimal. Therefore, Problem-Based Learning and conventional instruction demonstrated relatively equivalent effectiveness in improving Aqidah Akhlak learning outcomes. This study contributes to the development of scholarship in Islamic Religious Education and provides practical insights for teachers in selecting instructional strategies that align with subject characteristics and students' learning needs.

Keywords: Problem-Based Learning; Conventional Instruction; Learning Outcomes; Aqidah Akhlak; Quasi-Experimental Research

Abstrak

Penelitian ini dilatarbelakangi oleh terbatasnya studi yang membandingkan efektivitas Problem-Based Learning (PBL) dan pembelajaran konvensional dalam mata pelajaran Aqidah Akhlak di tingkat sekolah menengah atas. Tujuan penelitian ini adalah untuk menganalisis dan membandingkan hasil belajar Aqidah Akhlak siswa yang diajar menggunakan model PBL dengan siswa yang diajar melalui pembelajaran konvensional di SMA Muhammadiyah 3 Surakarta. Penelitian ini menggunakan pendekatan kuantitatif dengan desain kuasi-eksperimen Pretest–Posttest Control Group Design. Sampel penelitian terdiri dari 68 siswa yang dipilih melalui teknik purposive sampling, yang mencakup 34 siswa dalam kelompok eksperimen dan 34 siswa dalam kelompok kontrol. Data dikumpulkan menggunakan tes hasil belajar berbentuk pilihan ganda yang telah memenuhi kriteria validitas dan reliabilitas. Analisis data meliputi statistik deskriptif, uji normalitas dan homogenitas, uji-t sampel independen (Independent Samples t-test), uji-t sampel berpasangan (Paired Samples t-test), serta analisis ukuran efek (effect size). Temuan penelitian menunjukkan bahwa kedua model pembelajaran tersebut meningkatkan hasil belajar siswa. Skor rata-rata kelompok PBL meningkat dari 79,21 menjadi 90,15, sedangkan skor rata-rata kelompok pembelajaran konvensional meningkat dari 74,88 menjadi 89,26. Namun, uji-t sampel independen menunjukkan tidak adanya perbedaan yang signifikan secara statistik antara kedua kelompok tersebut ($p = 0,497 > 0,05$). Hasil ini didukung oleh nilai Cohen's d sebesar 0,20 (efek kecil) dan nilai Eta Squared sebesar 0,007 (efek yang dapat diabaikan), yang mengindikasikan bahwa pengaruh model pembelajaran terhadap perbedaan hasil belajar tergolong minimal. Dengan demikian, Problem-Based Learning dan pembelajaran konvensional menunjukkan tingkat efektivitas yang relatif setara dalam meningkatkan hasil belajar Aqidah Akhlak. Penelitian ini berkontribusi pada pengembangan keilmuan Pendidikan Agama Islam serta memberikan wawasan praktis bagi guru dalam memilih strategi pembelajaran yang sesuai dengan karakteristik mata pelajaran dan kebutuhan belajar siswa.

Kata Kunci: Problem-Based Learning; pembelajaran konvensional; hasil belajar; Aqidah Akhlak; penelitian kuasi-eksperimen

Introduction

Aqidah Akhlak is a distinctive component of Islamic education because it joins knowledge of belief with the language of moral judgment. At senior-secondary level, students are expected not only to recognize concepts such as tawhid, shirk, Asmaul Husna, and commendable or reprehensible conduct, but also to distinguish among concepts and apply them to situations that contain competing motives and consequences. When instruction is dominated by explanation and reproduction of definitions, students may achieve factual recall without receiving enough opportunity to articulate reasons, examine a moral situation, or test the coherence of an answer. The central educational problem is therefore not a simple opposition between teacher explanation and student activity. It is how to preserve doctrinal and conceptual accuracy while enabling students

to engage intellectually with the problems through which religious and moral knowledge acquires practical meaning.

Problem-Based Learning (PBL) is relevant to this problem because it uses a meaningful, often ill-structured problem as the organizing stimulus for learning. Students clarify what is known, identify what must be learned, gather information, propose explanations, and communicate a defensible response with peers. The teacher does not withdraw from instruction; the teacher designs the problem, monitors inquiry, prompts justification, and provides scaffolding when students' existing knowledge is insufficient (Hmelo-Silver 2004; Yew and Goh 2016). This combination of student inquiry and purposeful facilitation differentiates PBL from minimally guided discovery. Meta-analytic work on inquiry-based learning shows that guidance improves learning processes and outcomes, while research on scaffolding in PBL likewise reports benefits for conceptual and higher-order performance (Lazonder and Harmsen 2016; Kim, Belland, and Walker 2018).

The mechanisms proposed for PBL are cognitive, motivational, and social. A problem can activate relevant prior knowledge and expose gaps in understanding; subsequent inquiry promotes elaboration, organization, and retrieval of new information. Small-group explanation requires students to make their reasoning public, while a contextual problem may generate situational interest that sustains effort (Schmidt, Rotgans, and Yew 2011). Reviews further associate well-designed PBL with deeper approaches to learning, although the evidence is not uniform and depends on problem quality, assessment, facilitation, and the maturity of learners' self-directed learning skills (Dolmans et al. 2016). Positive social interdependence is also important: collaboration becomes productive when group members perceive that they have connected responsibilities rather than merely dividing a worksheet into isolated tasks (Shimizu et al. 2021). These mechanisms make PBL plausible for Aqidah Akhlak, where explanation, deliberation, and contextual judgment are pedagogically valuable.

In this study, PBL was represented by five observable instructional indicators: orientation to a relevant problem; organization of collaborative inquiry; collection and analysis of information; development and communication of a proposed solution; and reflection on both the solution and the reasoning process. These indicators are consistent with the broader PBL cycle, but they are not identical to Project-Based Learning. PBL is organized around understanding and responding to a problem, whereas Project-Based Learning normally culminates in a sustained product, performance, or artifact (Anazifa and Djukri 2017). Maintaining this distinction is important because the acronym PBL is sometimes used for both approaches. The present study concerns problem-based, not project-based, learning.

Learning outcomes are equally important to define. The intended outcomes of Islamic education can include knowledge, spiritual orientation, attitudes, habitual conduct, and participation in a moral community. The present research did not attempt to capture all those domains. Its dependent variable was cognitive Aqidah Akhlak achievement, operationalized through a 20-item multiple-choice test. The instrument

represented an Aqidah domain Islamic creed, Asmaul Husna, tawhid, shirk, and kalam and an Akhlak domain moral conduct, major virtues, commendable and reprehensible traits, moral development, and forms of virtuous behavior. The resulting scores indicate performance on curriculum-based cognitive tasks; they should not be treated as direct evidence of faith, moral internalization, or behavior. This construct boundary follows the principle that conclusions must remain aligned with the evidence produced by the measurement instrument (Ary et al. 2018).

The wider evidence base gives strong reasons to investigate PBL but insufficient reason to assume that it will always outperform structured instruction. Reviews and meta-analyses report positive average effects for science achievement, mathematical reasoning and problem solving, critical thinking, language learning, motivation, and several professional-education outcomes (Funa and Prudente 2021; Musna, Juandi, and Jupri 2021; Fitriyah, Putro, and Apino 2022; Liu and Pásztor 2022; Erdem 2024; Orhan 2025; Su et al. 2025; Wiratama et al. 2025). In nursing education, syntheses likewise indicate gains in critical thinking and other learning outcomes (Sharma et al. 2023; Wei et al. 2024). Yet a recent systematic review of reviews concludes that findings for theoretical knowledge are mixed and that instructor expertise, cultural context, subject matter, and outcome definitions moderate effectiveness (Amoa-Danquah and Carbonneau 2025). The pertinent question is consequently not whether PBL is universally effective, but under which instructional and measurement conditions its potential becomes visible.

Evidence from secondary education reinforces this conditional view. An experimental evaluation in secondary schooling showed that PBL effects can vary across student groups and outcomes, cautioning against transferring conclusions from university or medical curricula without examining school-level implementation (De Witte and Rogge 2016). Students who encounter PBL for the first time must also learn how to distribute responsibility, search selectively, and justify a response. Case evidence suggests that prior achievement, attendance, and familiarity with the model can shape what students gain from an initial PBL experience (Lee 2025). Thus, a short implementation may produce strong engagement or reasoning processes without necessarily generating a large advantage on a conventional cognitive test.

Research in Islamic education provides a more directly relevant but still developing evidence base. PBL has been associated with achievement, engagement, responsibility, empathy, social concern, and critical thinking in Islamic Religious Education and related settings (Rohman 2022; Dana 2024; Qorib 2025; Amanda and Ndzijia 2026; Hamid and Rahman 2026). Online PBL incorporating religious values has also been examined through a quasi-experimental design (Nawi and Zakaria 2019), while recent studies connect problem-centered learning with Islamic religious comprehension, ijihad-related reasoning in Fiqh, and integration with flipped learning (Abdul Khalil, Mohd Razif, and Rosele 2024; Nugraha, Abdel Rahim, and Lukman 2024; Amirudin, Muzaki, and Nurhayati 2025). These studies support contextual inquiry in religious education, but they differ in age group, subject, delivery mode, and outcome. Evidence specifically

comparing PBL and conventional instruction for senior-secondary Aqidah Akhlak remains limited.

That gap matters for two reasons. First, Aqidah Akhlak combines content that benefits from precise, structured explanation with situations that invite analysis and deliberation. Conventional instruction may therefore be efficient for organizing foundational knowledge, while PBL may add value through contextualization and explanation. Second, a between-group comparison in intact school classes must account for initial achievement differences. A posttest mean alone can conceal baseline non-equivalence, whereas a raw gain can exaggerate differences when one group starts lower and has more room to improve. Pretest measurement itself is also subject to error, which complicates automatic use of covariance adjustment in nonrandomized designs (Miyazaki et al. 2022). Transparent reporting of baseline, within-group change, final scores, effect sizes, and sensitivity checks is therefore essential.

The novelty of this study lies in its domain-specific and analytically cautious comparison. Rather than inferring effects from science, medicine, or general Islamic education, it examines curriculum-based cognitive achievement in Aqidah Akhlak at one Indonesian senior secondary school. It also separates three questions that are often conflated: whether achievement improved under PBL, whether achievement improved under conventional instruction, and whether final achievement differed between the two intact classes. The study addressed the following research questions: (1) To what extent did students' cognitive Aqidah Akhlak outcomes improve after PBL? (2) To what extent did they improve after conventional instruction? and (3) Was there a statistically significant difference in posttest outcomes between the PBL and conventional groups?

Method

The study used a quantitative quasi-experimental nonequivalent pretest-posttest control-group design. The instructional condition PBL or conventional instruction was the independent variable, and cognitive Aqidah Akhlak achievement was the dependent variable. Individual students were not randomly assigned because the study operated through existing classes in a natural school setting. Such a design can estimate patterns of change and group differences while retaining ecological validity, but it cannot provide the same protection from selection bias as individual randomization (Fraenkel, Wallen, and Hyun 2019). The design was represented as O1–X1–O2 for the PBL class and O1–X2–O2 for the conventional class, where O1 and O2 denote the common pretest and posttest and X denotes the respective instructional condition.

The analysis treated the two classes as nonequivalent from the outset. Pretest scores were therefore used to describe baseline comparability rather than to assume it. Within-group analyses addressed change under each instructional condition, whereas the primary between-group analysis addressed the observed posttest difference. These estimands answer different questions: a significant paired test shows that a group changed over time, but it does not show that its teaching model outperformed the other; a nonsignificant

posttest test indicates that a difference was not detected, but it does not prove statistical equivalence. This distinction guided both analysis and interpretation.

The research was conducted at SMA Muhammadiyah 3 Surakarta, Indonesia, during May–June 2025. The accessible population comprised 100 Grade 10 students distributed across classes XE1, XE2, and XE3. Two intact classes were purposively included in the outcome analysis because they studied the same grade-level Aqidah Akhlak curriculum in the same school context and were available for the instructional comparison. Class XE2 was designated as the PBL group ($n = 34$) and Class XE3 as the conventional group ($n = 34$). Class XE1 was used for pilot testing the achievement instrument and was excluded from the main outcome analysis.

The analytic sample therefore contained 68 students, evenly divided across conditions. Equal group sizes improve the stability of the independent-samples comparison when group variances are similar, but equal size does not remove selection bias. No student-level random assignment or matching was conducted. The sampling unit was effectively the intact class, and only one class represented each condition; consequently, class-specific influences cannot be cleanly separated from instructional-model influences. This design feature is reported as a boundary on causal generalization rather than hidden behind the label quasi-experiment.

Data were collected with a 20-item multiple-choice achievement test administered before and after instruction. The test was constructed from the Aqidah Akhlak curriculum and the learning objectives represented in the study. It sampled two broad domains: Aqidah, concerning conceptual understanding of Islamic belief, and Akhlak, concerning concepts and classifications of moral conduct. The same score scale was used for pretest and posttest, enabling direct description of change. Because the instrument was achievement-oriented and selected-response, it was best suited to estimating cognitive mastery and discrimination among concepts, not the quality of extended moral argument or the consistency of behavior outside the test situation.

Table 1. Curriculum Domains Represented in the Achievement Test

Domain	Indicators represented
Aqidah	Islamic creed; Asmaul Husna; tawhid; shirk; kalam
Akhlak	Moral conduct; major virtues; commendable and reprehensible traits; moral development; virtuous behavior

The instrument was reviewed for alignment with the intended content and piloted in the nonanalytic class. The archived item analysis reported a critical item-total correlation of $r = 0.339$. All 20 items exceeded that criterion, with observed coefficients from 0.467 to 0.663. The reported Cronbach's alpha was 0.846, indicating high internal consistency under the criterion used in the study (Hair et al. 2022). These statistics support use of the total score for group-level cognitive comparison. They do not, however, establish that the test measures affective commitment, transfer to authentic moral decisions, or stable behavior; those require additional methods and evidence

The PBL condition followed five connected phases. First, students were oriented to an Aqidah Akhlak problem and clarified its central issue. Second, the teacher organized collaborative work and helped students identify what information and concepts were needed. Third, students gathered and examined relevant information, relating it to the problem. Fourth, groups developed and communicated a proposed response supported by reasons. Fifth, students and teacher reflected on the response and evaluated the reasoning process. Teacher facilitation remained important for correcting conceptual misunderstanding, directing attention to relevant evidence, and keeping discussion aligned with the learning objectives.

The conventional condition addressed the same curricular content through structured explanation, question-and-answer interaction, and guided practice. This comparison was not a no-treatment control. It represented a credible instructional alternative capable of producing learning, especially where concepts require concise definition and authoritative clarification. Interpreting the study as PBL versus 'no learning' would therefore be incorrect. The design compared two active instructional conditions, each followed by the same achievement measure.

Table 2. Instructional Alignment and Contrasts

Component	PBL condition	Conventional condition	Shared element
Starting point	Contextual problem	Structured concept explanation	Aqidah Akhlak objectives
Student activity	Collaborative inquiry, analysis, solution communication	Question-answer interaction and guided practice	Engagement with the same subject content
Teacher role	Problem designer, facilitator, conceptual guide	Primary explainer and practice guide	Maintains conceptual accuracy
Assessment	Common pretest and posttest	Common pretest and posttest	Same 20-item achievement instrument

The surviving research records identify the instructional sequence and the May–June implementation window but do not document session-by-session duration, teacher fidelity ratings, observation rubrics, or contamination checks. The article therefore avoids inventing those details. Their absence is methodologically important because PBL effects can depend on problem quality, facilitation, student preparation, and implementation fidelity (Amoa-Danquah and Carbonneau 2025). Future replications should preserve lesson plans, exposure time, facilitator logs, and independent fidelity observations.

Descriptive analysis reported N, minimum, maximum, mean, standard deviation, and raw mean gain. Baseline comparability was evaluated with an independent-samples t-test on pretest scores and Cohen's d. Each research condition was then analyzed separately with a paired-samples t-test, including the mean change and its 95% confidence interval. The primary observed between-group comparison used an independent-samples

t-test on posttest scores at $\alpha = .05$. The magnitude of this difference was summarized with Cohen's d and eta squared. Calculations were independently checked from the final master dataset rather than copied from incompatible output files (Pallant 2020).

Assumptions and robustness were considered explicitly. Shapiro-Wilk tests assessed normality within each posttest group, and Levene's test assessed equality of variances. Because both posttest distributions departed from normality, a Mann-Whitney U test was used as a sensitivity analysis. The balanced sample sizes and similar variances make the t-test reasonably robust to moderate nonnormality, but convergence with a rank-based test gives a more defensible inference. A comparison of raw gain scores was reported as supplementary evidence only. In a nonequivalent design, gain differences are intertwined with starting level, possible ceiling effects, regression to the mean, and measurement error.

Pretest-adjusted analysis was considered but not promoted as a corrective device without assumption checking. In the master data, the group-by-pretest interaction was statistically significant, $F(1, 64) = 5.14$, $p = .027$, indicating that the homogeneity-of-regression-slopes assumption required by a conventional single-slope ANCOVA was not supported. Reporting a routine ANCOVA estimate as though it fully solved baseline imbalance would therefore be misleading. Moreover, pretest measurement error can bias adjusted treatment estimates in pretest-posttest designs (Miyazaki et al. 2022). The article consequently emphasizes transparent observed comparisons, robustness checks, and restrained causal language.

Result and Discussion

Results are organized around the three research questions. Statistical evidence is followed immediately by substantive interpretation, comparison with previous research, and discussion of the limits of each inference. This integrated structure prevents the results from being treated as isolated numbers and keeps the discussion aligned with what the study actually measured.

Preliminary Evidence: Distribution and Baseline
Table 3. Descriptive Statistics for Pretest and Posttest Scores

Group	N	Pretest range	Pretest (SD)	M	Posttest range	Posttest (SD)	M	Raw gain
PBL	34	72–85	79.21 (3.31)		80–100	90.15 (4.52)		10.94
Conventional	34	70–80	74.88 (2.57)		80–100	89.26 (4.46)		14.38

Table 3 shows high final achievement in both conditions, but it also reveals a consequential starting difference. The PBL group entered with a mean of 79.21, whereas the conventional group entered with 74.88. The 4.32-point baseline difference was statistically significant, $t(66) = 6.018$, $p < .001$, Cohen's $d = 1.46$. This is a large standardized difference and confirms that the intact classes were not exchangeable before instruction. Any claim that the classes were initially equivalent would contradict the data.

The baseline finding becomes the interpretive frame for both the final-score and gain-score comparisons.

Posttest scores were concentrated at a limited set of values, which is unsurprising for a 20-item test converted to a 0–100 scale. Shapiro-Wilk tests were significant in the PBL group, $W = 0.895$, $p = .003$, and the conventional group, $W = 0.885$, $p = .002$. Levene's test nevertheless supported equal variances, $F = 0.316$, $p = .576$. With equal sample sizes and almost identical posttest standard deviations, the independent t -test remains informative, while the Mann-Whitney result provides a distribution-free check. This dual reporting is preferable to declaring normality solely because a desired parametric analysis had been planned.

RQ1. Cognitive Outcomes after Problem-Based Learning

Table 4. Within-Group Pretest–Posttest Results

Group	Pretest (SD)	M Posttest (SD)	M Mean gain	95% CI for gain	t(df)	p
PBL	79.21 (3.31)	90.15 (4.52)	10.94	9.656–12.227	17.317 (33)	< .001
Conventional	74.88 (2.57)	89.26 (4.46)	14.38	12.591– 16.174	16.333 (33)	< .001

The first research question asked how far achievement improved under PBL. The PBL mean rose by 10.94 points, from 79.21 ($SD = 3.31$) to 90.15 ($SD = 4.52$). The paired comparison was statistically significant, $t(33) = 17.317$, $p < .001$, and the 95% confidence interval for the mean increase, [9.656, 12.227], excluded zero by a wide margin. The result is not a marginal threshold effect: it represents a substantial observed improvement across the instructional period. At the same time, the paired test establishes change within the PBL class, not the exclusive causal contribution of PBL. Repeated testing, ordinary classroom maturation, curricular exposure, and other concurrent influences are not removed by a within-group comparison.

The pattern is theoretically compatible with the PBL process. Presenting an Aqidah Akhlak issue as a problem can activate what students already know and make missing concepts visible. Collaborative explanation then requires students to translate terms into reasons and examine whether a proposed answer is consistent with relevant principles. Schmidt, Rotgans, and Yew (2011) identify prior-knowledge activation, elaboration, and situational interest as mechanisms through which PBL can support learning. Hmelo-Silver (2004) similarly emphasizes knowledge construction, self-directed inquiry, and flexible problem solving. The observed increase cannot verify each mechanism, because the study did not measure discourse or motivation directly, but it is consistent with a learning environment in which concepts are repeatedly retrieved and applied.

The result also corresponds with meta-analytic evidence linking PBL to cognitive and critical-thinking outcomes across educational domains (Liu and Pásztor 2022; Erdem 2024; Su et al. 2025). In science education, positive achievement effects have been

reported at secondary level, and in mathematics PBL has been associated with reasoning and problem-solving outcomes (Funa and Prudente 2021; Musna, Juandi, and Jupri 2021; Fitriyah, Putro, and Apino 2022). These cross-domain findings strengthen the plausibility of PBL as a learning design, but they do not determine the size of its effect in Aqidah Akhlak. Domain structure, the sophistication of the problem, teacher facilitation, and the assessment format remain consequential moderators.

Religious-education studies provide closer conceptual support. Problem-centered learning has been used to connect religious knowledge with comprehension, social responsibility, empathy, and critical reflection (Dana 2024; Qorib 2025; Amirudin, Muzaki, and Nurhayati 2025). Work in Fiqh emphasizes reasoning about how principles bear on cases, while integrated problem-based and flipped designs seek to prepare students for more active in-class inquiry (Abdul Khalil, Mohd Razif, and Rosele 2024; Nugraha, Abdel Rahim, and Lukman 2024). Aqidah Akhlak is not identical to those settings, but the shared pedagogical logic is that religious concepts become intellectually usable when students must relate them to a meaningful question.

One should nevertheless avoid treating PBL as a label that guarantees quality. Deep learning is more likely when students encounter appropriately challenging problems, receive guidance, participate genuinely, and are assessed in ways that value understanding (Dolmans et al. 2016). If a few students dominate group reasoning, if information search becomes superficial, or if the teacher withholds needed conceptual clarification, activity can increase without equivalent learning. The present test score increase shows that the PBL class learned; it does not provide fidelity evidence about which phase was most productive. That question requires classroom observation, student work, or process data in a future study.

RQ2. Cognitive Outcomes after Conventional Instruction

The second research question concerned change under conventional instruction. The conventional mean rose from 74.88 (SD = 2.57) to 89.26 (SD = 4.46), a raw increase of 14.38 points. The paired test was significant, $t(33) = 16.333$, $p < .001$, with a 95% confidence interval of [12.591, 16.174]. Conventional instruction was therefore associated with substantial cognitive improvement rather than serving as an inert control condition. This finding is central to a fair comparison: the relevant question is whether PBL added an observable advantage over an active, structured alternative, not whether PBL was better than absence of teaching.

Structured explanation can be effective when learners need an organized conceptual framework, precise terminology, and immediate correction of misunderstanding. Aqidah Akhlak contains propositions and classifications for which clarity matters. Teacher explanation followed by questioning and guided practice may reduce extraneous search and help students encode relationships among terms efficiently. A multiple-choice achievement test is also closely aligned with recognition and discrimination of such content. Accordingly, strong conventional-group improvement is educationally plausible and should not be dismissed as a methodological inconvenience.

The result illustrates why PBL research must specify the comparison condition. Meta-analyses often combine studies whose 'traditional' instruction ranges from passive lecture to interactive explanation with guided exercises. A well-implemented comparison condition can narrow the difference between models, particularly on immediate theoretical knowledge. Amoa-Danquah and Carbonneau (2025) report mixed evidence for theoretical knowledge even while finding broader support for engagement, critical thinking, and professional competencies. Trullàs et al. (2022) likewise show that PBL evidence varies with curriculum, implementation, and outcome. The present conventional condition may have been well aligned with the cognitive test even if PBL offered benefits on unmeasured discussion or reasoning processes.

The conventional group's larger numerical gain must be interpreted in relation to its lower starting point. A lower pretest mean creates more available distance to the upper score boundary. It can also interact with regression to the mean and with differences in how test items function across the score range. The gain is a genuine description of observed change, but it is not a randomized estimate of an instructional effect. Recognizing this distinction preserves the result without turning baseline imbalance into an unsupported claim that conventional instruction was superior.

RQ3. Between-Group Comparison and Robustness

Table 5. Primary Posttest Comparison

PBL (SD)	M	Conventional M (SD)	Mean difference	95% CI	t(df)	p	Cohen's d	Eta squared
90.15 (4.52)		89.26 (4.46)	0.882	-1.292–3.057	0.810 (66)	.421	0.20	.010

The third research question asked whether posttest achievement differed between conditions. The PBL mean was 90.15 and the conventional mean 89.26, producing an observed difference of 0.882 points. The independent-samples test was not significant, $t(66) = 0.810$, $p = .421$, and the 95% confidence interval $[-1.292, 3.057]$ included zero. Cohen's $d = 0.20$ indicated a small standardized difference, while eta squared = .010 indicated that the unadjusted group contrast accounted for about 1% of posttest variance. On the measured outcome and in this sample, the evidence did not detect a meaningful final-score advantage for PBL.

Table 6. Diagnostic and Sensitivity Evidence

Purpose	Statistic	Result	Interpretive role
Posttest normality: PBL	Shapiro-Wilk	$W = .895$, $p = .003$	Nonnormal distribution
Posttest normality: conventional	Shapiro-Wilk	$W = .885$, $p = .002$	Nonnormal distribution
Equality of posttest variances	Levene	$F = .316$, $p = .576$	Similar variances

Rank-based comparison	posttest	Mann-Whitney	U = 642, p = .409	Same inference as t-test
Raw-gain comparison		Independent test	t(66) = 3.175, p = .002, d = .77	Supplementary; baseline-sensitive
ANCOVA assumption	slope	Group × pretest	F(1,64) = 5.14, p = .027	Single-slope ANCOVA not supported

The rank-based sensitivity analysis reinforced the primary inference: Mann-Whitney U = 642.0, p = .409. Agreement between mean-based and rank-based tests is important because the posttest distributions were nonnormal and scores were discrete. The result is therefore not an artifact of using only a t-test. Nevertheless, 'not statistically significant' must not be rewritten as 'equivalent.' Formal equivalence requires a prespecified smallest effect of interest and equivalence testing. The present confidence interval is compatible with both a small disadvantage and a small advantage for PBL, so the defensible conclusion is that no posttest difference was detected.

The supplementary gain comparison produced a different surface impression. The conventional class gained 14.38 points and the PBL class 10.94, a 3.44-point difference, $t(66) = 3.175$, $p = .002$, $|d| = 0.77$. Yet the conventional group began 4.32 points lower. Because the classes were not randomized or matched, gain and baseline are structurally related. The significant group-by-pretest interaction also indicates that a conventional equal-slopes ANCOVA is not justified for these data. The gain finding is therefore informative about trajectories in these two classes, but it does not identify the superior instructional model.

Several explanations can account for substantial improvement in both groups and a negligible final difference. First, the measured outcome emphasized immediate cognitive achievement. PBL effects are often stronger for problem solving, application, critical thinking, collaboration, and motivation than for tests of theoretical knowledge (Amoa-Danquah and Carbonneau 2025; Wijnia et al. 2024; Wei et al. 2024). A selected-response test can capture important curricular mastery while remaining relatively insensitive to the quality of group explanation or moral reasoning. If PBL's distinctive contribution occurred in those processes, the present instrument would underestimate it.

Second, PBL has an adaptation cost. Students new to the model must learn how to formulate learning issues, participate equitably, evaluate information, and integrate multiple contributions. Lee (2025) found that outcomes among first-time PBL participants were associated with prior performance and attendance, illustrating that the model's benefits do not emerge independently of learner readiness. Productive interdependence also requires more than seating students in groups; task structure and facilitation must make each member's contribution meaningful (Shimizu et al. 2021). A limited intervention can therefore generate progress without yet producing a large advantage over familiar instruction.

Third, guidance may have reduced the contrast between conditions. PBL depends on scaffolding, especially when learners confront complex or unfamiliar content. Evidence from inquiry learning and computer-supported PBL shows that prompts, conceptual support, and procedural guidance improve learning (Lazonder and Harmsen 2016; Kim, Belland, and Walker 2018). Conventional instruction supplies guidance directly, whereas PBL distributes it through problems, teacher prompts, peer explanation, and resources. If both conditions provided adequate guidance and covered the same objectives, convergence in final cognitive scores is plausible even though the routes to learning differed.

Fourth, the content itself may favor a blended pedagogical logic. Some Aqidah Akhlak concepts require accurate exposition before students can analyze a case responsibly. PBL without sufficient conceptual grounding risks confident but incomplete reasoning; exposition without contextual use risks inert knowledge. The result therefore supports neither abandoning PBL nor retaining exclusively teacher-centered instruction. It suggests sequencing concise conceptual teaching with carefully designed problems, structured roles, evidence prompts, and reflective synthesis. Such a combination respects both normative precision and active reasoning.

The finding also contributes a useful counterweight to uniformly positive PBL narratives. Meta-analyses report average effects across heterogeneous studies, not guaranteed effects in every classroom. PBL outcomes vary across subject, learner, implementation, duration, comparison condition, and assessment (Gao et al. 2020; Wijnia et al. 2024; Amoa-Danquah and Carbonneau 2025). Evidence from secondary education likewise shows that effects may differ by outcome and student group (De Witte and Rogge 2016). A nonsignificant local comparison is therefore not a failure to reproduce a universal law; it is evidence about the conditions under which a detectable advantage did not emerge.

Educational Implications

For Aqidah Akhlak teachers, the practical implication is to treat PBL as a design framework rather than a replacement slogan. Problems should be authentic enough to require judgment, bounded enough to be addressed with available knowledge, and explicitly connected to targeted concepts. Group roles can distribute responsibility for defining the issue, locating relevant principles, evaluating alternatives, and presenting the final reasoning. Teacher prompts should ask students to justify connections rather than merely state preferences. The closing synthesis is essential: it consolidates correct concepts, addresses misconceptions, and relates the case back to the curriculum.

Assessment should also match the intended contribution. If PBL is expected to develop contextual reasoning, critical thinking, collaboration, or ethical deliberation, a multiple-choice test alone is insufficient. Future evaluations can combine a reliable cognitive test with scenario-based written responses, analytic rubrics, observation of group discourse, reflective journals, and delayed measures of transfer. Religious-education studies have already explored outcomes such as responsibility, empathy, social concern, and religious comprehension (Nawi and Zakaria 2019; Rohman 2022; Dana

2024; Qorib 2025; Amirudin, Muzaki, and Nurhayati 2025). Measuring several domains would show whether similar cognitive scores conceal different qualities of learning.

For school leaders and researchers, implementation documentation should become part of the intervention rather than an afterthought. A publishable PBL evaluation should record the number and length of meetings, problem materials, facilitator preparation, student attendance, exposure equivalence, fidelity indicators, deviations from plan, and examples of student products. Such information enables replication and explains variation across studies. It also prevents the instructional label from substituting for evidence that the defining PBL processes actually occurred.

Limitations and Directions for Further Research

Five limitations delimit the conclusions. First, the sample came from one school and contained only 68 students, restricting statistical precision and transfer to other institutional contexts. Second, one intact class represented each condition. The instructional effect is therefore confounded with any class-specific composition, peer climate, timetable, or teacher-related influence. Third, the large baseline difference demonstrates selection imbalance. Neither the final-score comparison nor raw gain eliminates that threat, and the observed group-by-pretest interaction prevented a routine equal-slopes ANCOVA from serving as a simple remedy.

Fourth, the outcome was a 20-item immediate cognitive test with discrete score values. The instrument had acceptable reported item correlations and internal consistency, but it did not measure extended reasoning, affective internalization, behavior, delayed retention, or transfer to unfamiliar moral situations. Fifth, the archive did not contain detailed exposure time, fidelity observations, or process measures. Consequently, the study cannot determine whether the lack of a between-group advantage reflects the model, the implementation dose, adaptation to PBL, assessment alignment, or a combination of these factors.

Future research should use several classes per condition, random assignment where feasible, or strong matching and longitudinal designs where randomization is not possible. Baseline adjustment should be planned before data collection, accompanied by checks of measurement quality and model assumptions rather than applied mechanically (Miyazaki et al. 2022). Researchers should prespecify a primary outcome and the smallest effect that would be educationally meaningful. If the goal is to establish that two approaches are practically similar, an equivalence or noninferiority design is more appropriate than interpreting $p > .05$ as equivalence. Multilevel models will also be needed when students are nested within multiple classes and teachers.

A stronger domain-specific program of research would compare alternative PBL designs in Aqidah Akhlak: unguided versus scaffolded inquiry, individual preparation versus flipped preparation, homogeneous versus structured heterogeneous groups, and cognitive-only versus multidomain assessment. Qualitative analysis of student explanations could identify how learners use religious concepts when resolving a case. Delayed follow-up could determine whether problem-centered learning produces more durable retrieval or transfer even when immediate posttest means are similar. These

extensions would move the evidence from a question of whether PBL 'works' toward an explanation of how, for whom, and for which outcomes it works.

Conclusion

This quasi-experimental study examined three distinct questions about cognitive Aqidah Akhlak learning. The PBL class improved significantly from $M = 79.21$ to $M = 90.15$, with a mean gain of 10.94 points. The conventional class also improved significantly, from $M = 74.88$ to $M = 89.26$, with a mean gain of 14.38 points. The observed posttest difference was small and not statistically significant, $t(66) = 0.810$, $p = .421$, $d = 0.20$; the Mann-Whitney sensitivity analysis produced the same inference. Thus, both instructional conditions were associated with substantial cognitive progress, while no posttest advantage for PBL was detected in this sample.

The result should not be described as proof of equivalent effectiveness, nor should the larger conventional-group gain be described as causal superiority. The intact classes differed markedly before instruction, the posttest was bounded and discrete, and the assumption required for a simple single-slope ANCOVA was not supported. The study's strongest conclusion is therefore descriptive and comparative: students in both classes learned, their final cognitive means were close, and the design cannot isolate a superior model.

Pedagogically, the evidence supports a complementary approach. PBL offers a principled way to connect Aqidah Akhlak concepts with contextual inquiry, collaborative explanation, and reflection; structured instruction remains useful for building precise conceptual foundations and correcting misconceptions. High-quality practice can combine these strengths through concise explanation, carefully bounded moral or religious problems, guided investigation, explicit reasoning prompts, and teacher-led synthesis. Future studies should test that alignment with multiple classes, better baseline control, documented fidelity, and measures of cognitive achievement, reasoning, affect, behavior, and delayed transfer.

Reference

- Abdul Khalil, S., Mohd Razif, N. F., & Rosele, M. I. (2024). Developing ijtihad skills for undergraduate students through problem-based learning in fiqh subjects: Present practices and the way forward. *Asia Pacific Journal of Educators and Education*, 39(2), 197–217. doi:10.21315/apjee2024.39.2.11
- Al-Hasyim, R., Kasran, & Rambe, M. (2024). Pengaruh model pembelajaran problem based learning terhadap hasil belajar siswa pada mata pelajaran Akidah Akhlak di Madrasah Aliyah Ikhwanul Muslimin Kabupaten Serdang Bedagai. *At-Tarbiyah: Jurnal Penelitian dan Pendidikan Agama Islam*, 2(1), 208–214.
- Amanda, D. A. S., & Ndziejia, A. N. (2026). Implementation of problem-based learning in enhancing student engagement and learning achievement in Islamic religious education for children. *Research Journal on Teacher Professional Development*, 4(1), 19–38. doi:10.21580/rjtpd.v4i01.24904

- Amirudin, A., Muzaki, I. A., & Nurhayati, S. (2025). Problem-based learning as a pedagogical innovation for transforming higher education students' Islamic religious comprehension. *Educational Process: International Journal*, 18, e2025412. doi:10.22521/edupij.2025.18.412
- Amoa-Danquah, P., & Carbonneau, K. J. (2025). A systematic review of reviews on problem-based learning and its effectiveness. *Current Issues in Education*, 26(2), 1–20. doi:10.14507/cie.vol26iss2.2293
- Anazifa, R. D., & Djukri. (2017). Project-based learning and problem-based learning: Are they effective to improve students' thinking skills? *Jurnal Pendidikan IPA Indonesia*, 6(2), 346–355. doi:10.15294/jpii.v6i2.11100
- Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education* (10th ed.). Boston, MA: Cengage Learning.
- Dana, C. (2024). Model pembelajaran PAI berbasis problem based learning dalam pembentukan karakter tanggung jawab peserta didik. *Oasis: Jurnal Ilmiah Kajian Islam*, 8(2). doi:10.24235/oasis.v8i2.13363
- De Witte, K., & Rogge, N. (2016). Problem-based learning in secondary education: Evaluation by an experiment. *Education Economics*, 24(1), 58–82. doi:10.1080/09645292.2014.966061
- Dolmans, D. H. J. M., Loyens, S. M. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: A review of the literature. *Advances in Health Sciences Education*, 21(5), 1087–1112. doi:10.1007/s10459-015-9645-6
- Erdem, C. (2024). A comparative meta-analysis of the effects of problem-based learning model on K–12 students' cognitive outputs. *Educational Studies*, 50(6), 1498–1519. doi:10.1080/03055698.2022.2103650
- Fitriyah, I. M., Putro, N. H. P. S., & Apino, E. (2022). Meta analysis study: Effectiveness of problem based learning on Indonesian students' mathematical reasoning ability. *Jurnal Riset Pendidikan Matematika*, 9(1), 36–45. doi:10.21831/jrpm.v9i1.46447
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). New York, NY: McGraw-Hill Education.
- Funa, A. A., & Prudente, M. S. (2021). Effectiveness of problem-based learning on secondary students' achievement in science: A meta-analysis. *International Journal of Instruction*, 14(4), 69–84. doi:10.29333/iji.2021.1445a
- Gao, J., Yang, L., Zhao, J., Wang, L., Zou, J., Wang, C., & Fan, X. (2020). Comparison of problem-based learning and traditional teaching methods in medical psychology education in China: A systematic review and meta-analysis. *PLOS ONE*, 15(12), e0243897. doi:10.1371/journal.pone.0243897
- Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2022). *Multivariate data analysis* (9th ed.). Boston, MA: Cengage Learning.
- Hamid, F., & Rahman, M. H. (2026). Implementation of problem-based learning methods in Islamic education to strengthen students' critical thinking skills at the Simalungun State Islamic School. *Multidisciplinary Indonesian Center Journal*, 3(1), 2199–2206. doi:10.62567/micjo.v3i1.2305

- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235–266. doi:10.1023/B:EDPR.0000034022.16470.f3
- Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. *Educational Psychology Review*, 30(2), 397–429. doi:10.1007/s10648-017-9419-1
- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86(3), 681–718. doi:10.3102/0034654315627366
- Lee, Y.-C. (2025). Changes in learning outcomes of students participating in problem-based learning for the first time: A case study of a financial management course. *The Asia-Pacific Education Researcher*, 34(1), 511–530. doi:10.1007/s40299-024-00873-y
- Liu, Y., & Pásztor, A. (2022). Effects of problem-based learning instructional intervention on critical thinking in higher education: A meta-analysis. *Thinking Skills and Creativity*, 45, 101069. doi:10.1016/j.tsc.2022.101069
- Magaji, A., Adjani, M., & Coombes, S. (2024). A systematic review of preservice science teachers' experience of problem-based learning and implementing it in the classroom. *Education Sciences*, 14(3), 301. doi:10.3390/educsci14030301
- Miyazaki, Y., Kamata, A., Uekawa, K., & Sun, Y. (2022). Bias for treatment effect by measurement error in pretest in ANCOVA analysis. *Educational and Psychological Measurement*, 82(6), 1130–1152. doi:10.1177/00131644211068801
- Musna, R. R., Juandi, D., & Jupri, A. (2021). A meta-analysis study of the effect of problem-based learning model on students' mathematical problem solving skills. *Journal of Physics: Conference Series*, 1882, 012090. doi:10.1088/1742-6596/1882/1/012090
- Nawi, A., & Zakaria, G. A. N. (2019). The impact of applying religious values through online problem-based learning among undergraduate students: A quasi-experimental study. *International Journal of Learning, Teaching and Educational Research*, 18(8), 192–213. doi:10.26803/ijlter.18.8.12
- Nugraha, T. C., Abdel Rahim, E.-S. E.-S. A., & Lukman, F. (2024). Integrating problem-based and flipped learning in Islamic religious education: A pathway to achieving sustainable development goals. *Jurnal Pendidikan Islam*, 10(1), 125–136. doi:10.15575/jpi.v10i1.35204
- Orhan, A. (2025). Investigating the effectiveness of problem based learning on academic achievement in EFL classroom: A meta-analysis. *The Asia-Pacific Education Researcher*, 34, 699–709. doi:10.1007/s40299-024-00889-4
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). London, England: Routledge.

- Qorib, F. (2025). Analisis model problem based learning pada fiqih puasa untuk menumbuhkan empati dan kepedulian sosial siswa madrasah ibtidaiyah. *Oasis: Jurnal Ilmiah Kajian Islam*, 10(1), 28–40. doi:10.24235/oasis.v10i1.21074
- Rohman, F. (2022). Problem based learning in Islamic religious education: The case of the Indonesian pesantren. *Global Journal Al-Thaqafah*, 12(1), 82–97. doi:10.7187/GJAT072022-5
- Schmidt, H. G., Rotgans, J. I., & Yew, E. H. J. (2011). The process of problem-based learning: What works and why. *Medical Education*, 45(8), 792–806. doi:10.1111/j.1365-2923.2011.04035.x
- Sharma, S., Saragih, I. D. E. T., Tarihoran, D. E. T. A. U., & Chou, F.-H. (2023). Outcomes of problem-based learning in nurse education: A systematic review and meta-analysis. *Nurse Education Today*, 120, 105631. doi:10.1016/j.nedt.2022.105631
- Shimizu, I., Matsuyama, Y., Duvivier, R., & van der Vleuten, C. (2021). Contextual attributes to promote positive social interdependence in problem-based learning: A focus group study. *BMC Medical Education*, 21, 222. doi:10.1186/s12909-021-02667-y
- Su, T., Liu, J., Meng, L., Luo, Y., Ke, Q., & Xie, L. (2025). The effectiveness of problem-based learning in enhancing critical thinking skills in medical education: A systematic review and meta-analysis. *Frontiers in Education*, 10, 1565556. doi:10.3389/feduc.2025.1565556
- Trullàs, J. C., Blay, C., Sarri, E., & Pujol, R. (2022). Effectiveness of problem-based learning methodology in undergraduate medical education: A scoping review. *BMC Medical Education*, 22, 104. doi:10.1186/s12909-022-03154-8
- Wei, B., Wang, H., Li, F., Long, Y., Zhang, Q., Liu, H., Rao, M. (2024). Effectiveness of problem-based learning on development of nursing students' critical thinking skills: A systematic review and meta-analysis. *Nurse Educator*, 49(3), E115–E119. doi:10.1097/NNE.0000000000001548
- Wijnia, L., Noordzij, G., Arends, L. R., Rikers, R. M. J. P., & Loyens, S. M. M. (2024). The effects of problem-based, project-based, and case-based learning on students' motivation: A meta-analysis. *Educational Psychology Review*, 36, 29. doi:10.1007/s10648-024-09864-3
- Wiratama, P. P., Dwandaru, W. S. B., Kuswanto, H., Laeli, S., & Kurniawan, A. (2025). Effectiveness of problem-based learning in enhancing critical thinking skills in science education: Meta-analysis. *Jurnal Eduscience*, 12(4), 1015–1028.
- Yew, E. H. J., & Goh, K. (2016). Problem-based learning: An overview of its process and impact on learning. *Health Professions Education*, 2(2), 75–79. doi:10.1016/j.hpe.2016.01.004