WOMEN, EDUCATION, AND SOCIAL WELFARE
VOL. 3 NO. 3 (2026)

ISSN: 3064-2469

Integrated Tafsir–Qur'anic Sciences Module for Qur'anic Literacy in a Female-Majority Pesantren

Yusri Yaldi*, Wedra Aprison, Yelfi Dewi S, Untung Famuji, and Muhammad Safdar Bhatti

To cite this article: Y. Yaldi, W. Aprison, Y. D. S, U. Famuji, and M. S. Bhatti, "Integrated Tafsir–Qur'anic Sciences Module for Qur'anic Literacy in a Female-Majority Pesantren," *Women, Educ. Soc. Welf.*, vol. 3, no. 3, pp. 1120–1135, 2026. https://doi.org/10.70211/wesw.v3i3.520

Published online: September 16, 2026

OPEN ACCESS

https://doi.org/10.70211/wesw.v3i3.520

Women, Education, and Social Welfare Vol 3 No 3 (2026)

Check for updates

Research Article

Integrated Tafsir–Qur'anic Sciences Module for Qur'anic Literacy in a Female-Majority Pesantren

Yusri Yaldi*, Wedra Aprison, Yelfi Dewi S, Untung Famuji, and Muhammad Safdar Bhatti

Received: June 24, 2025	Revised: July 23, 2026	Accepted: September 11, 2026	Online: September 16, 2026

Abstract

Teaching Tafsir separately from Qur'anic Sciences ('Ulumul Qur'an) can leave students able to reproduce meanings without routinely applying the interpretive principles that justify them. This study developed and evaluated an integrated Tafsir–'Ulumul Qur'an module for grade XII students in a female-majority Indonesian pesantren. A one-cycle Four-D Research and Development design was combined with a nonrandomized classroom outcome comparison. Evidence included curriculum documents, observation and interview sources, expert appraisal by six validators, practicality responses from six educators and 55 students, and learning outcomes from two classroom cohorts (18 existing-instruction and 20 module students). The 55-student practicality set and 38-student outcome set were distinct analytic samples rather than an attrition sequence. The 32-indicator expert appraisal produced an overall Aiken's V of 0.857. Educator and student practicality indices were 0.905 and 0.877, and the reported student N-gain was 0.739. Final scores were higher in the module cohort (M = 73.95, SD = 15.81) than in the comparison cohort (M = 56.17, SD = 13.07), with a mean difference of 17.78 points (95% CI [8.27, 27.29]), Welch's t(35.76) = 3.79, p = 0.0006, and d = 1.22. Because random allocation and group-specific baseline equivalence were not documented, the between-group result is interpreted as preliminary evidence rather than a causal effect. The module's main contribution is an infusive-complementary architecture that embeds contextual, structural, and methodological Qur'anic Sciences into aligned Tafsir tasks and assessment.

Keywords: Qur'anic Interpretive Literacy; Qur'anic Sciences; Pesantren Curriculum; Female Education.

INTRODUCTION

Tafsir learning becomes pedagogically fragile when students encounter Qur'anic meanings without routinely applying the interpretive instruments that justify those meanings. Studies of Qur'an teaching have repeatedly noted the persistence of recitation, memorization, and teacher-led explanation, while recent work on Qur'anic literacy calls for movement from textual decoding toward contextual understanding, application, and critical reflection [1], [2], [3]. Within Islamic education more broadly, critical inquiry is not antithetical to religious formation; it can be cultivated through dialogic, ethically bounded, and tradition-sensitive pedagogies [4], [5].

This problem is particularly consequential in Indonesian pesantren, where classical textual traditions coexist with pressures for curriculum modernization and stronger learning-outcome alignment. Pesantren remain distinctive because they combine scholarly transmission, communal discipline, local authority, and institutional continuity [6], [7]. For female learners, access to intellectually demanding religious study also has an educational-equity dimension, although equity should not be inferred from enrollment alone [8], [9]. Recent pesantren curriculum research therefore emphasizes modernization that preserves institutional identity while strengthening curricular coherence and quality assurance [10], [11], [12].

Qur'anic Sciences ('Ulumul Qur'an) provide the epistemic tools required to move Tafsir from descriptive explanation to methodologically responsible interpretation. In this study, the module did not attempt to cover every branch of 'Ulumul Qur'an. Components were selected when they had direct utility for interpreting the Tafsir units and could be converted into observable student tasks: contextual tools such as asbab al-nuzul and Makkan–Madinan classification; structural tools such as munasabah and verse sequence; and methodological tools including mantuq–mafhum, nasikh–mansukh, i'jaz, tafsir–ta'wil distinctions, and scholarly ethics of interpretation. These domains are consistent with scholarship that treats Qur'anic sciences as instruments for contextual, disciplined, and intellectually accountable reading [13], [14], [15], [16], [17]. Deliberative approaches to Muslim education likewise show how tadabbur and structured reasoning can be translated into secondary-level learning activities [18], [19].

The institutional needs analysis at Diniyah Limo Jurai Islamic Boarding School identified a specific curricular disconnection. Tafsir documents and classroom practice emphasized Arabic reading, language structure, translation, and explanation of thematic meaning, whereas 'Ulumul Qur'an was organized as a separate theoretical subject. Syllabus review, learning-material review, classroom observation, and interviews with institutional leaders and teachers converged on the same design problem: students encountered the two knowledge domains in parallel, but the curriculum did not systematically require them to use Qur'anic-sciences principles to justify a Tafsir conclusion. Because the archived materials do not include source-by-source coding counts, this convergence is treated as a design diagnosis rather than a frequency-based qualitative result.

The module was therefore grounded in constructive alignment: objectives, learning activity, and assessment were organized around the same interpretive performances [20]. The intended competence was not merely to remember a concept such as asbab al-nuzul or munasabah, but to identify when it is relevant, apply it to a verse or interpretive passage,

compare plausible readings, and justify a conclusion. Meta-analytic and framework studies show that critical thinking improves when learners explicitly analyze evidence, justify claims, and receive guided rather than minimally structured support [21], [22], [23], [24]. Feedback and self-regulated assessment further support the transition from first attempts to more disciplined reasoning [25], [26], while the active–constructive–interactive framework clarifies why interpretive tasks should move beyond passive reception [27].

Recent empirical literature provides important but incomplete precedents. Living-Qur'an studies link contextual engagement with critical reflection [28]; Qur'an-based pedagogy has been used to connect scriptural study with tolerance and contemporary social reasoning [29]; and problem-based or dialogic Islamic education has been associated with stronger analytical participation [30], [31]. Other curriculum integrations connect Qur'anic values to environmental learning [32]. These studies establish the pedagogical value of contextualization, inquiry, and integration, but they do not test the specific curricular move examined here: infusing 'Ulumul Qur'an as a recurring interpretive scaffold inside a Tafsir core subject and evaluating the resulting product through expert appraisal, practicality evidence, and classroom outcomes.

The study's novelty is architectural rather than merely topical. The Integrasi Multidisipliner Infusif-Komplementatif model retains Tafsir as the core subject while 'Ulumul Qur'an functions as a complementary analytic layer across contextual, structural, and methodological tasks. Active-learning evidence indicates that carefully designed participation can deepen understanding [33], [34], while motivation theory suggests that structured competence-building and meaningful participation are most productive when challenge is accompanied by support and a sense of agency [35], [36]. In a female-majority pesantren, this architecture is also relevant to equitable access to higher-order religious learning within the regular curriculum. The study did not, however, measure gender differences, agency, or confidence; gender is therefore treated as an important educational context rather than a tested causal mechanism.

Accordingly, this study asked four questions: (RQ1) What curriculum gap emerges when Tafsir and 'Ulumul Qur'an are taught as separate subjects in the target pesantren? (RQ2) How can 'Ulumul Qur'an be operationalized as an infusive-complementary scaffold within Tafsir learning? (RQ3) Does the developed module meet the study's predefined expert-validity and user-practicality criteria? and (RQ4) What learning-outcome pattern is observed after implementation, including the reported student N-gain and the final-score difference between the two nonrandomized classroom cohorts?

METHODS

Research Design

This study used a Research and Development (R&D) design organized as one Four-D development cycle: Define, Design, Develop, and Disseminate. The Four-D sequence structured product development; the classroom outcome component was a separate, nonrandomized evaluation embedded in the dissemination stage. This distinction is important because the study was not a randomized trial and was not a full design-based-research program with multiple theory-building iterations. Its classroom grounding shares features with design

research—authentic setting, iterative refinement of an educational product, and attention to use conditions [37], [38], [39], [40]—but the inferential claims in this article are limited to the single development cycle and the available outcome evidence.

Research Setting and Time

The study was conducted in the grade XII Takhasus Tafsir program at Diniyah Limo Jurai Islamic Boarding School, Sungai Pua, Agam Regency, West Sumatra, Indonesia. The module was implemented over one semester during the development cycle. The setting is described as female-majority because female students constituted the dominant learning context of the pesantren program; however, exact gender subgroup counts were not retained in the article-level dataset and no gender-comparative analysis was performed.

Participants and Sampling

Purposive and classroom-based sampling were used because each participant group had a distinct role in product development or evaluation. Six experts appraised the module, six educators assessed practicality and implementation feasibility, and 55 grade XII students completed the student-practicality questionnaire. The learning-outcome comparison used 38 students reported in two classroom cohorts: 18 in the existing-instruction comparison group and 20 in the module group. The 55-student practicality dataset and the 38-student outcome dataset were separate analytic sets; the numerical difference should therefore not be read as 17 cases of attrition. The retained report does not document individual random assignment or group-specific demographic and prior-achievement distributions, so the two outcome groups are treated as nonequivalent classroom cohorts and causal language is avoided.

Table 1. Participants, Data Sources, and Analytical Roles

Participant / source	N	Sampling / inclusion logic	Analytical role
Expert validators	6	Purposive; disciplinary/product expertise	Content, construct, language, curriculum, method, assessment, completeness, and evaluation appraisal
Educators / practitioners	6	Purposive; direct implementation role	Practicality and implementation feasibility
Student practicality respondents	55	Grade XII participation in product-use evaluation	Perceived clarity, usability, and feasibility
Learning-outcome comparison	38	Two classroom cohorts: 18 existing instruction; 20 module	Final-score comparison; separate from the practicality sample
Needs-analysis sources	Not reliably retained	Tafsir and ‘Ulumul Qur’an documents, classroom observation, institutional/teacher interviews	Triangulated curriculum-gap diagnosis; no frequency claims

Data Collection Procedure

During Define, the researchers reviewed the Tafsir and ‘Ulumul Qur’an syllabi and learning materials, observed classroom practice, and interviewed institutional leaders and teachers to identify the curriculum gap. The available project archive identifies these source categories but does not preserve reliable counts for every document, observation episode, or interview; consequently, the qualitative evidence is reported as source triangulation rather than frequency coding. During Design, learning outcomes and higher-order indicators were mapped to three integration layers—contextual, structural, and methodological—and converted into concept maps, guided analysis prompts, collaborative activities, reflective tasks, and assessment rubrics. During Develop, six experts assessed 32 indicators across content, language, presentation, learning method, assessment, completeness, and evaluation; the prototype was revised before practicality testing. During Disseminate, teachers received socialization/training and the module was implemented for one semester. Student learning was assessed with higher-order multiple-choice and essay tasks addressing contextual, structural, and methodological interpretation. The available archive does not include the item-level response matrix, the exact number of outcome-test items, or the original interview protocol, so these details are not reconstructed in this article.

Data Analysis

Expert appraisal was summarized with Aiken’s V [41]. The original study applied an institutional decision rule in which values above 0.81 were classified as very valid/practical and 0.61–0.80 as valid/practical. Because the available records do not preserve the number of response categories or item-level expert ratings, these thresholds are reported as the study’s predefined descriptive criteria rather than universal critical values; item-level confidence bounds cannot be reconstructed. Practicality was reported in the original dataset using a statistic labeled ‘Moment Kappa’. Because item-level response matrices and scale metadata are unavailable, the statistic is retained only as the original descriptive practicality index and is not reinterpreted as inter-rater reliability. Student improvement was summarized using normalized gain (N-gain) [42]. For the two classroom cohorts, the reported means, standard deviations, and sample sizes were used to verify a Welch independent-samples t-test, the mean difference with a 95% confidence interval, and Cohen’s d with an approximate 95% confidence interval [43]. The observed comparison was $t(35.76) = 3.79$, $p = 0.0006$; retrospective power for the observed $d = 1.22$ and $n = 38$ was approximately 0.95, but this post hoc value is not treated as evidence of design adequacy because baseline equivalence and random allocation were not established.

Ethical Considerations

The study was implemented as school-based curriculum development and reports only aggregated educational data; individual student identities are not included in the manuscript. The project documentation available for this article does not provide a separate research ethics committee approval number, and no approval number is asserted here. Because the participants were school students, future replications should explicitly archive the applicable institutional authorization, consent/assent procedures, data-protection protocol, and safeguards for minors so that ethical compliance can be independently audited.

RESULTS AND DISCUSSION

Results

Evidence Profile and Analytic Boundaries

The development evidence comprised four participant-based datasets plus qualitative source categories used for needs analysis. These datasets served different purposes and were not collapsed into a single sample. Table 1 makes the distinction explicit, especially between the 55 students who provided practicality responses and the 38 students included in the classroom outcome comparison.

Baseline Separation of Tafsir and Qur'anic Sciences

The needs analysis showed a consistent curricular separation between the two subjects. Tafsir learning materials organized study around selected verses and recognized Tafsir sources, with classroom work emphasizing reading, language structure, translation, and explanation. 'Ulumul Qur'an materials addressed concepts such as asbab al-nuzul, Makkan–Madinan classification, munasabah, nasikh–mansukh, tafsir–ta'wil, and the qualifications of a mufassir, but these concepts were not systematically embedded as required steps in Tafsir analysis. Table 2 reports only patterns supported by the retained source descriptions and avoids speculative claims about what students 'may' have learned.

Table 2. Triangulated Evidence of the Baseline Curriculum Gap

Evidence source	Observed / documented pattern	Curricular implication	Boundary of evidence
Tafsir syllabus and learning materials	Selected verses and recognized Tafsir sources; emphasis on reading, language structure, translation, and thematic meaning	Interpretive outcomes were not explicitly linked to 'Ulumul Qur'an application	Document pattern; no student-level causal inference
'Ulumul Qur'an syllabus and materials	Theoretical treatment of contextual, structural, and methodological concepts	Concepts were positioned beside, rather than inside, Tafsir analytical tasks	Separate-subject organization documented
Classroom observation	Reading, translation, and explanation dominated the described lesson sequence	Few explicit prompts required method-based interpretive justification	Observation-frequency counts were not retained
Institutional / teacher interviews	Need for closer linkage between Tafsir and 'Ulumul Qur'an was identified during needs analysis	Supported an integrated module as the design response	Interview count and verbatim quotation archive not retained in article dataset

Infusive-Complementary Module Architecture

The developed product operationalized Integrasi Multidisipliner Infusif-Komplementatif: Tafsir remained the core subject, while 'Ulumul Qur'an was infused as a complementary

analytic system rather than taught as an adjacent block of theory. The architecture had three recurrent layers. The contextual layer prompted learners to establish revelatory and historical context; the structural layer required attention to verse relationships and sequence; and the methodological layer required students to select and justify relevant interpretive principles. These layers were tied to learning outcomes, guided activities, formative feedback, and an assessment rubric so that the same competence was taught and measured. Table 3 details the architecture, and Figure 1 shows the development and revision flow.

Table 3. Architecture of the Integrated Tafsir–Qur'anic Sciences Module

Module element	Pedagogical function	Tafsir core	Infused Qur'anic-sciences layer
Learning outcomes	Defines observable interpretive competence	Selected thematic verse / Tafsir passage	Use relevant context, structure, and method to justify interpretation
Concept map	Makes relationships among ideas explicit	Theme, verse, meaning, ethical implication	Asbab al-nuzul, Makkan–Madinan, munasabah, verse sequence, methodological principles
Guided explanation	Provides disciplinary scaffolding	Meaning from recognized Tafsir sources	Why and when a Qur'anic-sciences principle is analytically relevant
Student activity	Makes reasoning visible	Read, translate, compare, explain	Select a principle, apply it, defend relevance, compare interpretations
Assessment rubric	Aligns scoring with intended competence	Accuracy and clarity of interpretation	Contextual adequacy, structural coherence, methodological justification
Reflection / feedback	Supports revision and self-monitoring	Refine interpretive conclusion	Identify assumptions, method limits, and ethical boundaries

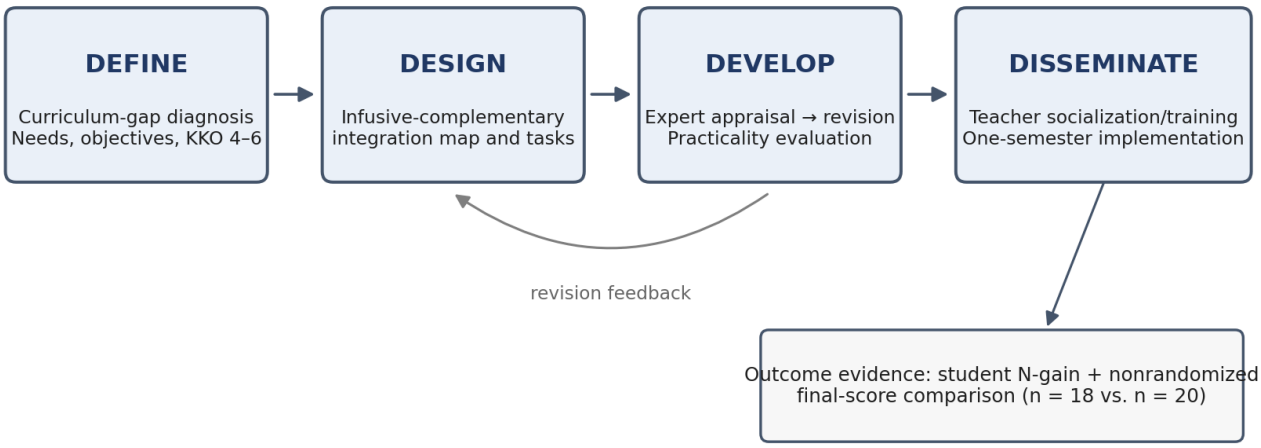

Figure 1. Four-D development flow, revision loop, and classroom outcome-evidence stage

Worked Replication Example

To make the architecture reproducible without presenting new empirical data, Table 4 gives an illustrative transposition of the three-layer procedure using Q. al-Baqarah 2:145. The example is pedagogical, not an additional observation from the outcome dataset: it demonstrates how a teacher can move from contextual and structural prompts to a method-based interpretive justification and a rubric-aligned response. The study's empirical claim concerns the module architecture and evaluation results, not the specific performance shown in this illustration.

Table 4. Illustrative Application of the Three-Layer Architecture to Q. al-Baqarah 2:145

Stage	Teacher / module prompt	Student learning opportunity	Aligned evidence
Tafsir anchor	Read the verse and selected recognized Tafsir explanation; state the interpretive question	Identify the central meaning before adding methodological claims	Accurate representation of source meaning
Contextual layer	Which revelatory/historical context is relevant, and what interpretive problem does it clarify?	Use contextual information only when it changes or constrains interpretation	Context is relevant and explicitly linked to the conclusion
Structural layer	How does the verse relate to surrounding verses and the sequence of the passage?	Explain textual coherence rather than treating the verse in isolation	Munasabah / sequence claim is textually defensible
Methodological layer	Which principle—e.g., mantuq–mafhum or another relevant rule—governs the inference?	Select and justify a method rather than merely naming a concept	Method is correctly applied and its limits are acknowledged
Synthesis	Write a concise interpretive conclusion and compare it with the initial reading	Integrate text, context, structure, and method into one justified claim	Rubric scores accuracy, coherence, evidence, and methodological responsibility

Expert Validity and Practicality

All seven expert-appraisal domains met the study's 'very valid' criterion. The overall Aiken's V was 0.857; aspect-level values ranged from 0.842 for language to 1.000 for the single evaluation indicator. Reporting the aspects separately is important because it shows the distribution across design domains rather than relying only on a global average. Educator and student practicality indices were 0.905 and 0.877, respectively, both above the predefined 0.81 criterion. The archived development record also notes that teachers found the guidance and activities workable while still requiring adjustment to the rhythm of pesantren instruction. The original report calculated an educator N-gain of 0.861; this revision does not use that value as learning-effectiveness evidence because educators were implementers and raters, not student outcome participants.

Table 5. Expert-Appraisal Results by Validation Aspect

Validation aspect	Indicators	Aiken's V	Study criterion
Content feasibility	6	0.875	Very valid
Language feasibility	5	0.842	Very valid
Presentation	9	0.861	Very valid
Learning method	5	0.850	Very valid
Assessment	6	0.847	Very valid
Completeness	1	0.875	Very valid
Evaluation	1	1.000	Very valid
Overall	32	0.857	Very valid

Learning-Outcome Evidence

The student N-gain reported in the original development dataset was 0.739. In the separate two-cohort final-score comparison, the module group (n = 20, M = 73.95, SD = 15.81) scored higher than the existing-instruction group (n = 18, M = 56.17, SD = 13.07). The mean difference was 17.78 points (95% CI [8.27, 27.29]), Welch's t(35.76) = 3.79, p = 0.0006, and Cohen's d = 1.22 (approximate 95% CI [0.52, 1.92]). The magnitude is educationally substantial, but the group-specific pretest means and allocation mechanism were not preserved in the article dataset. The result therefore supports a strong between-cohort association with module exposure, not a definitive causal treatment effect.

Table 6. Practicality and Student Learning-Outcome Evidence

Evidence domain	Source	Statistic / estimate	Interpretation boundary
Educator practicality	6 educators	Moment Kappa = 0.905	Original descriptive index; very practical by study criterion
Student practicality	55 students	Moment Kappa = 0.877	Original descriptive index; distinct from outcome sample
Student improvement	Development dataset	N-gain = 0.739	Descriptive pre/post improvement evidence
Existing-instruction final score	n = 18	M = 56.17; SD = 13.07	Nonequivalent classroom cohort
Module final score	n = 20	M = 73.95; SD = 15.81	Nonequivalent classroom cohort
Between-cohort contrast	n = 38	ΔM = 17.78; 95% CI [8.27, 27.29]; t(35.76) = 3.79; p = 0.0006; d = 1.22; approx. 95% CI [0.52, 1.92]	Large association; not a causal estimate because baseline equivalence/random allocation were not established

Discussion

The central contribution of this study is the conversion of 'Ulumul Qur'an from a parallel theoretical subject into an epistemic scaffold embedded within Tafsir performance. This differs

from simply adding more religious content. The module changes the sequence of learning: students encounter a verse and Tafsir account, identify which Qur'anic-sciences tools are relevant, apply those tools, justify the interpretive relationship, and receive feedback on the quality of that justification. This responds directly to concerns that Qur'an teaching can remain dominated by recitation, memorization, or explanation without sufficient interpretive participation [1], [2].

The three-layer architecture also clarifies how contextualization can be academically bounded. Recent Qur'an-Hadith and Living-Qur'an studies show the value of linking scriptural learning to reflection and lived contexts [3], [28], while deliberative tadabbur offers a compatible model of structured reasoning in Muslim education [18]. The present module extends this literature by specifying the disciplinary tools that mediate interpretation: context is not treated as free-form relevance, but is disciplined through asbab al-nuzul and Makkan–Madinan knowledge; structural interpretation is disciplined through munasabah and verse sequence; and methodological judgment is disciplined through established principles and interpreter ethics.

This design is consistent with constructive alignment and guided instruction. When learning objectives, classroom tasks, and assessment all require interpretive justification, students repeatedly practice the competence the curriculum intends to develop [20]. The critical-thinking literature further suggests that explicit argumentation, evidence evaluation, and scaffolding are more dependable than minimally guided discovery for complex novice learning [21], [22], [23], [24]. Feedback and reflection then help students compare initial reasoning with disciplinary criteria [25], [26]. The architecture is therefore theoretically coherent with the observed direction of the outcome evidence, even though the present design cannot isolate which component produced the score difference.

The module also differs from adjacent integration models. Pesantren curriculum modernization commonly focuses on balancing traditional and formal curricular structures [10], [11], while other contemporary Islamic-education integrations connect Qur'anic content with environmental or social themes [29], [32]. Those models demonstrate the feasibility of curriculum integration, but the present study locates integration inside the act of Tafsir itself: 'Ulumul Qur'an functions as a complementary method layer under Tafsir as the core disciplinary object. That distinction is substantive because it targets the validity of interpretive reasoning rather than only thematic enrichment.

The large final-score difference should nevertheless be interpreted conservatively. The observed d of 1.22 is larger than effects often seen in ordinary classroom interventions, and the confidence interval remains wide. Comparable inquiry-oriented Islamic education studies report gains in critical participation or reasoning [30], [31], but the present study used a small, nonrandomized two-cohort comparison and did not retain group-specific baseline statistics. Teacher effects, prior achievement, classroom composition, and selection could therefore contribute to the difference. The student N-gain of 0.739 offers convergent descriptive evidence of improvement, but it does not remove those threats to internal validity.

Practicality findings are important because an analytically ambitious module can fail if its instructional load exceeds the rhythm of pesantren learning. Both educator and student indices were above the study's predefined criterion, and the implementation record described the activities as usable with pacing adjustments. This is consistent with active-learning

evidence showing that participation is productive when tasks are structured rather than merely added to existing lessons [27], [33], [34]. Motivation theory likewise predicts that meaningful participation is more sustainable when learners experience growing competence and supported autonomy [35], [36]. Future work should quantify implementation fidelity, time-on-task, teacher adherence, and student participation rather than relying mainly on practicality judgments.

The female-majority context is educationally meaningful but should not be overinterpreted. Scholarship on women in Indonesian Islamic education shows that institutional design can shape opportunities for advanced religious learning and leadership development [8], [9]. In this setting, embedding higher-order interpretive work within the regular Tafsir curriculum may broaden routine access to academically demanding religious study. The present study, however, collected no gender-comparative outcome, agency scale, participation index, or confidence measure. The evidence therefore supports a narrower conclusion: the module was developed and found feasible in a female-majority pesantren. Whether the approach reduces gendered barriers or produces differential benefits for female learners remains a question for future gender-responsive research.

The findings also reinforce a broader position in Islamic education: disciplined reasoning and fidelity to tradition need not be treated as opposites. Contemporary Islamic-education scholarship argues for pedagogies that join revelation, reason, ethical formation, and critical engagement [5], [19], [44], [45], [46]. By operationalizing classical Qur'anic-sciences concepts as tools for analysis rather than isolated objects of memorization, the module offers one practical route for that synthesis. Its strongest contribution is therefore a replicable curriculum architecture with preliminary validity, practicality, and outcome evidence—not a universal claim of causal effectiveness.

For practice, Tafsir teachers can use the architecture as a planning rule: every unit should identify the interpretive question, select only the 'Ulumul Qur'an tools needed to answer it, design a student task that makes application visible, and score the justification with an aligned rubric. For institutions, the model suggests that curriculum integration can preserve classical sources while revising the relationship among subjects. For research, the next step is a preregistered multi-site study with baseline-equivalent or randomized classes, fully archived instruments, reliability evidence, fidelity measures, delayed retention tests, and explicit gender-responsive outcomes.

CONCLUSION

This study developed an integrated Tafsir–'Ulumul Qur'an module that reorganizes Qur'anic Sciences from a separate body of theory into contextual, structural, and methodological scaffolds for Tafsir interpretation. The product met the study's expert-validity and practicality criteria, and the retained learning evidence showed a high student N-gain together with a substantial final-score difference between the module and existing-instruction classroom cohorts. The contribution is strongest at the level of curriculum architecture: Tafsir remains the core subject, while Qur'anic-sciences concepts become tools that students must select, apply, and justify through aligned learning tasks and assessment. Because the outcome groups were nonrandomized and group-specific baseline equivalence was unavailable, the score

difference should be interpreted as promising preliminary evidence rather than a causal estimate. Within that boundary, the module offers a feasible, tradition-sensitive model for strengthening Qur'anic interpretive literacy in a female-majority pesantren.

LIMITATIONS

Interpretation is constrained by a single-site setting, a small nonrandomized outcome sample, unavailable group-specific pretest statistics, and the absence of individual-level demographic data needed to evaluate gender or prior-achievement differences. The available archive also lacks the full item-level validation/practicality matrices, rating-scale metadata, exact outcome-test item count, and reliability coefficients, so these statistics cannot be reconstructed. Implementation fidelity was described through practitioner feedback rather than a frequency-coded observation protocol, and the study did not test delayed retention or transfer to new Tafsir topics. These limitations restrict causal inference and generalizability; future research should use multi-site designs, archived and psychometrically evaluated instruments, explicit consent/assent documentation, fidelity measures, baseline-adjusted analyses, and longitudinal follow-up.

AUTHOR INFORMATION

Corresponding Author

Yusri Yaldi – Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi (Indonesia);
orcid.org/0009-0004-2794-8712
Email: yusriyaldi@gmail.com

Authors

Yusri Yaldi – Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi (Indonesia);
orcid.org/0009-0004-2794-8712
Wedra Aprison – Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi (Indonesia);
orcid.org/0009-0009-7879-2929
Yelfi Dewi S – Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi (Indonesia);
orcid.org/0000-0001-9963-8386
Untung Famuji – Universitas Islam Negeri Profesor Kiai Haji Saifuddin Zuhri Purwokerto (Indonesia);
orcid.org/0009-0006-7917-1321
Muhammad Safdar Bhatti – The Islamia University of Bahawalpur (Pakistan);
orcid.org/0000-0002-4535-2910

AUTHOR CONTRIBUTION

Y.Y., W.A., Y.D.S., U.F., and M.S.B. contributed to the conception and design of the study, development and evaluation of the learning module, interpretation of the findings, and preparation and critical revision of the manuscript. All authors reviewed and approved the final version and agree to be accountable for the integrity and accuracy of the work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DECLARATION OF USE OF AI IN SCIENTIFIC WRITING

During revision of this manuscript, the authors used ChatGPT (OpenAI) for language refinement, structural editing, and assistance with reference verification. The authors are responsible for reviewing and verifying the final content, interpretations, citations, factual claims, and compliance with the target journal's policies before submission.

REFERENCES

[1] Y. Sai, "Teaching Qur'an in Irish Muslim schools – curriculum, approaches, perspectives and implications," *British Journal of Religious Education*, vol. 40, no. 2, pp. 148–157, 2018 https://doi.org/10.1080/01416200.2016.1269723

[2] F. S. Tanjung, A. Kosasih, and C. Surahman, "Qur'anic literacy practices in Islamic religious education learning at senior high schools from lafdziyyah to amaliyah," *Asatiza: Jurnal Pendidikan*, vol. 7, no. 2, pp. 323–339, 2026 https://doi.org/10.46963/asatiza.v7i2.3687

[3] F. Aulia, M. N. A. Maulana, S. N. Asy-Syahida, Taufiqurrohman, and A. Nursobah, "Implementing Qur'an-Hadith learning to foster students' critical thinking skills," *Curricula: Journal of Curriculum Development*, vol. 4, no. 2, pp. 1781–1796, 2025 https://doi.org/10.17509/curricula.v4i2.92564

[4] H. K. Altinyelken, "Critical thinking and non-formal Islamic education: Perspectives from young Muslims in the Netherlands," *Contemporary Islam*, vol. 15, no. 3, pp. 267–285, 2021 https://doi.org/10.1007/s11562-021-00470-6

[5] C. Alkouatli, "Pedagogies in becoming Muslim: Contemporary insights from Islamic traditions on teaching, learning, and developing," *Religions*, vol. 9, no. 11, Art. no. 367, 2018 https://doi.org/10.3390/rel9110367

[6] R. Lukens-Bull, "Two sides of the same coin: Modernity and tradition in Islamic education in Indonesia," *Anthropology & Education Quarterly*, vol. 32, no. 3, pp. 350–372, 2001 https://doi.org/10.1525/aeq.2001.32.3.350

[7] F. K. Pohl, "Islamic education and civil society: Reflections on the pesantren tradition in contemporary Indonesia," *Comparative Education Review*, vol. 50, no. 3, pp. 389–409, 2006 https://doi.org/10.1086/503882

[8] E. Srimulyani, "Muslim women and education in Indonesia: The pondok pesantren experience," *Asia Pacific Journal of Education*, vol. 27, no. 1, pp. 85–99, 2007 https://doi.org/10.1080/02188790601145564

[9] S. A. Makruf, Sutrisno, and Z. Nuryana, "Islamic education for women in Madrasah Mu'allimat," *Jurnal Pendidikan Islam*, vol. 8, no. 2, pp. 191–206, 2022 https://doi.org/10.15575/jpi.v8i2.19452

[10] F. Sugiarto, "Kurikulum pendidikan madrasah di pesantren menjadi salah satu alternatif sistem pendidikan nasional," *El-Tsaqafah: Jurnal Jurusan PBA*, vol. 20, no. 1, pp. 65–81, 2021 https://doi.org/10.20414/tsaqafah.v20i1.2958

[11] C. R. Alid, Pardiman, and R. Basalamah, "Curriculum modernization strategy in Salaf Islamic boarding school based on Indonesian National Qualification Framework," *Nazhruna: Jurnal Pendidikan Islam*, vol. 5, no. 3, pp. 1118–1128, 2022 https://doi.org/10.31538/nzh.v5i3.2459

[12] F. A. Yahya, "Problem manajemen pesantren, sekolah dan madrasah: Problem mutu dan kualitas input-proses-output," *El-Tarbawi*, vol. 8, no. 1, pp. 93–109, 2015 https://doi.org/10.20885/tarbawi.vol8.iss1.art6

[13] H. Ahmad, "Integrasi Al-Qur'an dan ilmu sosial: Kontekstualitas Al-Qur'an dalam kehidupan bermasyarakat," *Ulumul Qur'an: Jurnal Kajian Ilmu Al-Qur'an dan Tafsir*, vol. 1, no. 2, pp. 1–15, 2021 https://doi.org/10.58404/uq.v1i2.69

[14] S. A. Acim and Suharti, "The concept of fiqh al-bi'ah in the Qur'an: A study of the Quranic verses on environment in the ulama's views of Lombok," *Ulumuna*, vol. 27, no. 1, pp. 115–140, 2023 https://doi.org/10.20414/ujis.v27i1.694

[15] S. Idris, "Sejarah perkembangan ilmu tafsir," *TAJDID: Jurnal Pemikiran Keislaman dan Kemanusiaan*, vol. 3, no. 2, pp. 174–187, 2019 https://doi.org/10.52266/tadjid.v3i2.294

[16] W. Wahyuddin and S. Saifulloh, "Ulum Al-Quran, sejarah dan perkembangannya," *Jurnal Sosial Humaniora*, vol. 6, no. 1, pp. 20–32, 2013 https://doi.org/10.12962/j24433527.v6i1.608

[17] S. M. Y. Gilani and T. Islam, "Qur'anic prerequisites for developing scientific mind," *Journal of Islam in Asia*, vol. 8, no. 3, pp. 229–253, 2012 https://doi.org/10.31436/jia.v8i0.259

[18] W. K. Abdul-Jabbar and Y. Makki, "Integrating intercultural philosophy into the high school curriculum: Toward a deliberative pedagogy of Tadabbur in diasporic Muslim education," *Religions*, vol. 15, no. 2, Art. no. 189, 2024 https://doi.org/10.3390/rel15020189

[19] M. Arif, "Fundamentals of Qur'anic Education Concept: A critical conceptualization with special reference to Al-Kilani's thought," *Global Journal Al-Thaqafah*, vol. 9, no. 2, pp. 15–26, 2019 https://doi.org/10.7187/gjat122019-2

[20] J. Biggs, "Enhancing teaching through constructive alignment," *Higher Education*, vol. 32, no. 3, pp. 347–364, 1996 https://doi.org/10.1007/BF00138871

[21] P. C. Abrami et al., "Strategies for teaching students to think critically: A meta-analysis," *Review of Educational Research*, vol. 85, no. 2, pp. 275–314, 2015 https://doi.org/10.3102/0034654314551063

[22] C. P. Dwyer, M. J. Hogan, and I. Stewart, "An integrated critical thinking framework for the 21st century," *Thinking Skills and Creativity*, vol. 12, pp. 43–52, 2014 https://doi.org/10.1016/j.tsc.2013.12.004

[23] P. A. Kirschner, J. Sweller, and R. E. Clark, "Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching," *Educational Psychologist*, vol. 41, no. 2, pp. 75–86, 2006 https://doi.org/10.1207/s15326985ep4102_1

[24] R. E. Mayer, "Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction," *American Psychologist*, vol. 59, no. 1, pp. 14–19, 2004 https://doi.org/10.1037/0003-066X.59.1.14

[25] J. Hattie and H. Timperley, "The power of feedback," *Review of Educational Research*, vol. 77, no. 1, pp. 81–112, 2007 https://doi.org/10.3102/003465430298487

[26] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies in Higher Education*, vol. 31, no. 2, pp. 199–218, 2006 https://doi.org/10.1080/03075070600572090

[27] M. T. H. Chi, "Active-constructive-interactive: A conceptual framework for differentiating learning activities," *Topics in Cognitive Science*, vol. 1, no. 1, pp. 73–105, 2009 https://doi.org/10.1111/j.1756-8765.2008.01005.x

[28] Maisyanah, A. Rahman, N. Kholis, M. Y. Abu Bakar, and A. A. Al Hasan, "Living Qur'an approach to improve critical thinking skills in Islamic Religious Education learning," *Al-Hayat: Journal of Islamic Education*, vol. 8, no. 4, pp. 1383–1398, 2024 https://doi.org/10.35723/ajie.v8i4.7

[29] U. Supriadi, M. Rahmat, G. E. Subakti, M. I. Firmansyah, E. Sumarna, and N. Nurjanah, "Qur'anic pedagogy for interreligious tolerance in Islamic Religious Education," *Jurnal Pendidikan Islam*, vol. 12, no. 1, pp. 48–61, 2026 https://doi.org/10.15575/jpi.v12i1.48465

[30] A. Fawait, K. Ayuni, and D. K. S. Wuyi, "The effect of problem-based learning on critical thinking skills in Islamic Religious Education: A quasi-experimental study in an Indonesian vocational high school," *Edukasi: Jurnal Pendidikan dan Pengajaran*, vol. 13, no. 1, pp. 396–413, 2026 https://doi.org/10.19109/ejpp.v13i1.34114

[31] A. S. Alam, Fadlillah, and M. R. Muhyiddin, "From dogma to dialogue: Reframing Islamic Religious Education through dialogic pedagogy in Indonesian Madrasah Aliyah," *Journal of Islamic Education Research*, vol. 6, no. 4, pp. 427–442, 2025 https://doi.org/10.35719/jier.v6i4.523

[32] R. Riyanto, Waharjani, R. Astari, D. Perawironegoro, and B. M. Bustam, "Implementation of Islamic Religious Education integrated with the Qur'an-based environment in SD Alam Pacitan," *TADRIS: Jurnal Pendidikan Islam*, vol. 20, no. 2, pp. 183–201, 2026 https://doi.org/10.19105/tjpi.v20i2.18993

[33] M. Prince, "Does active learning work? A review of the research," *Journal of Engineering Education*, vol. 93, no. 3, pp. 223–231, 2004 https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

[34] S. Freeman et al., "Active learning increases student performance in science, engineering, and mathematics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 23, pp. 8410–8415, 2014 https://doi.org/10.1073/pnas.1319030111

[35] E. L. Deci and R. M. Ryan, "The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior," *Psychological Inquiry*, vol. 11, no. 4, pp. 227–268, 2000 https://doi.org/10.1207/S15327965PLI1104_01

[36] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions," *Contemporary Educational Psychology*, vol. 61, Art. no. 101860, 2020 https://doi.org/10.1016/j.cedpsych.2020.101860

[37] S. A. Barab and K. Squire, "Design-based research: Putting a stake in the ground," *Journal of the Learning Sciences*, vol. 13, no. 1, pp. 1–14, 2004 https://doi.org/10.1207/s15327809jls1301_1

[38] A. L. Brown, "Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings," *Journal of the Learning Sciences*, vol. 2, no. 2, pp. 141–178, 1992 https://doi.org/10.1207/s15327809jls0202_2

[39] W. A. Sandoval and P. Bell, "Design-based research methods for studying learning in context: Introduction," *Educational Psychologist*, vol. 39, no. 4, pp. 199–201, 2004 https://doi.org/10.1207/s15326985ep3904_1

[40] T. Anderson and J. Shattuck, "Design-based research: A decade of progress in education research?" *Educational Researcher*, vol. 41, no. 1, pp. 16–25, 2012 https://doi.org/10.3102/0013189X11428813

[41] L. R. Aiken, "Three coefficients for analyzing the reliability and validity of ratings," *Educational and Psychological Measurement*, vol. 45, no. 1, pp. 131–142, 1985 https://doi.org/10.1177/0013164485451012

[42] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *American Journal of Physics*, vol. 66, no. 1, pp. 64–74, 1998 https://doi.org/10.1119/1.18809

[43] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," *Frontiers in Psychology*, vol. 4, Art. no. 863, 2013 https://doi.org/10.3389/fpsyg.2013.00863

[44] F. Panjwani, "Fazlur Rahman and the search for authentic Islamic education: A critical appreciation," *Curriculum Inquiry*, vol. 42, no. 1, pp. 33–55, 2012 https://doi.org/10.1111/j.1467-873X.2011.00574.x

[45] A. Sahin, "Critical issues in Islamic education studies: Rethinking Islamic and Western liberal secular values of education," *Religions*, vol. 9, no. 11, Art. no. 335, 2018 https://doi.org/10.3390/rel9110335

[46] J. M. Halstead, "An Islamic concept of education," *Comparative Education*, vol. 40, no. 4, pp. 517–529, 2004 https://doi.org/10.1080/0305006042000284510