JURNAL CENDEKIA : Media Komunikasi Penelitian dan Pengembangan Pendidikan Islam
Volume 08, Issue 02, October 2026, Pages 376-390 p-ISSN: 2086-0641 (Print)
e-ISSN: 2685-046X (Online)

DOI: https://doi.org/10.37850/cendekia.v18i02.1407
https://journal.faibillfath.ac.id/index.php/cendekia.

THE TAM-TAMAN TRADITION AS AN AUTHENTIC PROCESS-BASED ASSESSMENT OF CLASSICAL ISLAMIC TEXTS LEARNING IN PESANTREN

Ubaidillah[1]*, Zaenal Arifin[2], Ahmad Badi[3]
Universitas Islam Tribakti Lirboyo Kediri, Indonesia
email: Faruqubaidillah057@gmail.com[1)], zae.may13@gmail.com [2)], badivivin2010@gmail.com[3)]
* Corresponding Author

Received 18 August 2026; Received in revised form 18 September 2026; Accepted 30 October 2026

Abstract

This study examines tam–taman, the semester-end inspection of the completeness of students' interlinear glosses (word-by-word meanings written between the lines of the text), as an artifact-based assessment. A qualitative case study was conducted. We gathered data through non-participant observation, semi-structured interviews, and document analysis, and validated them through triangulation of source, technique, and time. The findings show that tam–taman runs in six stages: announcement, student preparation, text collection, inspection by correctors, determination of tam (complete) or naqis (incomplete) status, and re-inspection of incomplete texts. It serves as evidence-based verification of learning engagement, an academic gate to final examinations, a remedial cycle, and a means of habituating discipline within a culture that binds knowledge to writing. However, completeness of glossing does not guarantee accuracy of understanding. The study offers a critical account of how tradition-based, artifact-based evaluation verifies learning engagement in pesantren and where its validity ends, and it points to accuracy-verification mechanisms as a practical improvement. This study contributes to culturally responsive artifact-based assessment by conceptualizing tam–taman as a mechanism for verifying learning engagement and academic discipline, while identifying the need to complement completeness-based evaluation with accuracy verification to ensure actual student comprehension.

Keywords: Artifact-Based Assessment; Classical Islamic Texts; Interlinear Glossing; Pesantren; Learning Evaluation.

INTRODUCTION

Learning evaluation determines whether instructional objectives have been achieved and provides a basis for improving teaching (Rifkoh et al., 2025). In Islamic education, it extends beyond cognitive mastery to attitudes, skills, and character (Sholahudin et al., 2025). Studies of pesantren evaluation report that it is continuous, running from daily assessment to the end of each level, and covers knowledge, skills, and moral conduct (Maghfiroh & Nursikkin, 2023). Evaluation is reported to operate through student activeness, weekly exercises, and annual memorisation (Azka et al., 2024)

The study of classical Islamic texts (yellow book) is central to pesantren education and is characterized by diligence, continuity, and close student–teacher relations (Adim, 2026). These texts are printed without vowel marks, so reading

them depends on mastery of Arabic syntax and morphology (*nahwu* and *sharaf*), which is why these subjects are taught as a foundation from the beginner level (Ulum & Nuriyah, 2023). Students write meanings and grammatical notes between the lines of the text, a practice known as *maknani* or *makna gandul* (interlinear glossing), which together with commentaries and marginal notes forms part of how the text is transmitted and understood. The teaching of *kitab kuning* has been described as a cycle of planning, implementation, and evaluation (Ja'far, 2025; Kurniasari, 2022)

Assessment scholarship distinguishes test techniques from non-test techniques that evaluate attitudes, participation, and discipline (Sunaryati et al., 2024). Within the latter, authentic assessment judges real products and activities generated during learning rather than isolated test answers (Hidayati et al., 2025; Sakti, 2022), and portfolio assessment collects such products as documented evidence of learning over a period. These concepts offer a lens for examining a pesantren practice in which a learning artifact, the glossed text, is inspected.

Previous research can be grouped by what is evaluated. The first strand examines evaluation systems and outcomes in Islamic education and pesantren and shows that evaluation is continuous and multi-domain, but mostly at the level of the system or institution type (Achmadin et al., 2024; Maghfiroh & Nursikkin, 2023; Nopita & Saputra, 2025). The second strand applies program-evaluation frameworks, above all the Context, Input, Process, Product (CIPP) model, to a student-pesantren program. (Antariksa et al., 2022) Qur'an memorisation (*tahfiz*) programs (Azis, 2023; Riza et al., 2024) and an Islamic education study program (Hasriadi et al., 2024) A systematic review describes CIPP as a flexible framework covering context, input, process, and product (Riawan et al., 2025) These studies mostly assess program quality or outcomes. The third strand concerns evaluation within *kitab kuning* learning itself. It includes tests and daily observation in *nahwu–sharaf* instruction. (Jailani et al., 2024). and the individual text-reading method, where students read unvocalised texts fluently but their understanding of content remains limited and their evaluation is not yet systematic. (Fuadi et al., 2025). It also includes the Lirboyo system of activeness, *tamrin*, and *mukhafadhoh* (Azka et al., 2024), together with accounts of *kitab kuning* teaching as planned, implemented, and evaluated (Ja'far, 2025b; Kurniasari, 2022). One of these studies examines *tam–taman*, its mechanism, and what gloss inspection can and cannot verify.

What is known, then, is that pesantren evaluation is broader than examinations and that program models have been widely applied Antariksa et al., (2022); Azka et al., (2024); Maghfiroh & Nursikkin, (2023) What is missing is an account of how a tradition-based practice that inspects a learning artifact verifies the learning process, and where that verification stops. To the best of the authors' knowledge, no study has examined *tam–taman* in this way. This study therefore aims to (1) describe the practice and mechanism of *tam–taman* as an evaluation of

kitab kuning learning, and (2) analyze its functions and implications, including the limits of its validity. It contributes an empirically grounded, critical account of artifact-based assessment in a pesantren setting and clarifies the boundary between evidence of learning engagement and evidence of learning mastery.

METHOD

This study used a qualitative case study design. The case is the practice of tam-taman at Madrasah Hidayatul Mubtadiin (MHM), a madrasah within Pondok Pesantren Lirboyo, Kediri, East Java. MHM was chosen because tam-taman there is institutionalized: documents trace it from an informal *koreksian kitab* in the 1970s to an official program in 1984-1985, with a dedicated committee from 1989-1990 (Tim Dokumentasi Sejarah MHM, 2025). The practice is also embedded in a structured kitab kuning curriculum. Informants were selected purposively, in consultation with the madrasah administration, for their direct involvement in the practice: the madrasah head (MH), two mustahiq who serve as correctors (C1 and C2), and one student (S1). The small number reflects the study's focus on a single bounded practice. Three of the four informants are administrators or teachers and only one is a student, which is acknowledged as a limitation. Informants are identified only by role codes to protect confidentiality.

Data were collected between January and April 2026 through interviews, non-participant observation, and documentation. Semi-structured interviews followed a flexible guide and were held with MH (14 March 2026), C1 (20 April 2026), S1 (22 April 2026), and C2 (29 April 2026). We audio-recorded them with permission and transcribed them. We conducted non-participant observation on 3 January 2026 in Grade 1 Tsanawiyah and Grade 3 Aliyah classes. It covered kitab inspection, determination of tam or naqis status, corrector-student interaction, and students' readiness, and was recorded in field notes. Documents included the 2026 HSPK (Outcome of the Small Committee Meeting), schedules, rules, samples of stamped kitab, photographs, and the MHM history book. The pesantren leadership and the madrasah granted permission for the research, and recording and photography were carried out with permission.

Data were analyzed through data condensation, data display, and conclusion drawing and verification (Nisa et al., 2026). Transcripts and field notes were coded, and the codes were grouped into categories and themes under the two research questions. Credibility was pursued through triangulation of sources (head, correctors, student), techniques (interview, observation, documents), and time (observation in January, interviews in March-April) (Faot et al., 2026).

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

RESULTS AND DISCUSSION

RESULTS

Practices and Mechanisms of The *Tam-Taman* Tradition as An Evaluation Method for *Kitab Kuning* Learning

Tam–taman at MHM is a semester-end inspection of the completeness of students' interlinear glosses in the *kitab kuning* (classical Islamic texts) studied during the semester. The activity is conducted in each class before the final examinations. Three parties are involved: the madrasah head, who oversees the activity; teachers appointed by the madrasah as correctors, usually four to five teachers per class; and the students whose texts are inspected. The inspection schedule varies by educational level. For primary and lower-secondary students, it takes place from 07.00 to 11.00, whereas for upper-secondary and higher-level students, it is conducted from 19.00 to 23.00.

Historical documents indicate that the practice dates back to the 1970s, when class teachers informally inspected students' texts. During a madrasah meeting in the 1983–1984 academic year, the practice was proposed for formalisation in response to the increasing number of students whose texts were incomplete or insufficiently glossed. It was subsequently established as an official programme in the 1984–1985 academic year and has been coordinated by an inspection committee since 1989–1990 (Tim Dokumentasi Sejarah MHM, 2025). According to the madrasah head, this tradition is rooted in the teachings of senior religious leaders and the maxim *qayyidū al-ʿilma bil-kitābah* (bind knowledge with writing). This principle encourages students to preserve what they learn by recording and providing complete glosses in their learning texts.

The implementation of tam–taman follows six stages. First, a teacher appointed by the madrasah announces the inspection schedule, the texts that students must submit, and the re-inspection timetable several weeks before the activity. This advance notice allows students to identify the books to be examined and prepare them accordingly. MH explained:

> "Before tam–taman is held, students are usually already informed about the schedule and the requirements for the books that must be submitted" (MHM, Madrasah Hidayatul Mubtadiin Lirboyo, 2025).

Second, students prepare all the kitab kuning studied during the semester by ensuring that their interlinear glosses are recorded throughout the relevant texts. Third, several hours before the inspection begins, students submit their books to the designated collection points. Based on non-participant observation, students brought their books and arranged them individually in front of the assigned correctors. The books were organised according to the students and the corrector groups designated by the madrasah, with each group consisting of one or more teachers. This arrangement enabled the correctors to examine the submitted texts

systematically. Figure 1 illustrates how students prepared and arranged their books before the inspection.

Figure 1. Students preparing and arranging their *kitab kuning*

Document analysis further reveals that the books submitted for tam–taman contain printed Arabic passages with handwritten interlinear glosses. These handwritten annotations represent students' written records of the meanings of individual words in the texts studied during the semester. Their placement between the lines allows the correctors to examine the extent to which students have completed the required glosses across the assigned material. The examined page also contains handwritten notes and a red institutional stamp, which form part of the documentary features visible in the submitted text. Thus, the book serves not merely as a learning resource but also as a tangible record of students' engagement with the learning process. Figure 2 presents an example of a kitab kuning page containing handwritten interlinear glosses, as examined in the document analysis.

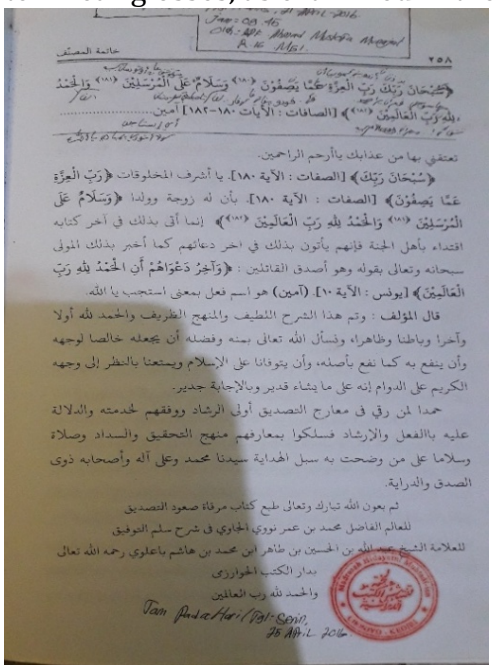

Figure 2. A page of kitab kuning showing handwritten interlinear glosses

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

The document analysis further reveals that the books submitted for tam–taman contain printed Arabic passages accompanied by handwritten interlinear glosses. These annotations represent students' written records of the meanings of individual words in the texts studied during the semester. Their placement between the lines allows the correctors to assess the completeness of students' glosses across the assigned material. Importantly, the inspected page also bears a red institutional stamp, which serves as tangible evidence that the assigned teacher (corrector) has examined the text. The stamp therefore functions as an administrative marker of the inspection process, linking the physical learning artifact to the institutional assessment procedure. In this regard, the submitted book serves not merely as a learning resource but also as documented evidence of students' engagement with learning and completion of the inspection process.

The Functions and Implications of The Tam–Taman Tradition in The Study of *Kitab Kuning*

The findings from interviews, classroom observations, and document analysis reveal three interconnected functions of the tam–taman tradition in the study of kitab kuning: documenting students' learning participation, providing continuous academic assessment and remediation, and reinforcing learning discipline and responsibility. These functions demonstrate that tam–taman is not merely a book-completion activity but is embedded in the madrasah's broader learning practices.

The first finding indicates that the glossed kitab serves as tangible evidence of students' participation in the learning process. Rather than relying solely on physical attendance, teachers use the presence and completeness of written glosses (*makna gandul*) to indicate students' engagement with the lessons.

MH explicitly emphasized this function:

> "Tam-taman ensures that students' participation is not only a matter of attendance, but there must be real evidence in their books." (MH, personal communication, March 14, 2026)

This statement suggests that participation is understood not simply as being physically present in the classroom but also as actively following the learning process through written annotations in the kitab. A similar perspective was expressed by C1, who explained:

> "A full book usually indicates that the student has followed the lessons well." (C1, personal communication, April 20, 2026)

These interview accounts are consistent with the document analysis, which identifies students' glossed kitab as one of the materials used to assess learning completion. The written annotations provide a visible record that teachers can

inspect when evaluating students' progress. Thus, the completeness of the kitab functions as an indicator of learning participation within the madrasah's established assessment practices. However, completeness should be understood as evidence of documented participation rather than definitive proof of students' mastery of the material.

The second finding shows that tam–taman functions as a continuous assessment mechanism that connects book inspection with remedial action and subsequent re-examination. The process involves checking students' glosses, determining whether their books have reached *tam* (complete) or *naqis* (incomplete) status, and requiring further completion when deficiencies are identified. MH described the procedure and its implications for students' academic progression:

> "If a student is not yet tam, his learning is not yet finished, so he cannot proceed to the exam. The solution is to patch the gaps quickly and then take the re-examination." (MH, personal communication, March 14, 2026)

This explanation illustrates that an incomplete kitab does not immediately result in a final judgment of failure. Instead, students are given an opportunity to address the identified gaps and undergo another inspection. In this respect, tam–taman incorporates a corrective process in which assessment results are followed by further learning activities. C2 also emphasized that the practice has a broader educational function:

> "Tam-taman is not only about passing or failing." (C2, personal communication, April 29, 2026)

This statement reinforces the understanding that tam–taman is not exclusively concerned with determining whether students meet a completion requirement. It also provides a mechanism to identify incomplete learning records and direct students to the necessary follow-up. The findings further indicate that tam status serves as an academic requirement for entering the final examination. Students whose books remain naqis must first complete the missing glosses and undergo re-examination before proceeding. Consequently, tam–taman occupies two related positions within the learning system: it functions as an assessment mechanism and as a procedural requirement for academic progression.

The third finding concerns the relationship between tam–taman and the formation of learning discipline and responsibility. According to the informants, the requirement to complete glosses throughout the learning process encourages students to maintain consistency rather than postponing their work until the inspection period. MH explained this relationship as follows:

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

> "Students become accustomed to discipline because if they do not gloss from the beginning, they will face difficulty at tam-taman." (MH, personal communication, March 14, 2026)

This account indicates that the inspection process encourages students to maintain their written annotations regularly. The possibility of receiving naqis status gives students an institutional reason to avoid accumulating unfinished work. In this sense, discipline is connected to the expectation that students take responsibility for documenting the lessons as they progress. C1 also situated the practice within the madrasah's educational tradition:

> "What is guarded is not only the final result but also the learning process." (C1, personal communication, April 20, 2026)

This perspective suggests that teachers understand tam–taman as part of a broader tradition of maintaining continuity in learning. Its significance lies not only in determining whether a book is complete but also in preserving an established practice through which students are expected to follow, record, and complete their lessons. Regarding students' understanding of the material, S1 described the perceived contribution of regular maknani activities:

> "If we often do maknani, gradually we understand it by ourselves, especially in nahwu and sharaf." (S1, personal communication, April 22, 2026)

This statement reflects one student's perception that repeated glossing helps develop an understanding of Arabic grammatical structures, particularly nahwu and sharaf. It indicates how students may perceive the relationship between written glossing and comprehension. Nevertheless, this finding represents an individual account rather than evidence of a measurable improvement in learning outcomes. The available data do not establish whether students who regularly engage in maknani demonstrate higher levels of understanding than those who do not.

An additional finding from the interviews and classroom observation concerns the distinction between the completeness and accuracy of students' glosses. According to C1 and C2, the inspection primarily focuses on whether the glossing is complete. This suggests that the assessment procedure emphasizes the presence of written meanings rather than systematically verifying the linguistic or conceptual accuracy of every annotation.

This distinction was also reflected in the classroom observation conducted on January 3, 2026. During the observed inspection, some students' written meanings differed from the explanations previously provided by the teacher. The discrepancy was visible when the students' annotations were compared with the teacher's explanations. In the context of the inspection, these students' books were still

presented for completion checking, although the differences raised questions about whether the written glosses accurately represented the material taught.

The observation suggests a potential limitation in an assessment procedure that prioritizes the completeness of written annotations. A book may include glosses across the required sections without demonstrating that the meanings the student wrote correspond to the teacher's explanation. Therefore, the designation tam should not automatically be interpreted as confirmation of students' conceptual or linguistic accuracy.

However, this observation was recorded on only one day and involved few instances. The available evidence does not establish how frequently such discrepancies occur, whether students deliberately produce them to obtain tam status, whether they copy glosses from peers, or whether teachers apply particular sanctions when they identify inaccuracies. Accordingly, the finding is presented as an observed discrepancy between written glosses and teacher explanations, rather than as evidence of systematic manipulation or a general weakness in the entire assessment process.

DISCUSSION

Positioning tam–taman in pesantren evaluation**.** The findings show that *tam–taman* turns a learning artifact, the glossed text, into evidence that learning took place across a semester, and links this evidence to a decision (*tam* or *naqis*) and a follow-up (re-inspection). Pesantren evaluation has been described as continuous and multi-domain (Maghfiroh & Nursikkin, 2023; Nopita & Saputra, 2025) and at Lirboyo it is reported to rely on student activeness, weekly exercises, and annual memorisation (Azka et al., 2024). *Tam–taman* adds a different element to this repertoire: a written trace of the whole semester's learning that can be inspected after the fact, whereas the techniques reported for *nahwu–sharaf* learning are tests and observation of performance (Jailani et al., 2024). The central question is therefore what kind of assessment this is and what it can validly claim.

Authentic assessment, with limits. The practice resembles authentic and portfolio-based assessment in two respects: it uses a real product of learning rather than a test, and it provides a path from judgement to improvement through re-inspection (Hidayati et al., 2025; Sakti, 2022). It departs from these approaches in important ways. Authentic assessment draws on meaningful tasks, explicit criteria, and feedback on quality (McTighe et al., 2020). The data show a single criterion (completeness) and no performance task. A glossed text is also not a portfolio in the strict sense, since portfolio assessment presupposes selection of work, reflection, and quality criteria. Within the non-test techniques described by Sunaryati et al., (2024), the practice is better described as artifact-based verification: it confirms that the activity of learning occurred rather than how well it was understood. This is a narrower claim than the authentic assessment label suggests, and it is the claim the data can support.

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Summative and formative functions. *Tam–taman* combines two functions that evaluation texts usually separate. It is summative in that it gates eligibility for the final examination, and formative in that *naqis* students are given a route to complete what is missing (Syaifudin et al., 2025). However, because the feedback concerns completion and not understanding, whereas evaluation is expected to support reflection and the improvement of learning (Isnadia et al., 2024). Glossing and the question of understanding. One student attributed growing understanding of *nahwu* and *sharaf* to repeated glossing. Interlinear glossing has long served to make Arabic texts accessible in pesantren, alongside commentaries and marginal notes (Zainuddin et al., 2026). Glossing can involve grammatical reasoning, as when students apply *nahwu* and *sharaf* to decide a word's meaning and function, but it can equally be copied from a teacher's dictation (Ulum et al., 2023). Research on glossing in second-language reading shows that learning gains depend on gloss format and on learner and text characteristics, so the mere presence of glosses is not evidence of learning (Yanagisawa et al., 2020). Because the *tam–taman* inspection checks only that glosses are present, it cannot tell which kind of learning has occurred. The student's account is consistent with a developmental effect, but one perception cannot establish it.

The most critical issue raised by the findings is the gap between completeness and accuracy. Authentic assessment is expected to meet principles of validity, objectivity, fairness, and accountability. Where the criterion is completeness, writing any meaning in the space can become a rational way to obtain *tam*. Academic cheating in pesantren is related to self-efficacy, religiosity, and academic integrity (Lestari et al., 2022), which suggests the risk is not trivial in this setting. This study cannot say how often it happens at MHM, but it identifies a weakness in the practice's design (Sani et al., 2025). Safeguards that could address it, though not tested here, include oral spot-checks on sampled passages (Nizar et al., 2022). A brief accuracy rubric and cross-checking against the teacher's dictation notes (Arif et al., 2025). Such measures would preserve the cultural meaning of *tam–taman* while strengthening what the stamp can claim.

The pesantren's own framing helps explain why the practice persists despite its limits. and glossing is a long-standing para-textual practice of the pesantren tradition (Adim, 2026). Seen this way, a text with complete glosses is a record of faithful learning. The discipline that informants described is consistent with research on habituation as a route to disciplined character (Arif, 2024), and the semester-end inspection supplies the accountability that makes such habituation stick. These accounts remain interpretations of informants' perceptions, not measurements of changes in character.

Implications. Theoretically, the study suggests artifact-based verification as a more precise category than portfolio-based authentic assessment for practices like *tam–taman*, and it makes explicit the boundary between evidence of engagement

and evidence of mastery (Hidayati et al., 2025). Practically, madrasah could retain *tam–taman* and re-inspection while adding an accuracy check such as oral sampling or a short rubric (Sakti, 2022), and could record *tam* and *naqis* data to monitor learning over time. For policy, quality assurance for pesantren could recognize tradition-based, evidence-based evaluation as a legitimate component alongside tests (Maghfiroh & Nursikkin, 2023; Nopita & Saputra, 2025), provided its validity limits are acknowledged.

CONCLUSION AND SUGGESTION

Tam-taman is a structured six-stage practice, from announcement to re-examination, in which correctors check whether the glossing in all kitab studied is complete and assign tam or naqis status. Informants described it as evidence of sustained learning, an academic gate to the final examination, a cycle of assessment and remediation, and a way to encourage discipline within a culture of binding knowledge with writing. Third, its implications are bounded. Because the criterion is completeness, a tam stamp signals engagement rather than mastery or accuracy. Tam-taman is therefore best understood as artifact-based verification that complements, and does not replace, assessments of understanding.

Madrasahs that keep the practice are advised to add an accuracy check, such as oral spot-checks on sampled passages or a short rubric, and to record tam and naqis data. Future research should compare tam-taman across pesantren, link tam and naqis status to examination results through mixed methods, include more student perspectives, and test verification mechanisms.

ACKNOWLEDGEMENTS

The authors express sincere gratitude to all who provided support, guidance, and assistance throughout the preparation of this article. Special appreciation is extended to Madrasah Hidayatul Mubtadiin (MHM) Lirboyo for granting permission to conduct this research and for facilitating the research process. The authors also gratefully acknowledge all informants for their valuable information and data. Finally, the authors thank their academic supervisor and everyone who offered guidance, constructive feedback, and continuous support that made completing this article possible.

AUTHOR CONTRIBUTIONS

The author contributed to all stages of this article, including formulating the research idea and focus, collecting data through observation and interviews, analyzing and interpreting data, conducting the literature review, preparing the manuscript, and revising and finalizing the article. The author is also responsible for the accuracy of the data, the substance of the discussion, and the overall content of the article.

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

CONFLICT OF INTEREST

The author declares that this article was prepared as part of the academic requirements. Nevertheless, the author declares that there are no conflicts of interest that could have influenced the research process, data analysis, or the presentation of the research findings. The research and preparation of this article were conducted objectively and in accordance with the data obtained from the field.

DECLARATION OF AI

The author used artificial intelligence (AI) tools, specifically ChatGPT (OpenAI) and Google Translate, to assist in preparing this manuscript. These tools supported language translation, improved the clarity and readability of the text, and refined the academic writing. The author conducted, verified, and approved all research design, data collection, data analysis, interpretation of the findings, and the final content of the manuscript. The author assumes full responsibility for the accuracy, originality, and integrity of the work.

REFERENCES

Achmadin, B. Z., Zuhriyah, I. A., & Barizi, A. (2024). Evaluation Model of Islamic Religious Education Learning Program at Madrasah'Aliyah Level in East Java. *Al-Riwayah: Jurnal Kependidikan*, *16*(2), 185–212. https://doi.org/https://doi.org/10.47945/al-riwayah.v16i2.1482

Adim, A. A. K. (2026). *Metode Pembelajaran Kitab Kuning di Pesantren Salaf dan Relevansinya di Era Modern*. *3*(2). https://doi.org/https://doi.org/10.62017/syariah.v3i2.6609

Antariksa, W. F., Fattah, A., & Utami, M. A. P. (2022). Evaluasi program pendidikan pesantren mahasiswa model CIPP (Context, Input, Process, Product). *Evaluasi: Jurnal Manajemen Pendidikan Islam*, *6*(1), 75–86. https://doi.org/http://doi.org/10.32478/evaluasi. v6i1. 848

Arif, M., Nasir, R., & Ma'arif, M. A. (2025). The Kitab Kuning Learning Model in the Development of Student Expertise in Pesantren-Based Higher Education. *Nazhruna: Jurnal Pendidikan Islam,* 8(1), 52–74. https://doi.org/10.31538/nzh.v8i1.8

Arif, S. (2024). *Penanaman Nilai-Nilai Pendidikan Karakter Berbasis Metode Pembiasaan Di Pondok Modern Darussalam Gontor Putri*. Uin Raden Intan Lampung Press.

Azis, A. B. (2023). *Evaluasi program tahfidzul Al-qur'an di pondok pesantren*. UIN Fatmawati Sukarno Press.

Azka, M. F., Masita, A., & Kibtiyah, A. (2024). Implementasi Metode Pembelajaran Dan Evaluasi Pembelajaran Di Pondok Pesantren Lirboyo. *Jurnal Studi Multidisipliner*, *8*(7). https://doi.org/https://doi.org/10.58578/tsaqofah.v4i3.3046

Faot, O., Tabuni, N., Demmy Tabuni, S., Ivona Dorobia, S., Rumawak, S., MAppLing, S., Nurwiati, S., Sos, M., Kaipman, P., & Rumaseb, S. H. (2026). *Metode penelitian kualitatif*. Goresan Pena.

Fuadi, M. N., Afifah, A. R., & Azizah, Y. L. (2025). Evaluasi Metode Sorogan Dalam Pembelajaran Kitab Kuning Menggunakan Model CIPP: Studi Kasus di Pondok Pesantren Alluqmaniyyah Yogyakarta. *At-Tadbir: Jurnal Manajemen Pendidikan Islam*, *5*(2), 102–112. https://doi.org/https://doi.org/10.51700/attadbir.v5i2.1051

Hasriadi, H., Pamessangi, A. A., Husna, M., & Armiyanti, A. (2024). Evaluasi Pembelajaran di Program Studi Pendidikan Agama Islam FTIK IAIN Palopo dengan Model Evaluasi CIPP. *IQRO: Journal of Islamic Education*, *7*(2), 551–565. https://doi.org/https://doi.org/10.24256/iqro.v7i2.6554

Hidayati, F., Nurhalisa, S., Radita, A., Candrika, A., Nisa, N., & Abdullah, S. (2025). Authentic assessment sebagai strategi evaluasi efektif dalam pembelajaran PAI berbasis karakter. *Jurnal Pendidikan Tambusai*, *9*(2), 18131–18140. https://doi.org/https://doi.org/10.31004/jptam.v9i2.28844

Isnadia, R., & Safitri, D. A. (2026). Management of the integration of kitab kuning in PAI teaching materials: Optimization of the sorogan and bandongan methods in contemporary madrasahs. *MANAGIERE: Journal of Islamic Educational Management,* 5(1), 55-64. https://doi.org/10.35719/managiere.v4i2.2454

Ja'far, J. (2025). Designing, Implementing, and Evaluating: Management of Teaching the Kitab kuning in Pesantren. *Jurnal Inovatif Manajemen Pendidikan Islam*, *4*(2), 40–49. https://doi.org/10.38073/jimpi.v4i2.2611

Jailani, M., Miranda, A., Taufikurrahman, Moh., & Al Faruqi, M. I. H. (2024). Implikasi dan Konsep Dasar Evaluasi dalam Pembelajaran Nahwu dan Shorof di Pesantren. *Journal of Education Research*, *5*(4), 4496–4506. https://doi.org/10.37985/jer.v5i4.1457

Kurniasari, A. (2022). Manajemen pembelajaran kitab kuning di pondok pesantren. *Jurnal Studi Islam Dan Kemuhammadiyahan (JASIKA)*, *2*(1). https://doi.org/https://doi.org/10.18196/jasika.v2i1.22

Lestari, W. I., Latifa, R., & Rahmah, M. (2022). Academic cheating in pesantren: How self efficacy, religiosity, and academic integrity predict academic cheating. *TARBIYA: Journal of Education in Muslim Society*, 1–18. https://doi.org/https://doi.org/10.15408/tjems.v9i1.24701

Maghfiroh, N., & Nursikkin, M. (2023). Development of the Evaluation System of Islamic Education in Indonesian Pondok Pesantren. *At Turots: Jurnal Pendidikan Islam*, 35–46. https://doi.org/https://doi.org/10.51468/jpi.v5i2.169

McTighe, J., Doubet, K. J., & Carbaugh, E. M. (2020). *Designing Authentic Performance Tasks and Projects: Tools for Meaningful Learning And Assessment*. ASCD.

Nisa, R., Abdullah, G., Mala, A., Rofi'ah, U. A., Sunariyah, A., Muhaimin, M. Z., Zuhriyah, N., & Puspita, A. G. (2026). *Metode Penelitian Praktis: Panduan Menyusun Karya Ilmiah*. Asta Cipta Mandiri indonesia.

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Nizar, M. J., Abror, T., & Samsuri, H. (2025). Implementasi Metode Al-Miftah Lil Ulum Dalam Meningkatkan Standar Mutu Kemampuan Membaca Kitab Kuning Santri Di PP Roudlotut Thohirin Sengon Pringgodani Bantur Malang. *AL-MIKRAJ Jurnal Studi Islam Dan Humaniora,* 5(2), 1067–1079. https://doi.org/10.37680/almikraj.v5i2.6808

Nopita, R., & Saputra, D. (2025). Evaluasi hasil belajar pendidikan islam di madrasah dan pondok pesantren. *IMTIYAZ: Jurnal Ilmu Keislaman*, *9*(2), 310–325. https://doi.org/https://doi.org/10.46773/imtiyaz.v9i2.1997

Riawan, I. M. O., Citariani, N. M., & Lokasanti, I. K. (2025). Efektivitas model evaluasi CIPP dalam program pendidikan: Kajian literatur sistematis. *Jurnal Evaluasi Pendidikan*, *16*(2), 108–115. https://doi.org/https://doi.org/10.21009/jep.v16i2.58457

Rifkoh, N., Hasanah, S., Atiullah, M., & Affandi, M. (2025). Pendekatan Evaluasi Pembelajaran dan Jenis Evaluasi Dalam Pembelajaran. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, *11*(04), 212–225. https://doi.org/https://doi.org/10.36989/didaktik.v11i04.10505

Riza, H., Maufur, M., & Basukiyatno, B. (2024). Evaluasi Progam Tahfid Al Quran sebagai Unggulan Pondok Pesantren Nurul Hayah Ketanggungan. *Journal of Education Research*, *5*(3), 3055–3071. https://doi.org/https://doi.org/10.37985/jer.v5i3.1412

Sakti, B. P. (2022). Types of Authentic Learning Assesment. *AT-TA`DIB*, 6(2), 342–350. https://doi.org/10.32832/at-tadib.v6i2.19545

Sani, A., & Majeed, M. K. (2025). Transforming kitab kuning literacy in the digital era: Challenges and future prospects for Islamic education. *Al'Adalah*, 28(1), 45-72. https://doi.org/10.35719/aladalah.v28i1.397

Sholahudin, T., Abid, I., Ikhwanudin, M., Arrizky, M. N., & Al-Ghozali, U. M. (2025). Evaluasi hasil Pembelajaran Pendidikan Agama Islam (PAI) Tinjauan terhadap Ayat Al-Qur'an dalam Aspek Kognitif, Afektif, dan Psikomotorik. *Ainara Journal (Jurnal Penelitian dan PKM Bidang Ilmu Pendidikan)*, *6*(1), 165–171. https://doi.org/10.54371/ainj.v6i1.808

Sunaryati, T., Laelly, T. A., Febriyanti, U., Noviyanti, N., & Apriliani, F. (2024). Analisis komprehensif terhadap jenis-jenis evaluasi pembelajaran sekolah dasar: Tinjauan literatur. *INNOVATIVE: Journal Of Social Science Research*, *4*(4), 13339–13354. https://doi.org/https://doi.org/10.31004/innovative.v4i4.13463

Syaifudin, M., Musadad, A., Fauzi, A.Z., Zahro, U.I., Nahidloh, S., & Pujiati, T. (2025). The Role of Al-Miftah lil Ulum Method in Enhancing Santri's Contextual Kitab Kuning Literacy Competence. *Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme.* 7(3); 106-123. https://doi.org/10.37680/scaffolding.v7i3.7865

Ulum, M., & Nuriyah, K. (2023). Implementasi Kitab Nubdzatul Bayan Dalam Pembelajaran Nahwu dan Shorrof Bagi Pemula. *Jurnal Educatio Fkip Unma*,

9(2), 1126–1132. https://doi.org/https://doi.org/10.31949/educatio.v9i2.5215

Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading?: A meta-regression analysis. *Studies in Second Language Acquisition*, *42*(2), 411–438. https://doi.org/https://doi.org/10.1017/s0272263119000688

Zainuddin, A., & Hikmah, H. (2026). From Sorogan to Makna Gandul: Arabic Reading Pedagogy Through Kitab Kuning in Indonesian Pesantren. *Senarai: Journal of Islamic Heritage and Civilization,* 2(4), 569-604. https://doi.org/10.66236/senarai.2026.2.4.569-604

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.