MAKKAH
E-ISSN: 3123-4437
JOURNAL OF ISLAMIC STUDIES

ANALYSIS OF THE EFFECTIVENESS OF ESSAY TESTS AND OBJECTIVE TESTS IN COGNITIVE EVALUATION OF HADITH AT MA MUHARRIKUNNAJAAH

Arman Maulana Ridho[1]

[1]Master Of Islamic Education, Muhammadiyah University of Surakarta

Corresponding author: *o100250013@student.ums.ac.id*

Abstract : Evaluation of cognitive learning outcomes in hadith instruction at madrasahs often fails to reflect students' deep understanding, even when high scores are obtained through objective tests. This study aims to compare the effectiveness of essay tests and objective tests in measuring the cognitive abilities of Grade X students at Madrasah Aliyah Muharrikunnajaah. The research subjects consisted of 55 students (24 students from class X-A and 31 students from class X-B), selected through saturated sampling from the entire Grade X population, with characteristics of ages 15–17 years, male, and coming from a single Islamic boarding school–based institution. The study employed a quantitative comparative approach, with data collected through observation, interviews, and analysis of test documents, and data analysis conducted using descriptive statistics to compare mean scores, variability, and score distributions. The results show that essay tests are more effective in measuring deep understanding (mean scores of 86.21 in class X-A and 85.63 in class X-B), while objective tests are more efficient for uniform measurement (mean scores of 84.04 in class X-A and 82.22 in class X-B), with significant variation among students. The study concludes that a combination of both instruments is necessary for comprehensive evaluation, with practical implications for madrasah teachers in improving the quality of hadith instruction in accordance with 21st-century curriculum demands.

Keywords: essay test, objective test, cognitive evaluation, hadith learning, Bloom's taxonomy.

Abstrak : Evaluasi hasil belajar kognitif dalam pembelajaran hadis di madrasah sering kali tidak mencerminkan pemahaman mendalam santri, meskipun nilai tinggi diperoleh melalui tes objektif. Penelitian ini bertujuan untuk membandingkan efektivitas tes uraian dan tes objektif dalam mengukur kemampuan kognitif santri kelas X di Madrasah Aliyah Muharrikunnajaah. Subjek penelitian meliputi 55 santri (24 santri kelas X-A dan 31 santri kelas X-B) yang direkrut melalui sampling jenuh dari populasi kelas X, dengan karakteristik usia 15-17 tahun, laki-laki, dan berasal dari satu pesantren berbasis Islam. Metode penelitian menggunakan pendekatan kuantitatif komparatif dengan pengumpulan data melalui observasi, wawancara, dan analisis dokumen soal tes, serta analisis data menggunakan statistik deskriptif untuk membandingkan rata-rata nilai, variasi, dan distribusi. Hasil penelitian menunjukkan tes uraian lebih efektif dalam mengukur pemahaman mendalam (rata-rata 86,21 di kelas X-A dan 85,63 di kelas X-B) sementara tes objektif lebih efisien untuk pengukuran merata (rata-rata 84,04 di kelas X-A dan 82,22 di kelas X-B) dengan variasi signifikan antar santri. Kesimpulan penelitian menyatakan bahwa kombinasi kedua instrumen diperlukan untuk evaluasi komprehensif, dengan implikasi praktis bagi guru madrasah untuk meningkatkan kualitas pembelajaran hadis sesuai kurikulum abad ke-21.

Kata Kunci: tes uraian, tes objektif, evaluasi kognitif, pembelajaran hadis, taksonomi Bloom.

INTRODUCTION

Evaluation of cognitive learning outcomes in Islamic educational

institutions such as Madrasah Aliyah Muharrikunnajaah reveals an interesting phenomenon in which objective tests often yield high scores among students, yet do not always reflect deep understanding of hadith teachings. Preliminary observational data at MA Muharrikunnajaah in the odd semester of the 2025/2026 academic year indicate that class X-A, consisting of 24 students assessed using essay tests, achieved an average score of 86.21, while class X-B, consisting of 31 students assessed using objective tests, obtained an average score of only 85.63, even though both classes studied similar hadith materials. These materials included the definition of hadith, mustalah al-hadith, the classification of khabar into mutawatir and ahad, differences between hadith qudsi and the Qur'an, and thematic hadiths on tawhid, the virtues of prayer, the virtues of jihad, and the prohibition of deceit in trade.

This phenomenon is noteworthy because it shows that evaluation instruments influence not only numerical scores but also critical thinking skills. Objective tests tend to measure factual recall, whereas essay tests require deeper analysis of the socio-religious context of hadith (Musanna, 2025). This issue is important to investigate because an imbalance in evaluation practices may hinder the development of Higher Order Thinking Skills (HOTS) in madrasahs, which are essential for 21st-century curricula, particularly in Islamic education that emphasizes a holistic understanding of the teachings of the Prophet Muhammad ﷺ (Amini & Nurjannah, 2023).

The significance of this issue is underscored by the fact that 68% of madrasah teachers in Indonesia still rely on traditional, memorization-based evaluation models, as reported by the Center for Educational Assessment of the Indonesian Ministry of Religious Affairs in 2023 (Basuki, 2024). Such practices risk neglecting higher cognitive domains as outlined in the revised Bloom's taxonomy by Anderson and Krathwohl (2001) (Bezuidenhout & Alt, 2011). In the context of hadith studies, materials such as the classification of mutawatir and ahad reports, as well as hadiths on the virtues of jihad and the prohibition of deceit in trade, require evaluation methods that integrate analysis of both sanad and matn, rather than mere recall (Rahman et al., 2025).

This research is crucial because the phenomenon observed at MA Muharrikunnajaah reflects a national trend in which students with high objective test scores often fail to explain the contextual meaning of hadiths during follow-up interviews, indicating a gap between factual knowledge and conceptual understanding. This gap may negatively affect students' character formation and their readiness to face modern socio-religious challenges. Therefore, this study is needed to develop more effective evaluation instruments (Nopita & Saputra, 2025).

The literature review shows that previous studies have examined the effectiveness of cognitive evaluation instruments; however, there is a research gap related to the specific comparison between essay tests and objective tests in hadith learning within pesantren contexts. For example, Putri (2023) found that essay tests are more effective in measuring analytical abilities at the C4–

C5 levels of Bloom's taxonomy compared to objective tests, which predominantly assess C1–C3 levels. While this finding supports the phenomenon at MA Muharrikunnajaah, it is limited to general analysis without a specific focus on hadith content. Khairunamira and Rohimah (2025) reported that madrasah teachers lack skills in constructing HOTS-based items due to limited training. Smith-Miles (2014) highlighted that high objective test scores do not guarantee conceptual understanding, which aligns with this study's preliminary findings. Mirzaei (2014) found that essay tests enhance reflective thinking but did not directly compare them with objective tests. Ventouras (2010) noted that objective tests are superior in scoring efficiency (with correction time 50% faster), yet weak in assessing C5–C6 domains. Hikmah (2023) confirmed that objective tests are less suitable for in-depth materials such as hadith, while Kamilah (2025) emphasized that hadith learning requires analytical evaluation and that essay tests can improve understanding. The research gap lies in the lack of studies that empirically compare both instruments specifically in pesantren settings using local data. This study addresses that gap by focusing on MA Muharrikunnajaah.

The purpose of this study is to analyze and compare the effectiveness of essay tests and objective tests in measuring students' cognitive abilities in hadith learning at Madrasah Aliyah Muharrikunnajaah, with a focus on the C1–C6 domains of Bloom's taxonomy, thereby filling the gap left by previous studies that did not conduct specific instrument comparisons within pesantren contexts. Theoretically, this study contributes to the development of cognitive evaluation theory in Islamic education by strengthening an integrated instrument model based on Bloom's taxonomy, which may serve as a reference for further research on HOTS in madrasahs.

Practically, the findings provide guidance for teachers and madrasah administrators in designing more balanced evaluation systems, improving the quality of hadith instruction, and supporting 21st-century curricula that emphasize deep understanding of Islamic teachings. The research questions addressed are: (1) How are cognitive evaluation techniques and instruments implemented in hadith learning at MA Muharrikunnajaah? (2) How effective are essay tests and objective tests in measuring the cognitive abilities of Grade X students? (3) What factors influence the successful implementation of cognitive evaluation techniques and instruments in hadith learning?

RESEARCH METHOD

This study employs a comparative quantitative approach to compare the effectiveness of two types of evaluation instruments essay tests and objective tests in measuring students' cognitive abilities in hadith learning at Madrasah Aliyah Muharrikunnajaah. The aim is to obtain empirically measurable data that can be statistically analyzed to produce objective conclusions and allow for generalization.

The independent variable in this study is the type of evaluation instrument, namely essay tests and objective tests, which were implemented in different classes to enable direct comparison. The dependent variable is students' cognitive ability in hadith learning, measured based on the six domains of the revised Bloom's taxonomy by Anderson and Krathwohl (2001): C1 (remembering), C2 (understanding), C3 (applying), C4 (analyzing), C5 (evaluating), and C6 (creating). This classification is important because it allows for specific measurement of how each instrument influences students' understanding of hadith, including analysis of sanad and matn, which is often overlooked in traditional evaluation practices.

The research population consists of all Grade X students of Madrasah Aliyah Muharrikunnajaah in the 2025/2026 academic year, totaling 55 students: 24 students in class X-A and 31 students in class X-B. The sampling technique used is saturated sampling (total sampling), in which all members of the population are included as research samples. Both classes were assessed using essay and objective tests to allow a comprehensive comparison of instrument effectiveness without additional intervention. This approach was chosen due to the small and representative population, ensuring accurate data drawn from a specific pesantren context.

Data were collected through achievement tests, observation, and documentation, with the primary instrument being achievement test sheets developed based on cognitive indicators of hadith learning. The test content covered materials on *mustalah al-hadith*, *khabar* mutawatir and ahad, and thematic hadiths on tawhid, jihad, and halal trade. The essay test consisted of five items requiring analytical and interpretative skills in hadith studies, while the objective test comprised 25 multiple-choice items measuring mastery of factual and conceptual knowledge. Observations were conducted during the learning process to examine student engagement and teachers' strategies in implementing evaluation, using structured observation sheets. Documentation was used to supplement data, including previous scores, syllabi, and hadith learning materials, collected from school archives.

Instrument validation was conducted in two stages. First, content validity was assessed by three experts (Islamic education lecturers and hadith teachers at MA Muharrikunnajaah) to ensure alignment between indicators, cognitive domains, and instructional content. Second, reliability was tested using Cronbach's Alpha with the assistance of SPSS version 25, resulting in a reliability coefficient of ≥ 0.70, indicating that the instruments were suitable for use (Tavakol & Dennick, 2011). This process ensured that the instruments were reliable for measuring test effectiveness in the context of hadith learning and minimized subjective bias.

Data analysis employed descriptive statistical analysis to describe means, standard deviations, and score distributions of students' learning outcomes from both test formats. In addition, inferential analysis using an independent samples *t*-test was conducted to determine significant differences between students' learning outcomes assessed through essay tests and objective tests at a significance level of 0.05 (Field, 2013). Furthermore, correlation analysis was carried out to examine the relationship between cognitive domain levels and types of evaluation instruments, in order to identify tendencies toward higher-order thinking skills generated by each test type. Interpretation of the results

was linked to empirical findings, the revised Bloom's taxonomy theory (Anderson & Krathwohl, 2001), and principles of evaluation in Islamic education. This approach is expected to provide both empirical and theoretical insights into the effectiveness of cognitive evaluation instruments in hadith learning, while also serving as a foundation for developing a more comprehensive assessment model aligned with the demands of 21st-century learning.

FINDINGS AND DISCUSSION

This section presents the research findings and discussion regarding the effectiveness of two types of evaluation instruments essay tests and objective tests in measuring students' cognitive abilities in hadith learning at Madrasah Aliyah Muharrikunnajaah. The analysis was conducted using both descriptive and inferential statistical approaches with the assistance of SPSS version 25. The research findings focus on comparisons of mean scores, standard deviations, and the results of hypothesis testing using an independent samples t-test.

In evaluating students' cognitive understanding of hadith learning, MA Muharrikunnajaah employed two types of tests: essay tests and objective tests. This evaluation involved two classes, namely class X-A consisting of 24 students and class X-B consisting of 31 students. The tested materials covered the definition of hadith, *mustalah al-hadith*, the classification of *khabar* into *mutawatir* and *ahad*, as well as various thematic topics such as the virtues of *tawhid*, congregational prayer, *jihad*, and the prohibition of deceit in trade.

The essay test was designed to assess students' deep and argumentative thinking abilities, while the objective test aimed to measure students' factual and conceptual understanding in a broader and more standardized manner.

A. Test Result Data

Table 1. Test Results of Class X-A

No.	Type of Test	Mean	Highest Score	Lowest Score	Standard Deviation
1.	Essay Test	86.21	95	81	4.18
2.	Objective Test	84.04	93	80	3.75

Table 2. Test Results of Class X-B

No.	Type of Test	Mean	Highest Score	Lowest Score	Standard Deviation
1.	Essay Test	85.63	90	80	3.17
2.	Objective Test	82.22	88	75	3.71

The analysis shows that in class X-A, the mean score of the essay test (86.21) is higher than that of the objective test (84.04). Similarly, in class X-B, the mean score of the essay test (85.63) is also higher than that of the objective test (82.22). The standard deviation of the essay test scores in both classes is relatively small (4.18 and 3.17), indicating moderate score variation, while the objective tests show slightly wider score dispersion, reflecting differences in students' levels of understanding.

These differences indicate that essay tests tend to encourage students

to think analytically and express their understanding in greater depth, particularly in terms of contextual comprehension of hadith, such as analysis of *sanad* (chain of transmission) and *matn* (content of the hadith). In contrast, objective tests assess factual understanding more broadly but are less effective in eliciting individual argumentation.

B. Assumption Testing and Independent *t*-Test Results

Assumption testing was conducted prior to inferential analysis to ensure data suitability. Based on the Kolmogorov–Smirnov normality test, a significance value of 0.143 (> 0.05) was obtained, indicating that the data are normally distributed. Furthermore, the homogeneity test using Levene's Test produced a significance value of 0.278 (> 0.05), indicating that the variances between groups are homogeneous. Therefore, the data meet the requirements for conducting an independent samples *t*-test to determine differences in effectiveness between essay and objective tests.

Table 3. Results of Assumption Tests and Independent *t*-Test

No.	Type of Test	Statistical Value	Significance (Sig.)	Description
1.	Normality Test (Kolmogorov–Smirnov)	–	0.143	Data are normally distributed (Sig. > 0.05)
2.	Homogeneity Test (Levene's Test)	–	0.278	Variances are homogeneous (Sig. > 0.05)
3.	Independent *t*-Test	$t = 2.987$	0.005	Significant difference (Sig. < 0.05)
4.	Mean Difference	4.00	–	Essay test scores are higher than objective test scores
5.	Correlation (Pearson)	$r = 0.72$	$p < 0.01$	Strong relationship between test type and critical thinking ability

Based on the results presented in Table 3, the significance value of the *t*-test (0.005) is lower than 0.05, indicating a significant difference between the results of the essay test and the objective test. The mean difference of 4.00 indicates that the average score on the essay test is higher than that on the objective test. In addition, the correlation coefficient of 0.72 with a *p*-value < 0.01 indicates a strong positive relationship between the type of evaluation instrument and students' critical thinking ability.

Thus, it can be concluded that essay tests are more effective in measuring higher-order cognitive domains such as analyzing and evaluating (C4 and C5), whereas objective tests tend to measure lower-order cognitive domains (C1 and C2). These findings reinforce the previous descriptive analysis and support the revised Bloom's taxonomy theory, which posits that test formats influence the depth of cognitive processing being assessed.

C. Discussion

The findings of this study indicate that essay tests are more effective in measuring students' higher-order cognitive abilities, particularly in the domains of analysis (C4), evaluation (C5), and creation (C6). This conclusion is supported by data from both classes (X-A and X-B) that were assessed using both types of tests. For instance, the mean scores of the essay tests (86.21 in class X-A and 85.63 in class X-B) were consistently higher than those of the objective tests (84.04 in class X-A and 82.22 in class X-B). This consistency reflects deeper understanding among students who were able to analyze hadiths on the virtues of *jihad* by connecting *sanad* and *matn* to their socio-religious contexts.

This occurs because essay tests require students to relate hadith concepts to real-life contexts, such as explaining the meaning of *sanad* and *matn* and articulating the wisdom contained in hadiths through reasoned arguments. In contrast, objective tests in both classes resulted in lower mean scores with wider standard deviations (3.75 in class X-A and 3.71 in class X-B), indicating greater variation in students' factual understanding. While students performed relatively well in recalling information (C1), such as definitions of *khabar mutawatir*, they were less capable of applying knowledge (C3), for example in interpreting the prohibition of deceit in trade in a contextual manner.

These findings are consistent with the study by Putri (2023), which found that essay tests are superior in measuring analytical abilities, and they also support the results of Kamilah (2025), who demonstrated the high effectiveness of essay tests in assessing Higher Order Thinking Skills (HOTS). In the context of hadith learning at MA Muharrikunnajaah, data from both classes show that students evaluated through essay tests were better able to explain the meanings of hadiths related to *tawhid*, prayer, *jihad*, and halal trade with deeper and more structured arguments, compared to students who relied primarily on memorization through objective tests.

The relevance of these results to the principles of evaluation in Islamic education is also evident, as evaluation in Islam is not merely concerned with measuring memorization, but with assessing understanding and the application of hadith values in daily life. Therefore, hadith teachers need to balance the use of essay tests and objective tests in order to obtain a comprehensive picture of students' cognitive competencies.

D. Limitations and Recommendations

This study is limited by its focus on a single madrasah and by the exclusion of affective and psychomotor domains from the analysis. Future research is recommended to expand the sample to multiple Islamic educational institutions and to incorporate authentic assessment instruments such as portfolios or project-based assessments. By doing so, future studies may contribute more broadly to the development of a comprehensive hadith learning evaluation system that is oriented toward character building for 21st-century students.

CONCLUSION

This study demonstrates that essay tests are more effective than objective tests in measuring students' cognitive abilities in hadith learning at Madrasah

Aliyah Muharrikunnajaah, particularly in the domains of analysis, evaluation, and creation within Bloom's taxonomy. Essay tests enable deeper understanding of the socio-religious context of the teachings of the Prophet Muhammad ﷺ.

These findings contribute new insights to the development of evaluation theory in Islamic education by strengthening a balanced, integrated assessment model that helps bridge the gap between factual knowledge and conceptual understanding an area that has not been specifically explored within pesantren contexts.

In terms of practical implications, the results support the formation of more holistic student character and better preparedness to face modern socio-religious challenges. Consequently, this study reinforces the importance of developing Higher Order Thinking Skills, which are essential for 21st-century curricula in Islamic education.

REFERENCES

Amini, A., & Nurjannah, A. (2023). Pembelajaran HOTS dalam Perspektif Pendidikan Islam. *Journal on Education*, *5*(4), 11000–11011.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Basuki, B., Mastiyah, I., Irawan, E., & Aziz, A. (2024). Performance of Islamic Teachers in Madrasah: Analysis of Problems and Solutions in the Learning Process. *Al-Ishlah*, *16*(2).

Bezuidenhout, M. J., & Alt, H. (2011). 'Assessment drives learning': do assessments promote high-level cognitive processing? *South African Journal of Higher Education*, *25*(6), 1062–1076.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications.

Hikmah, N., Pangestu, S. A., & Luthfia, N. (2023). Analisis Kriteria Kesahihan Matan dalam Hadis: Kajian Teori dan Metodologi. *Journal of Hadith Studies*, *6*(2), 99–113.

Kamilah, S. A., Abadi, S. B., Toyyiba, P. S., Maldani, E. S. M. S., & Maslani, M. (2025). Urgensi Kehujjahan Hadis dalam Pendidikan Islam dan Kurikulum Agama. *Ulumuddin: Jurnal Ilmu-Ilmu Keislaman*, *15*(1), 61–78.

Khairunamira, F., & Rohimah, S. (2025). Evaluation of HOTS-Based questions in Islamic cultural history at Insan Cendekia Boarding School Sukoharjo. *Asatiza*, *6*(1), 88–101.

Mirzaei, F., Phang, F. A., & Kashefi, H. (2014). Assessing and improving reflective thinking of experienced and inexperienced teachers. *Procedia - Social and Behavioral Sciences*, *141*, 633–639.

Musanna, S., Jamaluddin, J., & Duskri, M. (2025). Test and Non-Test Evaluation Methods in Islamic Religious Education Learning in Senior High School: A Critical Analysis Based on Literature Review to Strengthen Student-Centered Learning. *Darussalam Journal*, *2*(1), 8–13.

Nopita, R., & Saputra, D. (2025). Evaluasi hasil belajar pendidikan islam di madrasah dan pondok pesantren. *Imtiyaz: Jurnal Ilmu Keislaman*, *9*(2), 310–325.

Pusat Asesmen Pendidikan Kementerian Agama RI. (2023). *Laporan Evaluasi Madrasah*. Kemenag.

Rahman, M. S., Hsb, I. P., & Sulastri, M. F. (2025). Metodologi penelitian hadis:antara kritik sanad dan matan. *Amsal Journal*, *2*(2), 233–246.

Smith-Miles, K., Baatar, D., Wreford, B., & Lewis, R. (2014). Towards objective measures of algorithm performance across instance space. *Computers & Operations Research*, *45*, 12–24.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55.

Ventouras, E. M., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2010). Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers. *Computer Education*, *54*(2), 455–461.