Conceptualization of a HOTS-Based Tashbih Assessment Model: Instrument Gap Analysis in The Balāghah Textbook

Konseptualisasi Model Penilaian Tashbih Berbasis HOTS: Analisis Kesenjangan Instrumen dalam Buku Teks Balāghah

Muhamad Subhi[1*], Nirmala Luthfia Hamaz[1], Ahmad Noval[2], Zamakhsari[1]
[1] Universitas Islam Negeri Sunan Kalijaga, Yogyakarta, Indonesia
[2] Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia
[*] E-mail: subhi10011996@gmail.com

Abstract

Developing critical thinking skills in Arabic language learning is currently a curriculum requirement, but the quality of the assessment instruments in textbooks is often suboptimal in stimulating these skills. This study aims to identify gaps (needs analysis) in the assessment instruments for Tashbih (التشبيه) material in the Balāghah Grade XII Madrasah Aliyah Religious Specialization textbook as a basis for formulating an innovative assessment model based on Higher-Order Thinking Skills (HOTS). The method used was Qualitative Content Analysis of 28 Tashbih questions in the textbook, which was carried out by classifying the questions according to Bloom's Revised Taxonomy (C1-C6) and the measurement's focus. The results revealed a significant gap: the majority of assessment instruments (85%) were dominated by low cognitive levels (LOTS/C1-C2) and intermediate levels (MOTS/C3), focusing on memorization and simple identification of the pillars of Tashbih. Conversely, the measurement of the ability to analyze rhetorical functions, evaluate context, and create (HOTS: C4, C6) only reached 15%. Based on these findings, this study recommends the conceptualization of a Tashbih Rhetorical Analysis Project Assessment Model. This prototype model focuses on HOTS-based instruments that require students to compare, evaluate, and create Tashbih in authentic texts (Al-Qur'an/Hadith), supported by an assessment rubric that measures the depth of meaning analysis (Wajhu Shabh). This conceptualization is expected to serve as a guide for assessment innovation to improve students' critical thinking skills in understanding Balāghah.

Keywords: Assessment innovation, content analysis, higher-order thinking skills, HOTS, tashbih.

Abstrak

Pengembangan kemampuan berpikir kritis dalam pembelajaran bahasa Arab saat ini menjadi tuntutan kurikulum, namun kualitas instrumen penilaian dalam buku teks sering kali belum optimal dalam menstimulasi keterampilan tersebut. Penelitian ini bertujuan mengidentifikasi kesenjangan *(needs analysis)* dalam instrumen penilaian materi Tashbih (التشبيه) pada buku teks Balāghah Kelas XII Madrasah Aliyah Peminatan Keagamaan sebagai landasan perumusan model penilaian inovatif berbasis keterampilan berpikir tingkat tinggi *(Higher-Order Thinking Skills)* HOTS. Metode yang digunakan adalah Analisis Isi Kualitatif terhadap 28 butir soal Tashbih di dalam buku ajar tersebut, yang dilakukan dengan mengklasifikasikan soal berdasarkan Taksonomi Bloom Revisi (C1-C6) dan fokus pengukurannya. Hasil menunjukkan adanya kesenjangan signifikan, pada mayoritas instrumen penilaian (85%) didominasi oleh level kognitif rendah (LOTS/C1-C2) hingga menengah (MOTS/C3) yang berfokus pada hafalan dan identifikasi rukun Tashbih secara sederhana. Sebaliknya, pengukuran terhadap kemampuan analisis fungsi retoris, evaluasi konteks, dan penciptaan (HOTS: C4, C6) hanya mencapai 15%. Berdasarkan temuan kesenjangan ini, penelitian ini merekomendasikan Konseptualisasi Model Penilaian Proyek Analisis Retoris Tashbih. Prototipe model ini berfokus pada instrumen berbasis HOTS yang menuntut siswa untuk membandingkan, mengevaluasi, dan menciptakan Tashbih dalam teks otentik (Al-Qur'an/Hadis), serta didukung oleh rubrik penilaian yang mengukur kedalaman analisis makna (Wajhu Shabh). Konseptualisasi diharapkan dapat menjadi panduan inovasi penilaian guna meningkatkan kemampuan berpikir kritis siswa dalam memahami ilmu Balāghah.

Kata Kunci: Inovasi penilaian, analisis isi, keterampilan berpikir tingkat tinggi, HOTS, tashbih.

1. Introduction

Education at Madrasah Aliyah (MA) with a Religious Specialization emphasizes deepening Islamic knowledge, with mastery of Arabic as the primary focus. One of the essential branches of Arabic is the science of Balāghah (البلاغة) (Mustofa & Hasan, 2023). Balāghah is the science of Arabic rhetoric, teaching the principles of composing eloquent (fasih) and effective (baligh) speech (Jurjani, 2001), namely speech that is appropriate to the situation, audience, and purpose (Ariza, Prasetyo, et al., 2025). Mastery of Balāghah is fundamental for MA Religious Studies students because it is a vital tool for understanding the depth of meaning (i'jaz) and rhetorical beauty of the Qur'an and Hadith, which are the main sources of Islamic teachings (Suryani, 2019).

The material on Tashbih (التشبيه, simile) is the core of the science of Bayan and a basic rhetorical element in Balāghah (Ardyanto et al., 2025). Tashbih functions not only to compare two things but also as a rhetorical technique to clarify, beautify, and create a deep emotional impact (Abdul-Raof, 2006; Ali, 2023). Therefore, students' ability in Tashbih (mushabbah, mushabbah bih, etc.) must reach a level of in-depth analysis of the metaphor's rhetorical function or secret (sirr al-balāghah) (Aritonang et al., 2024).

Although the essence of Balāghah lies in the analysis and appreciation of deep meaning, classroom assessment practices often do not reflect this goal (Razak et al., 2023). The assessment of Tashbih material tends to focus on memorizing definitions (LOTS - Low-Order Thinking Skills) or on textual identification of pillars (Raji et al., 2024). The dominant assessment instruments only measure mastery of theory, without the ability to apply it to critical analysis, evaluation, and rhetorical creation (Monteiro et al., 2025). This condition creates a cognitive gap in LOTS/MOTS (C1, C2, C3) that focuses on memorizing rules and simple identification. C3 (Middle-Order Thinking Skills/Applying) only requires the use of rules in a structured context, which is still insufficient for analyzing the rhetorical function of Tashbih. HOTS (C4, C5, C6) are minimal or non-existent in instruments that require the ability to analyze rhetorical functions (sirr al-balāghah), evaluate the effectiveness of context, or create effective Tashbih.

This gap emphasizes the urgency of developing assessment instruments that require higher-order thinking skills (HOTS), such as analyzing (C4), evaluating (C5), and creating (C6) (Aisah & Pahlevi, 2020). The textbook "Arabic Language (Balāghah) Grade XII MA Religious Specialization," published by the Ministry of Religious Affairs, must be analyzed to determine the extent to which its assessment instruments align with these HOTS objectives. This research offers a contemporary contribution by bridging the gap between classical Balāghah concepts and modern educational assessment standards (HOTS). In the long term, this study provides a framework for future curriculum developers to ensure that Balāghah is taught as an applicable and contextual instrument of appreciation, rather than merely a classification of theoretical and normative rules (Aflisia et al., 2022; Umam et al., 2025).

Based on these issues, this study aims to: (1) analyze the cognitive level and measurement focus in the Tashbih assessment instrument in the Grade XII MA Balāghah textbook; and (2) formulate a model or prototype of HOTS-based Tashbih assessment as a recommendation to overcome this gap so that the assessment is better able to measure the ability to analyze rhetorical functions optimally.

2. Literature Review

Previous studies have consistently highlighted systemic problems in Arabic language learning and Islamic Religious Education in Indonesia. Research by Verawati et al. (2022) found

that the tasks in the Ministry of Religious Affairs' Arabic textbooks were dominated by Low-Order Thinking Skills (LOTS), thus failing to foster students' analytical abilities. This is reinforced by Haniefa (2022), who argues that the implementation of HOTS-based assessment in Arabic language learning is still not optimal, as teachers tend to focus on questions that only require memory and comprehension.

In the broader context of Islamic education, Nurhaliza and Mustopa (2023) identify that school exam questions for Islamic Religious Education (PAI) subjects also operate at a low cognitive level. Specifically, the need for Higher-Order Thinking Skills (HOTS) is essential in learning 'Ilmu Balāghah (Arabic Rhetoric). Unlike structural Arabic subjects (Nahwu and Sharf), the main objective of Balāghah is to appreciate and analyze the rhetorical beauty of the Qur'an and Hadith, which are inherently at cognitive levels C4 (Analyzing), C5 (Evaluating), and C6 (Creating).

In this context, research shows that students' rhetorical competence correlates significantly with higher-order cognitive abilities. For example, a study by Umam et al. (2025) explicitly highlights that the fundamental objective of Balāghah instruction (namely, achieving an understanding of sirr al-balāghah or i'jaz) can only be achieved through assessment instruments that require students to apply the rules of Tashbih to new contexts and evaluate their effectiveness, rather than simply identifying their components.

Furthermore, in the field of educational evaluation, research by Vlachopoulos & Makri (2024) confirms that the main challenge in developing HOTS instruments for language subjects is creating authentic, contextually relevant stimuli. This applies directly to Tashbih material, where instruments must compel students to analyze similes in the context of specific verses or Hadith and explain the rhetorical reasoning behind them a demand that far exceeds Low-Order Thinking Skills (LOTS).

Although previous studies, such as those conducted by Verawati (2022) and Haniefa (2022), have mapped the levels of students' thinking skills in Arabic language learning in general, there are fundamental differences with this study. Previous studies tended to focus on broadly measuring students' cognitive abilities, whereas this study focuses specifically on the critical analysis of the quality of assessment instruments for Tashbih material in textbooks. Furthermore, this study does not stop at identifying cognitive levels (LOTS/HOTS), but goes further by formulating an innovative assessment model that integrates rhetorical depth (sirr al-balāghah) with high cognitive evaluation indicators. This study fills this gap by providing an evaluation model that is relevant to the distinctive characteristics of Balāghah, which previous studies have not explored in depth.

3. Research Method

This study uses a qualitative content analysis to examine the structure, measurement focus, and cognitive level of the Tashbīh assessment instrument in the Arabic Language (Balāghah) textbook for Grade XII of the Madrasah Aliyah (MA) Religious Specialization, published by the Ministry of Religious Affairs (Kania & Kusumah, 2025). This approach was chosen because it allows for a systematic examination of the characteristics of assessment items, both in terms of the representation of Balāghah concepts and their relationship to the Higher Order Thinking Skills (HOTS) assessment framework (Zakaria & Abdelrahiem, 2025).

The research objects consisted of all questions, assignments, and learning activities that specifically focused on the Tashbīh material (Jannah et al., 2023). These units were analyzed as documents, so this study did not involve human respondents. The analysis focused on the

cognitive demands posed by the assessment instruments rather than on students' learning performance (Ekalia et al., 2025).

The research instrument used was a document analysis sheet based on the Revised Bloom Taxonomy (C1–C6) (Pangestuti et al., 2024; Sausan et al., 2023). This instrument served as a non-test instrument in the content analysis approach and was used to classify each unit of analysis into LOTS (C1–C2), MOTS (C3–C4), and HOTS (C5–C6) categories (Tuela & Palar, 2022). Classification is based on operational indicators derived directly from the conceptual and operational definitions of Bloom's Revised Taxonomy (Listiani & Rachmawati, 2022).

The instrument's validity is established through content validity (Yusoff, 2019), as it is not psychometric and is not intended to measure students' latent constructs. Content validity is ensured by the consistency between cognitive-level classification indicators, the unit of analysis, and the research objectives (Nabil et al., 2022; Syafitri et al., 2023). Therefore, statistical validity tests such as item-total correlation or Cronbach's Alpha are not applied in this study.

The validity of the content is reflected in the instrument's ability to classify all Tashbīh items into cognitive categories C1–C6 clearly and unambiguously (Puspitasari & Febrinita, 2021). All 28 items analyzed can be consistently mapped to the relevant cognitive level, so the instrument is considered representative in describing the cognitive level construct being studied (Nabil et al., 2022).

The instrument's reliability is maintained through the formulation of detailed, operational classification indicators at each cognitive level (Bolognesi et al., 2017). The clarity of these indicators allows the coding process to be carried out consistently across all units of analysis. This consistency is demonstrated by the absence of overlapping or unclassified items, thereby ensuring adequate internal reliability in the context of document analysis (Krippendorff, 2004).

Data analysis was conducted through three main stages. The first stage was the classification of each item's cognitive level using Bloom's Revised Taxonomy to map LOTS, MOTS, and HOTS tendencies (Rochmat et al., 2025). The second stage is a focus measurement analysis to distinguish between structural assessments (Tashbīh), such as component and type recognition, and functional-rhetorical assessments (sirr al-balāghah) (Istianah et al., 2024). The third stage involves calculating frequencies and percentages to obtain the proportional distribution for each category (Khan et al., 2025).

Conclusions were drawn descriptively and interpretively, by interpreting data distribution patterns based on dominant trends and category limitations that emerged, without making statistical generalizations (Krippendorff, 2004; Miles & Huberman, 1994). In addition, this study was supplemented with a conceptual analysis to formulate a prototype of the Tashbīh assessment model based on HOTS, by synthesizing content analysis results with classical Balāghah principles and theoretical HOTS studies (Creswell, 2013). The entire research process was carried out over one month and resulted in a theoretical assessment model design without empirical testing.

4. Result and Discussion

4.1. Research Result

4.1.1. Analysis of Tashbih Assessment Instrument Requirements

Content analysis of the Tashbīh (Simile) assessment instrument in the Madrasah Aliyah (MA) Grade XII textbook reveals a systemic gap between the objectives of the higher-level curriculum and the actual measurement tools provided. The findings are categorized into cognitive distribution and measurement focus as follows:

A. Cognitive Dimension: Dominance (LOTS/MOTS)

The distribution of cognitive levels in the assessment items shows a heavy skew toward basic cognitive processes. Of the total items analyzed, 85% fall into the LOTS/MOTS category, which is dominated by the levels of Remembering (C1), Understanding (C2), and Applying (C3). These items typically ask students to define terms such as Tashbīh, Mushabbah, and Wajh Shabh, or to identify these components in short, isolated sentences without literary context.

Table.1. Analysis of Tashbih Assessment Instrument Items Based on Cognitive Level (Revised Bloom's Taxonomy)

No.	Cognitive Level Category	Number of Items	Percentage
1.	C1 (Remembering)	8	28.57%
2.	C2 (Understanding)	9	32.14%
3.	C3 (Applying)	6	21.43%
4.	C4 (Analyzing)	4	14.29%
5.	C5 (Evaluating)	1	3.57%
6.	C6 (Creating)	0	0%
	Total Item Analyzed	**28**	**100%**

The analysis of 28 questions focusing on Tashbīh material shows that most questions are at a low to medium cognitive level. There are 17 questions (60.71% in categories C1 (remembering) and C2 (understanding), indicating that the assessment is dominated by mastery of basic concepts and a definitional understanding of Tashbīh. Meanwhile, 10 questions (35.72%) were at levels C3 and C4, indicating an effort to direct students to apply and analyze the elements of Tashbīh in the text. As for HOTS questions, they remain very limited, with only one in the C5 category and none in the C6 category. This distribution shows that the assessment of Tashbīh in textbooks emphasizes structural aspects and conceptual understanding rather than the development of evaluative and creative abilities.

B. Measurement Focus: Structural vs. Functional Orientation

Further details of the instrument's focus reveal a strong emphasis on structural classification. Approximately 92% of the questions are designed to test the identification of Rukun and Types (Tashbīh Mursāl, Mu'akkad, Mujmal). In contrast, only 8% of the items discuss Rhetorical Functions or Sirr al-Balāghah (the secret/wisdom underlying simile).

Table.2. Distribution of Measurement Focus

No.	Assessment Focus	Number of Items	Percentage
1.	Komponen dan Jenis Tashbih (Fokus Struktural)	26	92%
2.	Fungsi Retoris (Sirr al-Balāghah)	2	8%
	Total Items Analyzed	**28**	**100%**

This focus on technicalities (92%) shows that the assessment treats Balāghah as a set of rigid labels. Students are trained to label similes rather than analyze linguistic beauty. The near absence of questions on Sirr al-Balāghah (only 8%) means that the spirit of Balāghah the reason why certain comparisons are chosen for certain effects is largely ignored in the assessment process.

4.1.2. Development of HOTS-Based Comparative Prototypes

To bridge the identified gap, this study developed a prototype instrument based on the Comparative Analysis Model. Unlike traditional questions that ask for a single label, this model presents students with two competing linguistic expressions to trigger higher-order cognitive processing (C4, C5, and C6). Prototype Design Specifications:

1. Stimulus: The instrument provides two contrasting quotations that carry the same core meaning but differ in rhetorical intensity (e.g., literal simile vs. dense/mature simile).
2. Cognitive Task: Analysis (C4): Students are asked to deconstruct the structural differences between expressions. Evaluation (C5): Students provide reasoned justifications regarding which expression is more effective in a particular context, such as in praise (madḥ) or criticism (dhamm). Creation (C6): Students are challenged to construct new Tashbīh expressions that are rhetorically equivalent based on different contextual situations.

Example of Prototype Development:

1. Analysis & Evaluation Stage (C4 & C5): "Analyze and compare the following two expressions: (1) عَلِيٌّ كَالْأَسَدِ فِيْ الشُّجَاعَةِ (Ali is like a lion in courage Tashbīh Mufaṣṣal) and (2) عَلِيٌّ أَسَدٌ (Ali is a lion Tashbīh Balīgh). In the context of extraordinary praise (madḥ), justify which expression has greater rhetorical power and explain the Sirr al-Balāghah behind the removal of adāt (tools) and wajh shabh (aspects of comparison)!".
2. Creation Stage (C6): "Based on your understanding of the power of Tashbīh Balīgh in the example above, create an original Tashbīh expression in Arabic to describe 'extensive knowledge' using the Tashbīh Balīgh model to give a stronger aesthetic impact on the listener!".

Table.3. Comparison of Lots Questions vs HOTS Prototypes

No.	Comparison Dimensions	Old Model Questions (Ministry of Religious Affairs Arabic Textbook)	HOTS Prototype Model
1.	Cognitive Level	LOTS/MOTS (C1-C3)	HOTS (C4-C6)
2.	Measurement Focus	Structural (Labeling the pillars of Tashbīh)	Functional-Rhetorical (Sirr al-Balāghah)
3.	Example of Question Instructions	"State the pillars of Tashbīh in the following sentence!"	"Compare the following two verses. Why is the use of Tashbīh Balīgh more appropriate in this context than Tashbih Mursil?"
4.	Text Source	Simple man-made sentence	Authentic text (the Qur'an, Hadith, or classical poetry).
5.	Student Output	Short answers/terminology memorization	Critical argumentation, aesthetic evaluation, and original text production.

This prototype shifts students' roles from passive identifiers to literary critics and language creators. By demanding justification and creativity, this instrument forces students to connect linguistic forms with rhetorical intent, thereby comprehensively measuring their actual competence in Balāghah.

4.2. Discussion

4.2.1. Implications: Reduction of Balāghah to Classification

The dominance of LOTS-based instruments (85%) implies a significant shift in the essence of Balāghah learning (Nuha et al., 2023). Balāghah is essentially the science of appreciating beauty (jamāl) and effectiveness (balāghah), not merely the study of structural linguistics (Mahmudi & Abu Bakar, 2025). When 92% of assessments focus on classification, Balāghah is reduced to a "rhetorical taxonomy." This paradigm shift indicates that learning success is measured solely by students' ability to label linguistic phenomena rather than by their experience of them.

This creates a "pedagogical paradox" in which students are theoretically fluent (fāṣiḥ) but rhetorically incompetent (balīgh) (Ariza, Harahap, et al., 2025). They become adept at matching patterns, such as identifying Tashbīh Mujmal, without understanding the psychological or aesthetic reasons why a speaker, especially in the context of the miracles of the Qur'an, chooses that specific metaphor to influence the audience (Adawiah et al., 2023; Alsaied & Farag, 2025). By focusing on low-level cognitive tasks, the curriculum inadvertently hinders students from engaging in the critical analysis necessary to dissect complex Arabic literature (Kumar et al., 2023). As a result, Balāghah loses its essence as a tool for "meaning-making" and becomes merely an extension of rigid grammatical rules (Aflisia et al., 2022).

4.2.2. Failure in Achieving Sirr al-Balāghah and Creativity

Focusing solely on the pillars (mushabbah, mushabbah bih, adāt al-tashbīh, and wajhu shabah) tests only the "tools" of rhetoric (Lahiani et al., 2023). This leads to failure in achieving Sirr al-Balāghah (the secret of rhetoric). In this structuralist approach, students treat Tashbīh like mathematical equations rather than fluid artistic expressions. Although they master the terminology, they remain unable to evaluate the effectiveness of an expression or to create new expressions relevant to the context (Ramadhan, 2025; Yuningsih & Devi, 2024). This results in:

1. Minimum Evaluation Skills: Without HOTS stimuli, students are never challenged to justify why one poem or verse is more persuasive than another (C5) (Nasarudin et al., 2025).
2. Lack of Critical Criteria: Students lack criteria for assessing "linguistic beauty" beyond the mere presence of linguistic elements, rendering them passive consumers of texts rather than active critics (Ariza, Prasetyo, et al., 2025).
3. Stagnation in Creative Ability (C6): Because assessments never require students to construct original Tashbīh (C6), students fail to transform theoretical knowledge into productive competencies (Monteiro et al., 2025). They may memorize the definition of Tashbīh Balīgh, but are unable to use the technique of removing elements to create powerful and aesthetic messages in their own writing or speeches. The absence of level C6 in assessments keeps Balāghah a "dead science" rather than a living, creative communication tool (Raji et al., 2024).

4.2.3. Conceptualization of a Comparative and Productive Analysis Model

The failure of LOTS instruments calls for radical innovation in assessment (Azizah & Nugraha, 2025). The focus of assessment must shift from Structural Identification (what type is this?) to Functional-Evaluative Analysis (why is this effective?) and Creative Production (how to build it?). This shift is the only way to validate students' understanding of Sirr al-Balāghah and of contextual accuracy (muqtaḍā al-ḥāl) (Delima et al., 2025; Wahdi et al., 2025; Zuriyati et al., 2024). The Comparative Analysis Model integrated with construction tasks is proposed as a solution because it comprehensively triggers higher-order thinking:

1. Analysis & Evaluation (C4 & C5): By comparing two nearly identical expressions (Tashbīh Mursāl vs. Tashbīh Balīgh), students are forced to deconstruct structural differences and provide evidence-based assessments of why one expression is more suitable for a particular context (Yulistio, 2022).

2. Creating (C6): By requiring students to create a Tashbīh that is rhetorically equivalent to the analyzed model, this assessment ensures the internalization of aesthetic values. This creative process forces students to use Balāghah theory as a lens for original expression, rather than simply relying on memory banks (Yarun et al., 2023).

Operationally, the proposed prototype assessment model operates through three main integrated mechanisms:

1. Deconstruction-Evaluation Mechanism (C4-C5)
    In this model, students are not given single-choice questions but are presented with two authentic texts (e.g., two verses from the Qur'an with different Tashbīh patterns). Students are asked to analyze why the removal of adāt al-tashbīh in Tashbīh Balīgh gives a stronger impression of emphasis (mubālaghah) than Tashbīh Mursāl. Here, the instrument tests the sharpness of students' reasoning in uncovering the 'secrets' behind the structure of the Arabic language (Masrifah et al., 2025; Melati et al., 2024).

2. Contextual Transformation Mechanism (C6)
    This prototype requires students to produce language. Students are given a simple message (prose) and asked to transform it into a Tashbīh form with high balāghah value in accordance with a specific situation (muqtaḍā al-ḥāl) determined by the question. This ensures that knowledge does not remain in the mind but manifests itself in the application of skills (Ahmad & Ghafar, 2025; Hakim et al., 2024; Nurhamidah et al., 2025).

3. Holistic Assessment Rubric
    Unlike the rigid answer keys in LOTS instruments, this prototype is supported by a rubric that measures three criteria: (1) Structural accuracy of Tashbīh; (2) Depth of rhetorical argument; and (3) Originality of diction. Thus, assessment becomes a diagnostic process that measures the extent to which students have acquired "sense of language" (al-dzauq al-adabī) (Al-Nafjan et al., 2025; Ubaidillah et al., 2025).

For example, in the test questions in the textbook on page 20, question 7 only asks students to identify the Rukun-rukun Tashbīh (LOTS) (Hafidz, 2020). In contrast, the proposed comparative analysis model prototype frames the question as an evaluative challenge, presenting students with two sentences with similar meanings but different Tashbīh structures and asking them to justify which has a stronger emotional impact (at-ta'tsir al-nifsi). Through this mechanism, students not only master the 'tools' of rhetoric mechanistically but also delve into the depth of meaning, as emphasized by Al-Jurjani in Asrār al-Balāghah (Jurjani, 2001).

Ultimately, this model transforms assessment from a "memory test" to an "appreciation and creation exercise," ensuring that MA Religious Studies students achieve the depth of understanding necessary to engage with the aesthetic miracles of the Qur'an (Haqqar & Hamzah, 2026; Hasanah et al., 2022).

5. Conclusion

This study concludes that the assessment instruments for Tashbīh material in the Grade XII Balāghah textbook for Madrasah Aliyah (MA) Religious Specialization are still dominated by lower-order cognitive skills (LOTS), with 85% of the questions focusing on memorization and

structural identification (C1-C3). Only 15% of the instruments touch on HOTS, and even then, they are limited to simple analysis. This condition causes Balāghah learning to become stuck on mastering rhetorical labels, neglecting its main essence: the appreciation of the beauty of meaning (sirr al-balāghah) and the creation of effective expressions.

As a solution, this study offers a prototype of a HOTS-based assessment model that integrates Comparative Analysis and Creative Production. This model requires students not only to distinguish rhetorical structures (C4) and evaluate the contextual effectiveness of an expression (C5), but also to construct original Tashbīh with high rhetorical power (C6). By fully engaging cognitive levels C4 to C6, the assessment instrument can be transformed from a mere memorization test into a functional, creative tool for developing rhetorical competency.

The implications of this study underscore the need to thoroughly revise the evaluation instruments in Arabic textbooks to align them with modern educational standards and the essence of Balāghah as a discipline of taste and logic. The development of instruments that emphasize Sirr al-Balāghah and the ability to create texts will ensure that students of Madrasah Aliyah Religious Specialization have the depth of analysis necessary to understand and produce eloquent Arabic texts, especially in interacting with the beauty of the language of the Qur'an.

6. References

Abdul-Raof, H. (2006). *Arabic Rhetoric: A Pragmatic Analysis* (1. publ). Routledge.

Adawiah, D. R., Ambarawati, P., Marfuah, P., Hidayat, W., & Fauzi, A. (2023). Aspek-Aspek Epistemologi Dalam Manajemen Pendidikan Islam. *Benchmarking*, *7*(2), 84. https://doi.org/10.30821/benchmarking.v7i2.18369

Aflisia, N., Hendrianto, & Kasmantoni. (2022). Teaching Balaghah for the Purpose of Appreciation of Al-Quran Language. *Lughawiyyat: Jurnal Pendidikan Bahasa Dan Sastra Arab*, *4*(2), 156–172. https://doi.org/10.38073/lughawiyyat.v4i2.537

Ahmad, H., & Ghafar, N. A. (2025). Stylistic Variation and Linguistic Strategies in Quranic Discourse: A Rhetorical, Phonetic, and Translational Analysis. *International Journal of Research and Innovation in Social Science*, *IX*, 5957–5964. https://doi.org/10.47772/IJRISS.2025.90400426

Aisah, S., & Pahlevi, T. (2020). Pengembangan Instrumen Penilaian Higher Order Thinking Skills (HOTS) pada Mata Pelajaran Korespondensi Kelas X OTP di SMK Negeri 1 Jombang. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, *8*(1), 146–156. https://doi.org/10.26740/jpap.v8n1.p146-156

Ali, M. K. (2023). Tasybih Dalam Ilmu Balaghah: Balaghah. *Al Muallaqat*, *2*(2), 15–25. https://jurnal.stainmajene.ac.id/index.php/almuallaqat/article/view/576

Al-Nafjan, E., Alhawsawi, S., & Nasser Abu Dujayn, A. (2025). Exploring Self-Assessment in EFL Academic Writing: A Study of Undergraduate Saudi Students Utilizing a Rubric. *Language Testing in Asia*, *15*(1), 62. https://doi.org/10.1186/s40468-025-00397-2

Alsaied, M. A., & Farag, H. H. A. (2025). Analysis of the Alignment of Speech with the Requirements of Context in the Qur'an A Methodological and Foundational Study. *Journal of Posthumanism*, *5*(4). https://doi.org/10.63332/joph.v5i4.1078

Ardyanto, A., Rohanda, R., & Abdul Halim, M. (2025). Tasybih Dalam Kitab Al-Alqiro'ah Al-Rasyidah Karangan Abdul Fatah Shobary Dan Ali Umar Kajian Ilmu Bayan. *Jurnal Kajian Bahasa, Sastra Dan Pengajaran (KIBASP)*, *8*(3), 393–411. https://doi.org/10.31539/x6wbac62

Aritonang, P. Y., Nurul Aulia Ersa Putri, Said Fahrezi, & Harun Al Rasyid. (2024). Tasybīh Al-Tamsīl Dalam Al-Qur'an: Analisis BalagĀh Pada Surah Al-Kahfi Ayat 45. *Al Furqan: Jurnal Ilmu Al Quran Dan Tafsir*, *7*(1), 142–158. https://doi.org/10.58518/alfurqon.v7i1.2672

Ariza, F. N., Harahap, R. F., Perdana, T., Harahap, S. K., Hayat, Z., & Siahaan, U. P. (2025). Penerapan Balaghah dalam Pembelajaran Bahasa Arab di Pesantren. *MUDABBIR Journal Research and Education Studies*, *5*(2), 711–718. https://doi.org/10.56832/mudabbir.v5i2.1234

Ariza, F. N., Prasetyo, M. H., Dermawan, E., Lubis, M., Siagian, A. T. A., Ramadhan, M. F., & Ardiansyah, H. (2025). Ilmu Balaqhah Al-Quran, Pengertian Fashahatul Mufrad: Tanafuri Al-Huruf, Gharabah, Mukhalafatul Qiyas. *Hikmah: Jurnal Studi Pendidikan Agama Islam*, *2*(2), 342–351. https://doi.org/10.61132/hikmah.v2i2.1097

Azizah, F. N., & Nugraha, M. S. (2025). Kesesuaian Butir Soal Buku Al-Qur'an Hadis Kelas X Madrasah Aliyah Dengan Pendekatan HOTS Dan LOTS Dalam Kurikulum 2013.: The Alignment of Question Items in the Al-Qur'an Hadith Textbook for Grade X of Madrasah Aliyah with HOTS and LOTS Approaches in the 2013 Curriculum. *Epistemic: Jurnal Ilmiah Pendidikan*, *4*(3), 395–409. https://doi.org/10.70287/epistemic.v4i3.278

Bolognesi, M., Pilgram, R., & van den Heerik, R. (2017). Reliability in Content Analysis: The Case of Semantic Feature Norms Classification. *Behavior Research Methods*, *49*(6), 1984–2001. https://doi.org/10.3758/s13428-016-0838-6

Creswell, J. W. (with Internet Archive). (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Los Angeles : SAGE Publications. http://archive.org/details/qualitativeinqui0000cres_d2m7

Delima, D. A., Hikmah, H., & Masrun, M. (2025). Analisis Higher Order Thinking Skills (hots) Pada Soal Ujian Asesmen Madrasah Aliyah Tahun 2024. *Maharaat Lughawiyyat: Jurnal Pendidikan Bahasa Arab*, *4*(2), 118–125. https://doi.org/10.18860/jpba.v4i2.15872

Ekalia, Y. J., Jemadi, F., & Beda, R. (2025). Questions: A Bloom's Taxonomy Study of the Grade. *IJEAL (International Journal of English and Applied Linguistics)*, *5*(2). https://doi.org/doi.org/ijeal.v5i2.6701

Hafidz, M. (2020). *Bahasa Arab (Balaghah) Kelas XII MA Peminatan Keagamaan*. Direktorat KSKK Madrasah, Direktorat Jenderal Pendidikan Islam, Kementerian Agama RI.

Hakim, A. R., Huda, M., & Nasrullah. (2024). Rhetoric as a Tool for Controlling Conflict: A Discourse Analysis Perspective. *LiNGUA: Jurnal Ilmu Bahasa dan Sastra*, *19*(2), 256-267. https://doi.org/10.18860/ling.v19i2.30746

Haniefa, R. (2022). Implementasi Model Penilaian Hots (Higher Order Thinking Skills) Pada Penilaian Empat Keterampilan Berbahasa Arab. *Ta'limi | Journal of Arabic Education and Arabic Studies*, *1*(1), 49–71. https://doi.org/10.53038/tlmi.v1i1.11

Haqqar, M. F., & Hamzah. (2026). A New Direction in Arabic Morphosyntactic Studies: Integrating Classical Theory, Modern Linguistics, and Pedagogical Applications. *Multidisciplinary Indonesian Center Journal (MICJO)*, *3*(1), 1212–1220. https://doi.org/10.62567/micjo.v3i1.1972

Hasanah, M., Syarofah, A., & Sari, R. R. (2022). *Pragmatic Thinking in the Book of Dalail Al-I'jaz Abdul Qahir Al-Jurjani:* International Symposium on Religious Literature and Heritage (ISLAGE 2021). https://doi.org/10.2991/assehr.k.220206.007

Istianah, I., Arrohmah, N., Nursafitri, N., & Makruf, I. (2024). Item Analysis of Multiple Choice Questions of Nahwu Class XII Madrasah Aliyah Pondok Pesantren Islam Al Mukmin Ngruki Sukoharjo: Analisis Butir Soal Pilihan Ganda Mata Pelajaran Nahwu Kelas XII Madrasah Aliyah Pondok Pesantren Islam Al Mukmin Ngruki Sukoharjo. *Civilization Research: Journal of Islamic Studies*, *3*(1), 60–75. https://doi.org/10.61630/crjis.v3i1.44

Jannah, I. K., Mahanal, S., & Mashfufah, A. (2023). Analisis Tingkat Kognitif Soal Asesmen Sumatif Akhir Semester I (ASAS I) IPA Berbasis Jenis Soal AKM Berdasarkan Taksonomi Bloom Di Kelas V SD Swasta Kota Malang. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, *6*(2). https://doi.org/10.54371/jiip.v6i2.1633

Jurjani, 'Abd al-Qahir ibn 'Abd al-Rahman. (2001). *Asrar Al-Balaghah*. al-Maktabah al-'Asriyah.

Kania, N., & Kusumah, Y. S. (2025). The Measurement of Higher-Order Thinking Skills: A Systematic Literature Review. *Malaysian Journal of Learning and Instruction*, *22*(1), 97–116. https://doi.org/10.32890/mjli2025.22.1.6

Khan, H. F., Qayyum, S., Beenish, H., Khan, R. A., Iltaf, S., & Faysal, L. R. (2025). Determining the Alignment of Assessment Items with Curriculum Goals Through Document Analysis by Addressing Identified Item Flaws. *BMC Medical Education*, *25*(1), 200. https://doi.org/10.1186/s12909-025-06736-4

Krippendorff, K. (with Internet Archive). (2004). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, Calif. : Sage. http://archive.org/details/contentanalysisi0000krip_y7k5

Kumar, T., Soozandehfar, S. M. A., Hashemifardnia, A., & Mombeini, R. (2023). Self Vs. Peer Assessment Activities in EFL-Speaking Classes: Impacts on Students' Self-Regulated Learning, Critical Thinking, and Problem-Solving Skills. *Language Testing in Asia*, *13*(1), 36. https://doi.org/10.1186/s40468-023-00251-3

Lahiani, R., Alfalasi, T., Alzaabi, A., & Alhassani, N. (2023). Translations of Overtones in Classical Arabic Poetry: Optimality Theory as a Quality Assessment Framework. *Ampersand*, *11*, 100143. https://doi.org/10.1016/j.amper.2023.100143

Listiani, W., & Rachmawati, R. (2022). Transformasi Taksonomi Bloom Dalam Evaluasi Pembelajaran Berbasis HOTS. *Jurnal Jendela Pendidikan*, *2*(03), 397–402. https://doi.org/10.57008/jjp.v2i03.266

Mahmudi, M. 'Ainul F., & Abu Bakar, M. Y. (2025). Konstruksi Keilmuan Balaghoh: Sebuah Tinjauan Dari Perspektif Filsafat Ilmu. *Perspektif : Jurnal Pendidikan Dan Ilmu Bahasa*, *3*(1), 228–249. https://doi.org/10.59059/perspektif.v3i1.2116

Masrifah, K., Ahsanuddin, M., & Ainin, M. (2025). The Effectiveness of HOTS-Based Practice Questions in Enhancing Students' Critical Thinking in Arabic Learning. *ALSUNIYAT:*

Jurnal Penelitian Bahasa, Sastra, Dan Budaya Arab, *8*(2), 360–369. https://doi.org/10.17509/alsuniyat.v8i2.87847

Melati, A., Ziihah, R. A. N., & Setiyawan, A. (2024). Higher Order Thingking Skills (HOTS)-Based Assessment as an Innovation in Balaghoh Learning: Penilaian Berbasis HOTS Sebagai Inovasi Dalam Pembelajaran Balaghoh. *Edulab: Majalah Ilmiah Laboratorium Pendidikan*, *9*(1), 106–119. https://doi.org/10.14421/edulab.2024.91.07

Miles, & Huberman. (1994). *Qualitative Data Analysis* (Second). SAGE Publication: International Educational and Professional Publisher.

Monteiro, F., Nascimento, L. B., Leitão, J. A., Santos, E. J. R., Rodrigues, P., Santos, I. M., Simões, F., & Nascimento, C. S. (2025). Optimizing Working Memory Assessment: Development of Shortened Versions of Complex Spans, Updating, and Binding Tasks. *Psychological Research*, *89*(2), 65. https://doi.org/10.1007/s00426-025-02083-7

Mustofa, A., & Hasan, Moh. A. K. (2023). Peran Bahasa Arab Dalam Pendidikan Islam Di Ma'had Aly Ar-Rasyid Wonogiri: Tinjauan Terhadap Pembelajaran Bahasa Arab Sebagai Media Akses Ilmu Agama. *Tatsqifiy: Jurnal Pendidikan Bahasa Arab*, *4*(2), 88–94. https://doi.org/10.30997/tjpba.v4i2.8642

Nabil, N. R. A., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis Indeks Aiken Untuk Mengetahui Validitas Isi Instrumen Asesmen Kompetensi Minimum Berbasis Konteks Sains Kimia. *Paedagogia*, *25*(2), 184–191. https://doi.org/10.20961/paedagogia.v25i2.64566

Nasarudin, N., Munir, M. M., Munjiah, M., Alrasi, F., Yunus, M., Ma'arif, A. S., Ma'ali, A., Suroiyah, E. N., Habibi, R., Hamim, H., Charis, M. A., Wasil, M., & Kirom, M. (2025). *Bahasa Arab: Teori, Praktik, Dan Konteks Pembelajaran Modern*. Padang: Gita Lentera.

Nuha, M. A. U., Nisa, I. F., Anwar, S., & Musyafaah, N. (2023). SAVI Learning Model in Improving Students' Learning Interest and Thinking Ability in Balaghah Learning. *International Journal of Arabic Language Teaching*, *5*(01), 15. https://doi.org/10.32332/ijalt.v5i01.6491

Nurhaliza, A. S., & Mustopa, M. F. (2023). Analisis Tingkat Kognitif Soal Ujian Sekolah Pada Mata Pelajaran Pendidikan Agama Islam Berdasarkan Taksonomi Bloom Revisi 2001. *Journal of Research and Thought on Islamic Education (JRTIE)*, *6*(1), 71–84. https://doi.org/10.24260/jrtie.v6i1.2696

Nurhamidah, I., Santosa, R., Djatmika, D., & Yustanto, H. (2025). A Qur'anic Discourse Analysis: Exploring Values in Nouman Ali Khan's Amazed by the Qur'an. *3L: Southeast Asian Journal of English Language Studies*, *31*(2), 133. https://doi.org/10.17576/3L-2025-3102-10

Pangestuti, A. F. N., Ana, S., Syahra, E. F. A., & Sari, S. N. (2024). Analisis Distribusi Soal Higher Order Thinking Skills (HOTS) Dan Lower Order Thinking Skills (LOTS) Dalam Buku Ajar Bahasa Indonesia. *Indonesian Journal of Action Research*, *3*(1), 25–31. https://doi.org/10.14421/ijar.2024.31-03

Puspitasari, W. D., & Febrinita, F. (2021). Pengujian Validasi Isi (Content Validity) Angket Persepsi Mahasiswa Terhadap Pembelajaran Daring Matakuliah Matematika Komputasi.

Journal Focus Action of Research Mathematic (Factor M), *4*(1), 77–90. https://doi.org/10.30762/factor_m.v4i1.3254

Raji, S., Alikhani, M., De Melo, G., & Stone, M. (2024). A Corpus of Persian Literary Text. *Language Resources and Evaluation*, *58*(2), 409–425. https://doi.org/10.1007/s10579-023-09689-6

Ramadhan, R. W. (2025). Sejarah Dan Peran Tokoh Dalam Perkembangan Ilmu Balaghah. *Al-Iftah: Jurnal Tafsir Al-Qur'an Dan Hadist*, *1*(2), 147–155. https://e-jurnal.publikasiakademikgroup.com/index.php/AIJTAH/article/view/16

Razak, R. A., Mat Yusoff, S., Hai Leng, C., & Mohamadd Marzaini, A. F. (2023). Evaluating Teachers' Pedagogical Content Knowledge in Implementing Classroom-Based Assessment: A Case Study Among Esl Secondary School Teachers in Selangor, Malaysia. *Plos One*, *18*(12), e0293325. https://doi.org/10.1371/journal.pone.0293325

Rochmat, C. S., Masturoh, F., Subakir, F., & Sadiah, Z. H. (2025). Item Analysis of Insya' Assessment Based on Bloom's Taxonomy at Islamic Senior High School Darussalam Rajapolah Tasikmalaya. *An Nabighoh*, *27*(1), 201–222. https://doi.org/10.32332/an-nabighoh.v27i1.201-222

Sausan, K., Rohmatillah, R., & Rini, Y. P. (2023). HOTS Based on Revised Bloom's Taxonomy: Analysis in English Examination Test Items Used in High School Level. *English Education: Jurnal Tadris Bahasa Inggris*, *16*(1), 102–118. https://doi.org/10.24042/ee-jtbi.v16i1.16103

Suryani, K. (2019). Keunggulan Bahasa Al-Qur'an Di Bidang Sastra (al-Balaghah) Dalam Pandangan Ibn Asyur. *Dar El-Ilmi: Jurnal Studi Keagamaan, Pendidikan Dan Humaniora*, *6*(2), 220–245. https://doi.org/10.52166/dar

Syafitri, A., Anggraini, D. P., Parinduri, W. M., Rambe, T. R., & Rambe, N. (2023). Analisis Validitas Isi Pada Instrumen Penilaian Akhir Semester Mata Pelajaran Ipas Di Sd. *School Education Journal Pgsd Fip Unimed*, *13*(4), 344–349. https://doi.org/10.24114/sejpgsd.v13i4.54309

Tuela, A. I., & Palar, Y. N. (2022). Analysis of Higher Order Thinking Skills (HOTS) Based on Bloom Taxonomy in Comprehensive Examination Questions. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran, Dan Pembelajaran*, *8*(4), 957–971. https://doi.org/10.33394/jk.v8i4.5885

Ubaidillah, M., Haddad, A., Wadi, N., & Qomari, N. (2025). Digital Portfolio as an Authentic Assessment Instrument in Arabic Language Learning: A Literature Review. *Jurnal Penelitian Ilmu Pendidikan Indonesia*, *4*, 838–846. https://doi.org/10.31004/jpion.v4i2.486

Umam, R., Setiyawan, A., & Hassan, A. R. T. S. (2025). Balaghah Learning Innovation Based on Problem-Based Learning. *Jurnal Al-Maqayis*, *12*(1), 1–16. https://doi.org/10.18592/jams.v12i1.15251

Verawati, H., Febriani, E., Muflihah, I., Hasanah, U., Susanti, A., & Fitriani, F. (2022). HOTS Analysis of Task Instructions in Bahasa Arab Madrasah Aliyah Textbook Published by The Ministry of Religious Affairs. *Edukatif: Jurnal Ilmu Pendidikan*, *4*(1), 944–951. https://doi.org/10.31004/edukatif.v4i1.1930

Vlachopoulos, D., & Makri, A. (2024). A Systematic Literature Review on Authentic Assessment in Higher Education: Best Practices for the Development of 21st Century Skills, and Policy Considerations. *Studies in Educational Evaluation*, *83*, 101425. https://doi.org/10.1016/j.stueduc.2024.101425

Wahdi, S., Raswan, & Dardiri, A. (2025). Pandangan Kritis Amin Al-Khuli Terhadap Tradisi Balaghah Klasik. *Journal of Literature Review*, *1*(2), 401–409. https://doi.org/10.63822/xeqb9b38

Yarun, A., Hasan, M. Z. A., Syahri, S., & Markus, M. (2023). Higher Order Thinking Skill (HOTS) Students in Indonesia and Saudi Arabia (Comparative Study of Assessment Application in Schools). *Journal of Contemporary Islamic Education*, *3*(2), 96–109. https://doi.org/10.25217/jcie.v3i2.3787

Yulistio, D. (2022). Kemampuan Mahasiswa Menganalisis Tes (Soal) Bahasa Indonesia Sesuai Aspek Ranah Kognitif Model Anderson Dan Krathwohl. *Silampari Bisa: Jurnal Penelitian Pendidikan Bahasa Indonesia, Daerah, Dan Asing*, *5*(1), 1–21. https://doi.org/10.31540/silamparibisa.v5i1.1498

Yuningsih, T., & Devi, W. S. (2024). Dinamika Pembelajaran Retorika Dan Berpikir Kritis Pada Mahasiswa Pendidikan Bahasa Dan Sastra Indonesia Universitas Muhammadiyah Jakarta. *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, *10*(2), 152. https://doi.org/10.29210/020243706

Yusoff, M. S. B. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal*, *11*, 49–54. https://doi.org/10.21315/eimj2019.11.2.6

Zakaria, M., & Abdelrahiem, S. I. (2025). Content Analysis of the Objectives of the Rhetoric Curriculum in Islamic Universities in Aceh in Light of the Integrative Approach to Teaching Arabic Language: تحليل مضمون أهداف منهج البلاغة في الجامعات الإسلامية بأتشيه في ضوء المدخل التكاملي لتعليم اللغة العربية. *EL-MAQALAH: Journal of Arabic Language Teaching and Linguistics*, *6*(1), 46–80. https://doi.org/10.22373/maqalah.v6i1.7670

Zuriyati, Boeriswati, E., Rafli, Z., Nuruddin, Rasyid, Y., Kamal, M., Nugrahawan, A. R., Munibi, A. Z., Hamdani, H., & Mariani. (2024). Peningkatan Hasil Akreditasi Prodi Pendidikan Bahasa Arab Stai Nurul Iman Bogor Melalui Pendampingan Telaah Kurikulum. *Prosiding Seminar Nasional Pengabdian Kepada Masyarakat*, *5*(1), SNPPM2024P352-SNPPM2024P364.