

## Alignment of *Aqidah Akhlak* Items with Basic Competency Indicators: Mapping and Cognitive Analysis

\*Melinda Nur Ekawati<sup>1</sup>, Suwadi<sup>2</sup>

<sup>1,2</sup>Universitas Islam Negeri Sunan Kalijaga, Jl. Laksda Adisucipto, Sleman,  
Daerah Istimewa Yogyakarta, Indonesia

\*melindaekha01@gmail.com

**ABSTRACT:** *This study evaluates the alignment of Grade X Aqidah Akhlak assessment instruments with Basic Competencies in the Islamic Religious Education curriculum. It examines how far test items correspond to curriculum demands and maps their cognitive levels. This study used a quantitative approach with a descriptive-evaluative design. Item to competency mapping was applied to analyse the suitability of 180 end-of-year Aqidah Akhlak assessment items with Basic Competencies. Bloom's Taxonomy guided the identification of Lower Order Thinking Skills (LOTS) and Higher Order Thinking Skills (HOTS) proportions. The findings showed that the alignment reached a moderate level of 64.4%. The analysis revealed uneven representation of Basic Competencies, as some competencies were fully accommodated while others received limited or no measurement. Cognitively, the instruments were dominated by LOTS at 70%, whereas HOTS reached only 13%. This indicates that teachers' assessment practices still emphasise conceptual recall rather than analytical abilities required by the curriculum. Given its focus on Grade X end-of-year Aqidah Akhlak assessment items from four Madrasah Aliyah, this study's findings should be understood as context-specific rather than broadly generalizable. The study highlights the need to improve teachers' assessment literacy, especially in developing valid instruments aligned with Basic Competencies and curriculum-required cognitive levels. This study contributes by combining item to competency mapping with Bloom-based cognitive analysis, expanding assessment literature in Islamic Religious Education, particularly Aqidah Akhlak.*

Studi ini mengevaluasi keselarasan instrumen penilaian Aqidah Akhlak Kelas X dengan Kompetensi Dasar dalam kurikulum Pendidikan Agama Islam. Studi ini meneliti sejauh mana butir-butir soal sesuai dengan tuntutan kurikulum dan memetakan tingkat kognitifnya. Studi ini menggunakan pendekatan kuantitatif dengan desain deskriptif-evaluatif. Pemetaan butir soal ke kompetensi diterapkan untuk menganalisis kesesuaian 180 butir soal penilaian Aqidah Akhlak akhir tahun dengan Kompetensi Dasar. Taksonomi Bloom memandu identifikasi proporsi Keterampilan Berpikir Tingkat Rendah (LOTS) dan Keterampilan Berpikir Tingkat Tinggi (HOTS). Hasil penelitian menunjukkan bahwa keselarasan mencapai tingkat sedang sebesar 64,4%. Analisis mengungkapkan representasi Kompetensi Dasar yang tidak merata, karena beberapa kompetensi diakomodasi sepenuhnya sementara yang lain menerima pengukuran terbatas atau tidak sama sekali. Secara

---

<sup>1</sup> orcid id: <https://orcid.org/0009-0000-6197-8203>

<sup>2</sup> orcid id: <https://orcid.org/0000-0001-5800-3763>

kognitif, instrumen didominasi oleh LOTS sebesar 70%, sedangkan HOTS hanya mencapai 13%. Hal ini menunjukkan bahwa praktik penilaian guru masih menekankan pada ingatan konseptual daripada kemampuan analitis yang dibutuhkan oleh kurikulum. Mengingat fokusnya pada butir-butir penilaian Aqidah Akhlak akhir tahun kelas X dari empat Madrasah Aliyah, temuan penelitian ini harus dipahami sebagai spesifik konteks dan bukan dapat digeneralisasikan secara luas. Penelitian ini menyoroti perlunya peningkatan literasi penilaian guru, terutama dalam mengembangkan instrumen yang valid yang selaras dengan Kompetensi Dasar dan tingkat kognitif yang dibutuhkan kurikulum. Penelitian ini berkontribusi dengan menggabungkan pemetaan butir soal ke kompetensi dengan analisis kognitif berbasis Bloom, memperluas literatur penilaian dalam Pendidikan Agama Islam, khususnya Aqidah Akhlak.

**Keywords:** *Aqidah Akhlak Assessment, Item to Competency Mapping, Basic Competencies, Cognitive Analysis, Bloom's Taxonomy.*

*Received: January 03, 2026; Revised: April 11, 2026; Accepted: May 24, 2026*

## I. INTRODUCTION

Islamic Religious Education (PAI) has a central position in the Indonesian education system because it is not only oriented toward the mastery of religious knowledge but also toward character formation, value strengthening, and the development of students' religious behaviour. This orientation shows that religious education does not merely lead to cognitive achievement, but also to self-balance as a manifestation of *insan kamil* (Awali, 2020). In the broader discourse of Islamic Religious Education, the curriculum ideally integrates *ta'lim*, *ta'dib*, and *tarbiyah* so that learning outcomes, learning activities, and assessment practices are not reduced to cognitive mastery alone (Ansyari, 2023). PAI, as a subject that encompasses *aqidah*, worship, and morals, therefore requires a balanced learning process that affects students' cognitive, affective, and psychomotor domains. Assessment provides a formal means of measuring changes in cognitive, affective, and psychomotor domains. It should operate continuously through meaningful tasks and performance-based activities rather than being restricted to a particular realisation. To realise this objective, teachers should develop assessment instruments carefully to ensure that they measure memorisation but also students' appreciation and practice of moral values in daily life. In this context, a well-designed assessment helps ensure that students meet the intended learning objectives. The quality of assessment is therefore one of the important indicators that influences the success of PAI learning in shaping students' character in accordance with curriculum demands (Zahro', 2022).

In PAI learning, assessment serves as a key component for evaluating the overall achievement of learning objectives. However, previous studies have shown that the assessment instruments used by PAI teachers are still dominated by questions that require students to remember and understand religious information, such as mentioning definitions or citing textual arguments, without requiring the analysis to interpret values or solve moral problems (Samsudin & Iffah, 2019). Similar problems have also been found in curriculum and assessment studies outside the PAI context. Khan et al. (2025) emphasised that assessment items need to be aligned with curriculum objectives because

misalignment between content objectives, cognitive processes, and assessment strategies may weaken the quality of evaluation and hinder the development of critical thinking and problem-solving skills. This confirms that the alignment between test items and curriculum competencies is not merely an administrative requirement, but a core element of assessment validity.

Previous studies have also indicated that the tendency of PAI assessment to emphasise low-level cognitive skills is not merely a technical issue, but is closely related to teachers' assessment literacy (Ahmad & Sukiman, 2019; Gaikwad et al., 2023; Guci, 2023). Several studies have reported that PAI teachers still experience difficulties in constructing high-quality assessment instruments, particularly in formulating operational verbs, preparing test specifications, developing assessment grids, and aligning test items with Basic Competencies (KD) (Nst et al., 2025; Samsudin & Iffah, 2019). These difficulties are also influenced by limited time, administrative workload, and the continued use of traditional assessment practices. As a result, many assessment instruments are prepared without adequate KD analysis and without systematic item specifications. This condition causes the resulting test items to insufficiently represent the competency structure required by the curriculum and weakens the accuracy of assessment as a tool for measuring students' learning achievement in PAI (Gunartha, 2024; Sarikah et al., 2025).

The issue of cognitive demand is central in curriculum alignment studies because assessment tasks should reflect not only content coverage but also the expected level of thinking. Lemu et al. (2026) show that competency-based curriculum reforms may be weakened when learning materials and assessment tasks remain oriented toward factual retention rather than analytical rigour, while Saad et al. (2025) emphasise that Islamic education, including the *Tahfiz* Integrated Curriculum, should focus on memorisation by integrating Higher Order Thinking Skills (HOTS). This concern is also evident in PAI assessment practices, where previous studies report the dominance of cognitive and low-level assessment, limited evaluation of affective and psychomotor domains, and teachers' difficulties in developing authentic and HOTS-oriented instruments (Ahmad & Sukiman, 2019; Marfuah & Febriza, 2019; Samsudin & Iffah, 2019). Recent findings further indicate that many teachers still struggle to formulate operational verbs, map Basic Competencies, prepare assessment grids, and ensure item–material alignment, causing assessment instruments to remain dominated by LOTS and insufficiently supportive of students' analytical and critical thinking skills (Gunartha, 2024; Sarikah et al., 2025).

The preparation of assessment instruments in subjects within the PAI subject group, including *Aqidah Akhlak*, is still dominated by low-level cognitive questions at the remembering (C1) and understanding (C2) levels (Ahmad & Sukiman, 2019; Muhayimana et al., 2022). The question items are generally formulated in the form of instructions to explain definitions, mention terms, or cite textual arguments, without requiring students to analyse the meaning of faith values or apply concepts in real-life contexts (Ahmad & Sukiman, 2019). This pattern indicates that the abilities measured by teachers are categorised as Lower Order Thinking Skills (LOTS), because the evaluation instruments mainly require the reproduction of information and have not encouraged students to think critically, analyse moral values, or demonstrate concrete actions as manifestations of faith and moral understanding. The dominance of LOTS in examination questions has also been reported by Muhayimana et al. (2022), who used Bloom's Taxonomy to evaluate examination questions and found that lower-order

cognitive levels were far more dominant than higher-order thinking skills. This evidence strengthens the argument that the imbalance between LOTS and HOTS is a recurring problem in educational assessment and needs to be examined more specifically in *Aqidah Akhlak* assessment (Anderson et al., 1956).

Constructive Alignment Theory, initiated, emphasises that learning effectiveness can be achieved only when objectives, learning activities, and assessments are consistently aligned. This means that when the Basic Competencies of *Aqidah Akhlak* require analysing the meaning of faith values, interpreting moral teachings, and emulating commendable behaviour, the assessment instrument should not stop at measuring memorisation. Instead, it should measure students' analytical and internalise values, and demonstrate moral practices in daily behaviour (Biggs, 1996). Bloom, Anderson, and Krathwohl's taxonomic framework provides a basis for assessing whether the question items created by teachers have reached higher-order thinking levels, such as analysis, evaluation, and creation, or remain limited to remembering and understanding (Anderson et al., 1956). In addition, the affective domain in *Aqidah Akhlak* needs to be understood through Krathwohl's affective taxonomy, which describes the process of internalisation, from receiving and responding to organising and internalising values as inherent character traits. Alignment between learning objectives and assessment is especially important in religious subjects because religious competence requires the complete integration of knowledge, attitudes, and actions (Krathwohl et al., 1964). Thus, Constructive Alignment Theory and the Bloom Krathwohl taxonomic framework are important foundations for evaluating whether the *Aqidah Akhlak* assessment system truly reflects the competency demands formulated in the curriculum.

This study aims to analyse the level of conformity between the *Aqidah Akhlak* test items and the Basic Competencies stipulated in the curriculum. Specifically, this study maps each question item to the relevant KD indicator using an item-to-competency mapping technique to obtain an objective picture of the extent to which teachers' assessment instruments align with curriculum demands. Through this analysis, the research is expected to identify competency areas that are adequately accommodated and those that are neglected in the test items. Therefore, this study provides accurate information about the quality of assessment and supports efforts to improve the design of *Aqidah Akhlak* learning evaluation in a more targeted, systematic, and comprehensive manner.

## **II. METHOD**

This study employed a quantitative descriptive evaluative analysis to analyse the alignment of teacher-made *Aqidah Akhlak* assessment instruments with the Basic Competencies (KD) stated in the curriculum (Creswell & Guetterman, 2019). The alignment analysis in this study was conceptually grounded in Biggs' Constructive Alignment Theory, which emphasises that assessment tasks must be consistently aligned with the intended learning outcomes, competency demands, and cognitive processes expected in the curriculum (Biggs, 1996). Therefore, the item-to-competency mapping conducted in this study was not merely an administrative comparison between test items and curriculum documents, but a systematic analysis of the extent to which each assessment item represented the content scope and cognitive demand of the relevant KD (Khan et al., 2025; Webb, 1997). The research data consisted of 180 written Year-End Assessment (PAT) items for Grade X *Aqidah Akhlak* collected from four Madrasah Aliyah schools. Each item was treated as a unit of analysis and examined by referring

to the official curriculum document, particularly the Basic Competencies and their corresponding competency indicators.

The researchers followed a staged analytical procedure. They first identified the main material, concept, or value tested in each question item. They then compared the competency measured by each item with the relevant KD and competency indicator. Finally, they analysed the Operational Verb (*Kata Kerja Operasional/KKO*) in each to determine its alignment with the cognitive demand stated in the KD. Fourth, the cognitive level of each item was classified based on the revised Bloom's Taxonomy, covering C1 remembering, C2 understanding, C3 applying, C5 evaluating, and C6 creating (Anderson et al., 1956; Krathwohl et al., 1964). This study evaluated item alignment through the correspondence among content coverage, competency demand, and cognitive level, rather than through topic similarity alone (Chavda et al., 2024; Webb, 1997). Through this procedure, the researchers assessed item alignment based on both topic relevance and the item's ability to measure the competence required by the KD (Al-Aarifin Ismail et al., 2020; Webb, 1997).

To make the classification more transparent, this study used an item-alignment rubric consisting of three categories: aligned, partially aligned, and not aligned (Al-Aarifin Ismail et al., 2020; Webb, 1997). A category was categorised as aligned when the tested content, competency focus, and cognitive level were consistent with the relevant KD (Chavda et al., 2024; Webb, 1997). A category was categorised as partially aligned when it addressed the same material or topic as the KD but did not fully represent the cognitive demand or operational verb required by the KD (Anderson et al., 1956; Newton et al., 2020). For example, an item related to a KD using the operation "analyse" was considered partially aligned if it only required students to identify or define a concept. A category is categorised as not aligned when it tests material outside the scope of the KD, measures a different competency, or cannot be linked to any KD in the curriculum document. If a question item appeared to cover more than one KD, the item was mapped to the most dominant KD based on its main stimulus, answer key, and intended competency. When the dominant KD could not be determined clearly, the item was discussed by the coders until an agreement was reached.

The coding process was conducted by two independent raters with expertise in Islamic Religious Education assessment and curriculum analysis (McHugh, 2012; Stemler, 2004). The two raters independently coded the assessment items using a common coding sheet that recorded item number, tested material, relevant KD, competency indicator, operational verb, cognitive level, and alignment category. Afterwards, they compared the results to identify discrepancies in KD mapping, KKO interpretation, cognitive-level classification, and alignment decisions. The raters resolved coding discrepancies through consensus discussions guided by the official curriculum document, the revised Bloom's Taxonomy, and the item-alignment rubric. To ensure coding reliability, the researchers calculated inter-rater agreement using the percentage agreement between the two raters. This procedure was used to reduce subjectivity in interpreting KKO and classifying cognitive levels, considering that Bloom's Taxonomy classification involves interpretive judgment (Anderson et al., 1956; Stemler, 2004). To analyse descriptively, the researchers calculated the frequency and percentage of item alignment with KD (Creswell & Guetterman, 2019). They then examined the cognitive-level distribution to identify the dominant thinking level represented in the *Aqidah Akhlak* assessment instruments. The results of the analysis were presented in tables and graphs to provide a comprehensive view of the alignment between teacher-made assessment instruments

and curriculum demands, as well as the cognitive quality of the test items (Khan et al., 2025; Webb, 1997).

### **III. RESULT AND DISCUSSION**

#### **Curriculum Coverage and Item Competency Alignment in *Aqidah Akhlak* Assessment**

This study analysed 180 Year-End Assessment (PAT) items for Grade X *Aqidah Akhlak* obtained from four Madrasah Aliyah, consisting of three private madrasahs and one public madrasah. Teachers officially used all assessment instruments during one academic year and therefore represent the summative assessment practice at the education-unit level. The items consisted mainly of multiple-choice questions, supplemented in several madrasahs by essay and true-false items. These variations indicate that teachers use different assessment formats to measure students' factual knowledge, conceptual understanding, reasoning ability, and religious-moral attitudes. In classroom-based assessment, the diversity of item formats is important, but it does not automatically guarantee that assessment practices are aligned with curriculum policy, competency standards, and intended learning outcomes (Hays et al., 2024; Yusoff et al., 2025).

Since *Aqidah Akhlak* teachers developed the test items independently within their respective madrasahs, the analysed instruments functioned as analytic artefacts that reflected teachers' interpretation of Basic Competencies (KD), competency indicators, and assessment demands. Accordingly, the researchers applied item-to-competency mapping to evaluate each item's representation of the content scope and cognitive demand of Grade X *Aqidah Akhlak* KD. Previous curriculum alignment studies emphasise that assessment instruments need to be aligned not only from the perspective of topic coverage, but also from the extent to which they represent curriculum goals, competency structures, and cognitive processes (Chan & Lee, 2021; Khan et al., 2025; Lemu et al., 2026).

The suitability of the test items with KD was determined using an item-alignment rubric developed from the principles of constructive alignment, the official curriculum document, and the revised Bloom's Taxonomy. The rubric contained three main indicators: content coverage, competency focus, and cognitive demand. Content coverage refers to whether the material tested in the item corresponds to the material scope stated in the KD. Competency focus refers to whether the item measures the type of competence required by the KD. Cognitive demand captures the correspondence between the item's Operational Verb (*Kata Kerja Operasional/KKO*), its required thinking level, and the cognitive level demanded by the KD.

An item demonstrated alignment when it fulfilled the categorised components. By contrast, an item lacked alignment when it only showed categorisation, did not assess the required competence, operated below the KD's cognitive demand, or tested material outside the relevant KD. Thus, item suitability in this study was not determined merely by topic similarity, but by the degree of correspondence between tested material, competency demand, and cognitive level. This procedure is consistent with studies that recommend the use of content-validity logic, rubrics, assessment blueprints, and alignment matrices to evaluate whether assessment items properly represent learning

objectives and cognitive levels (Al-Aarifin Ismail et al., 2020; Burns et al., 2023; García-Beltrán & Cantero-Sandoval, 2025; Rahayu & Hariwibowo, 2021).

**Table 1. Conformity of Class X Moral Aqidah Questions with Basic Competencies**

KD	KD Description	Total Questions	Conform	Inappropriate	Percentage Match
3.1	Analyse the meaning, causes, and negative impacts of reprehensible traits, including <i>hubbuddunya</i> , <i>hasad</i> , <i>ujub</i> , arrogance, <i>riya'</i> , and their related behaviors	8	5	3	62,5%
3.2	Analyse the obligatory attributes of Allah, including <i>nafsiyah</i> , <i>salbiyah</i> , <i>ma'ani</i> , and <i>ma'nawiyah</i> , as well as the possible attributes of Allah	22	21	1	95,5%
3.3	Analyse the meaning, conditions, and position of repentance as the foundation of the spiritual journey	0	0	0	0%
3.4	Analyse the virtues and manners of honouring parents and teachers based on religious evidence and scholars' opinions	18	2	16	11%
3.5	Analyse the exemplary lessons from the story of Prophet Luth (a.s.)	21	16	5	76,2%
3.6	Analyse the meanings of selected Asmaul Husna, including <i>al-Karim</i> , <i>al-Analyse al-Wakil</i> , <i>al-Matin</i> , <i>al-Jami'</i> , <i>al-Hafiz</i> , <i>al-Rafi'</i> , <i>al-Wahhab</i> , <i>al-Raqib</i> , <i>al-Mubdi'</i> , <i>al-Muhyi</i> , <i>al-Hayyu</i> , <i>al-Qayyum</i> , <i>al-Akhir</i> , <i>al-Mujib</i> , and <i>al-Awwal</i>	18	13	5	72,2%
3.7	Analyse the meaning, textual evidence, and characteristics of Islam. Analyse or moderate Islam, as well as the characteristics of radical religious understanding	18	12	6	66,6%
3.8	Analyse the nature of desire and anger, as well as ways to control the Analysegh <i>mujahadah</i> and <i>riyadhah</i> as part of self-purification	18	7	11	38,8%
3.9	Analyse the meaning and virtues of the main moral qualities, namely <i>hiAnalyseffah</i> , <i>syaja'ah</i> , and <i>'adalah</i>	36	24	12	66,6%
3.10	Analyse the causes, examples, and ways to avoid cunning, greed, and injustice. Analyse discriminatory behavior	15	10	5	66,6%
3.11	Analyse the manners and wisdom of visiting the sick	6	6	0	100%
Total		180	116	64	64,4%

$$\text{Conformity Percentage} = \frac{\text{Number of Conforming Questions}}{\text{Total Questions}} \times 100\%$$

Figure 1. Formula for Calculating Basic Competency Suitability

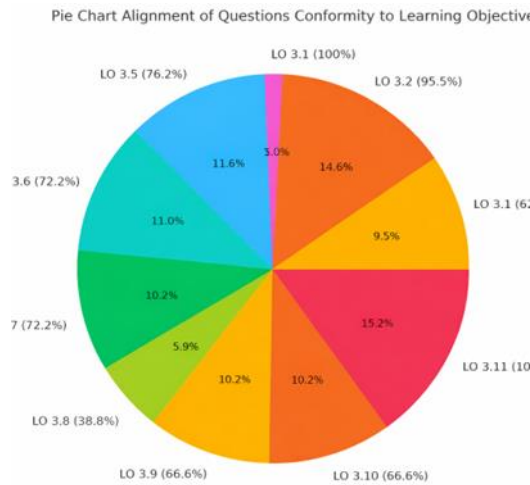


Figure 2. Pie Chart Alignment Questions Against KD

Based on table 1, the level of item alignment with Basic Competencies shows considerable variation across KD. Of the 180 items analysed, 116 items, or 64.4%, were categorised as aligned with KD, while 64 items, or 35.6%, were categorised as not aligned. KD 3.11 obtained the highest alignment, followed by KD 3.2 at 95.5%. This indicates that in several KD areas, teachers were able to construct assessment items that corresponded to the expected competency demands. However, the distribution was not balanced. KD 3.4 obtained only 11% alignment, while KD 3.3 was not represented at all in the assessment instruments. These findings indicate that the alignment of *Aqidah Akhlak* assessment items with KD is still uneven and that several competencies are underrepresented in the Year-End Assessment. From the perspective of content validity, this imbalance is problematic because assessment scores cannot be interpreted as a comprehensive representation of students' mastery when several KD are weakly represented or not measured. Content validity requires that the domain of competencies formulated in the curriculum be adequately represented in assessment instruments. When some KD are overrepresented while others are neglected, the assessment tends to measure only selected parts of the curriculum rather than the full structure of expected competence. Similar problems have been identified in assessment alignment studies, which show that item flaws, weak correspondence between objectives and questions, and insufficient representation of cognitive demands may reduce the validity and usefulness of assessment results (Divjak et al., 2023; García-Beltrán & Cantero-Sandoval, 2025; Khan et al., 2025).

In the framework of constructive alignment, assessment should directly measure the intended learning outcomes formulated in the curriculum. When KD requires students to analyse, evaluate, or interpret religious and moral values, the assessment instrument should not merely ask students to recall definitions or identify terms. The finding that

the overall alignment level only reached 64.4% indicates that teacher-made assessment instruments have not fully reflected the competency structure required in the *Aqidah Akhlak* curriculum. This condition suggests that the problem is not only a technical item-writing weakness, but also a curriculum-alignment issue that affects the validity and usefulness of the assessment as a tool for measuring learning achievement. In Islamic Religious Education, this alignment becomes more important because curriculum design should connect religious purposes, learning content, teaching practice, and assessment in a coherent structure (Ansyari, 2023).

The uneven distribution of KD representation also indicates a gap between the planned curriculum and the assessed curriculum. Curriculum mapping studies show that the relationship among learning outcomes, learning materials, and assessment tasks must be systematically documented so that no essential competency is ignored in the assessment process (Frost & Ackrill, 2025; Geshere et al., 2026). In this study, KD 3.3 was not represented at all, while KD 3.4 and KD 3.8 showed low alignment. This suggests that value-oriented and reflective competencies are more difficult to translate into test items than conceptual-theological competencies. Therefore, assessment development in *Aqidah Akhlak* requires not only item writing but also systematic blueprinting, curriculum mapping, and validation of item-objective correspondence (Al-Aarifin Ismail et al., 2020; Chavda et al., 2024).

### **Cognitive Demand of Teacher-Made Items: The Dominance of LOTS over HOTS**

**Table 2. Distribution of Cognitive Level Question Items Based on Bloom's Taxonomy**

Cognitive Level (Bloom)	Number of Questions	Percentage (%)	Remarks
C1 (Remembering)	40	22,2%	The question focuses on memorising facts or terms, not yet demanding complex conceptual elaboration
C2 (Understand)	86	47,7%	The most dominant level, the majority of questions only ask for explanations or simple descriptions, but still show LOTS
C3 (Apply)	30	16,6%	It has started to measure the ability to implement concepts, but the proportion is still limited.
C4 (Analyse)	17	9,4%	Analytical questions began to emerge, but they were not comparable to the demands of KD, who used a lot of analytical KKO
C5 (Evaluate)	4	2,2%	There are very few evaluative questions; the ability to assess and give consideration has not been adequately accommodated in the assessment instruments.
C6 (Create)	3	1,6%	Questions that demand creativity and the preparation of new ideas are almost unused
Total	180	100%	The composition of the questions is dominated by LOTS (C1–C2 = 70%), while HOTS (C4–C6) is only about 13%, indicating that the assessment has not encouraged high-level thinking

$$\text{Percentage} = \frac{\text{Number of Questions}}{180} \times 100\%$$

Figure 3. Cognitive level percentage calculation formula

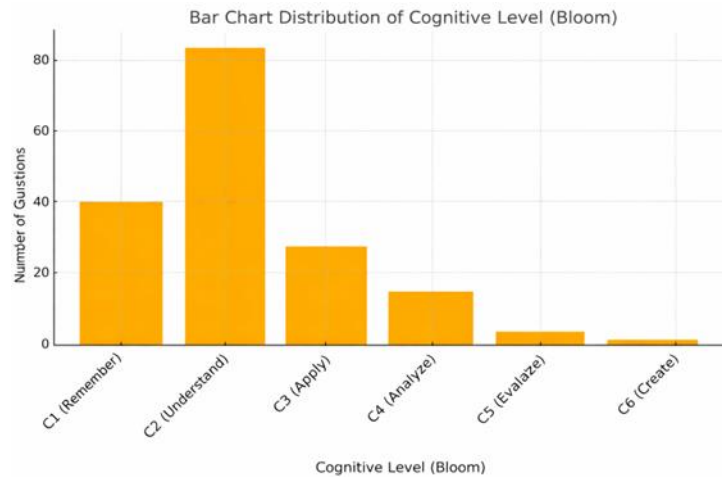


Figure 4. Bloom Cognitive Level Distribution Bar Diagram

The distribution of cognitive levels in Table 2 shows that Grade X *Aqidah Akhlak* assessment items are predominantly located in the Lower Order Thinking Skills (LOTS) domain. Of the 180 items analysed, the largest proportion was at C2, or understanding, with 86 analysed, equivalent to 47.7%. This was followed by C1, or remembering, with 40 items, equivalent to 22.2%. Taken together, C1 and C2 account for 70% of the total items. This means that most items still focus on the ability to remember factual information, explain concepts, identify terms, or understand basic religious information. Studies that use Bloom's Taxonomy to analyse examination questions similarly show that teacher-made and for-analyse test items are often concentrated at lower cognitive levels, while analytical, evaluative, and creative levels are less represented (Damayanti et al., 2020; Khoy, 2025; Muhayimana et al., 2022). The C3 level, or applying, consisted of 30 items, equivalent to 16.6%. This indicates that some items had begun to measure students' ability to apply *Aqidah Akhlak* concepts in certain contexts. However, this proportion remained limited compared with the curriculum demand that frequently uses analytical operational verbs. Meanwhile, HOTS-oriented items were very limited. C4, or analysing, accounted for only 17 items, or 9.4%; C5, or evaluating, accounted for only 4 items, or 2.2%; and C6, or creating, accounted for only 3 items, or 1.6%. Overall, HOTS items only reached approximately 13% of the instrument.

This condition indicates a mismatch between the cognitive demand stated in the KD and the cognitive level measured by the assessment items. Many KD in *Aqidah Akhlak* use analytical operational verbs, such as analysing the meaning of values, identifying causes and consequences of behaviour, interpreting religious arguments, and examining the wisdom of moral teachings. However, most items only require students to recognise, remember, or understand information. Similar findings have been recognised in studies on HOTS and LOTS, which argue that the transformation from lower-order to higher-order thinking requires deliberate pedagogical design, clear operational verbs, and

assessment tasks that require reasoning, evaluation, and creation (Newton et al., 2020; Prakash & Litoriya, 2022; Susantini et al., 2025).

The dominance of LOTS has important pedagogical implications. When assessment emphasises remembering and understanding, teachers tend to direct class construction toward verbal explanation, concept memorisation, and closed-ended practice questions (Biggs, 1996; Brookhart, 2018). Students may consequently view *Aqidah Akhlak* as a subject centred on memorising definitions, examples, and textual arguments rather than developing moral reasoning and reflective understanding. Memorising narrows the space for moral reasoning, ethical reflection, and contextual analysis, even though these abilities are central to the learning objectives of *Aqidah Akhlak* (Firdaus et al., 2026; Maqi, 2025). In the context of Islamic education, Ahmad & Sukiman (2019) found that Dirasah Islamiyah examination items were still dominated by lower-level cognitive processes, while Saad et al. (2025) emphasised that Islamic learning, including tahfiz-based curricula, is not limited to memorisation by strengthening HOTS and deeper understanding.

From an memorisation literacy perspective, the dominance of LOTS also reflects the tendency of teachers to use item formats that are easier to construct and score objectively. Multiple-choice questions can be useful for measuring factual and conceptual knowledge. However, they become limited when the curriculum requires students to analyse moral problems, evaluate ethical choices, or formulate reflective responses. Research on teachers' test construction competence indicates that many teachers still experience difficulty designing assessment items that represent HOTS, particularly when they must connect learning objectives, cognitive levels, item formats, and scoring criteria (Agama & Agboro, 2025; Epinasti et al., 2021; Guci, 2023). Therefore, the cognitive imbalance found in this study suggests that teachers need stronger support in translating analytical KD into assessment tasks that require higher-order thinking.

### **Evidence of Item Competency Misalignment in *Aqidah Akhlak* Assessment**

The misalignment between items and KD becomes clearer when examining several examples of question items. In KD 3.1, students are expected to analyse the meaning, causes, and negative impacts of reprehensible traits such as *hubbuddunya*, *hasad*, *ujub*, arrogance, *riya'*, and their derivative behaviours. However, one item asks students to identify the type of reprehensible trait reflected in a contextual situation: "Based on the perspective of Islamic ethics, the behaviour of officials who refuse to sit on a par with students reflects ...." The answer options include *kibr*, *ujub*, and *riya'*. Although the item uses a contextual stimulus, the cognitive task only requires students to identify the correct term. The item does not require students to analyse the causes, implications, or moral consequences of the behaviour. The item is more accurately classified at C2, or understanding, rather than C4, or analysing, as demanded by the KD.

This finding illustrates that the usual analysis of the stimulus does not automatically make an item HOTS-oriented. An item can appear contextual but remain at a lower cognitive level if students only need to recognise a term or select a definition. Studies on Bloom-based assessment confirm that the cognitive level of an item should be determined by the thinking process required to answer it, not merely by the presence of a stimulus or narrative context (García-Beltrán & Cantero-Sandoval, 2025; Muhayimana et al., 2022; Newton et al., 2020). Therefore, alignment analysis must

consider both the content being tested and the cognitive operation required from students.

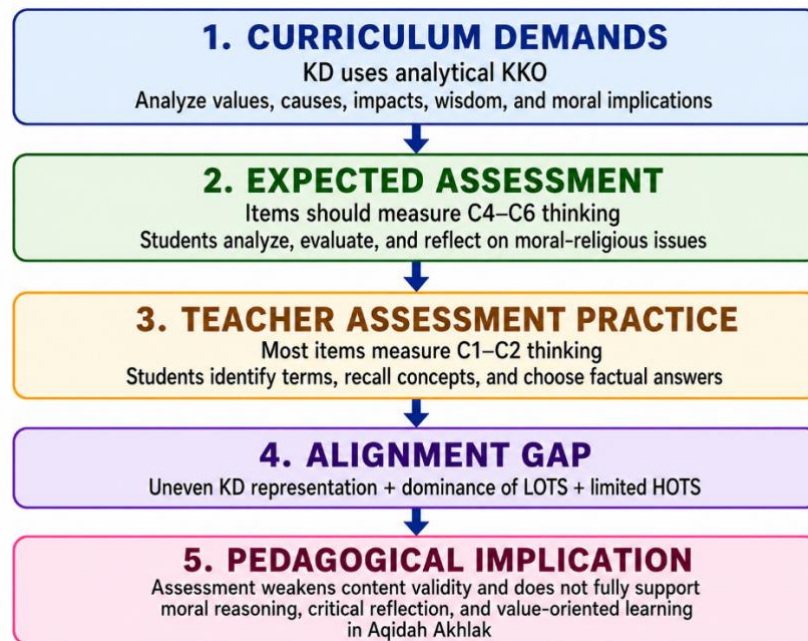
A similar inconsistency was found in KD 3.4, which requires students to analyse the virtues and manners of filial piety to parents and teachers, based on evidence and scholarly opinions. One item associated with this KD asks students to choose an attitude that corresponds to the hadith stating that Allah's pleasure depends on parents' pleasure. The item only requires students to select an example of appropriate behaviour. It does not require students to analyse the wisdom of the hadith, compare scholarly arguments, or explain the ethical implications of filial piety in a specific context. Consequently, the item measures understanding rather than analysis, which explains the low alignment level of KD 3.4. A more fundamental form of misalignment appears in KD 3.3, which requires students to analyse the nature, conditions, and position of repentance as the foundation of the spiritual journey. The mapping results show that no item explicitly measured this KD. No question tested the meaning of repentance, its conditions, its theological position, or its role in *tazkiyatun nafs*. The omission of this KD from the Year-End Assessment suggests that the assessment failed to capture a competency explicitly required in the curriculum. This finding indicates a clear disconnect between the planned curriculum and the assessed curriculum.

These examples confirm that misalignment occurs in two forms. First, several items are topically related to KD but do not measure the cognitive demand required by the KD. Second, some KD are not represented at all in the assessment instrument. Both forms of misalignment reduce the content validity of the instrument because the assessment does not fully measure the competencies that students are expected to achieve. Curriculum mapping studies similarly show that weak alignment among objectives, learning materials, and assessment tasks may produce incomplete competency coverage and weaken the accuracy of educational evaluation (Geshere et al., 2026; Khan et al., 2025; Lemu et al., 2026).

In addition, the examples above show that assessment alignment cannot be separated from the quality of teachers' interpretation of operational verbs. The verb "analyse" in the KD requires students to identify relationships, explain analyses, examine consequences, compare arguments, or interpret value implications. If the item only asks students to recall a term or recognise a correct example, the item does not sufficiently represent the recognitional demand of the KD. This is why the use of Bloom's Taxonomy must be supported by clear action verbs and item validation procedures (Chavda et al., 2024; Khoy, 2025; Newton et al., 2020).

### **The Gap between Curriculum Demands and Teacher Assessment Practice**

Theoretically, the *Aqidah Akhlak* curriculum requires alignment among KD, analytical operational verbs, learning activities, and assessment instruments. The curriculum expects students not only to master religious concepts but also to analyse values, interpret moral teachings, and reflect on ethical behaviour; however, the findings of this study reveal a gap between curriculum expectations and teacher assessment practice. The average alignment level of 64.4% indicates that assessment items do not fully correspond to KD. At the same time, the dominance of LOTS items shows that the cognitive quality of assessment remains lower than the analytical demand of the curriculum. The gap can be illustrated as follows:



*Figure 5. Gap Between Curriculum Demands and Teacher Assessment Practices*

This diagram shows that the gap is not merely caused by the number of inappropriate items, but by the mismatch between the intended curriculum, the expected cognitive process, and the actual assessment format. KD that are conceptual and theological in nature, such as KD 3.2, obtained a high alignment level of 95.5%. In contrast, KD that demand value reflection and moral internalisation, such as KD 3.4 and KD 3.8, obtained much lower alignment levels, namely 11% and 38.8%. KD 3.3 was not accommodated at all. This pattern indicates that teachers are more successful in measuring conceptual-theological material than reflective and analytical moral competencies.

These findings are consistent with previous studies on HOTS-based assessment and curriculum alignment. Ahmad (2019) found that Islamic studies examination items were dominated by C1 and C2 levels, with limited representation of analysis and evaluation. Similar patterns are found in many formal exams and teacher-made tests that still focus on remembering and understanding (Chavda et al., 2024; Damayanti et al., 2020) Muhayimana et al., 2022). In a more recent context, Susantini et al. (2025) emphasised the need for systematic support to help teachers design HOT emphasised assessment questions based on the revised Bloom's Taxonomy. Therefore, the findings of this study reinforce the argument that the dominance of LOTS is not an isolated case in *Aqidah Akhlak* but is part of a broader assessment-literacy problem in classroom evaluation.

Several factors may explain this gap. First, teachers' assessment literacy may still be insufficient to translate analytical KD into valid assessment items. Teachers may understand the curriculum conceptually, but still experience difficulty in constructing items that measure analysis, evaluation, and value reflection. This explanation is supported by studies showing that teachers often need further training in assessment literacy, test construction, assessment standards, and the use of cognitive taxonomies in designing valid assessment tasks (Ahmadizadeh et al., 2026; Gaikwad et al., 2023; Gucci, 2023; Yildirim et al., 2024). Second, administrative workload and limited preparation time may encourage teachers to use technically efficient items, especially multiple-choice questions that are easier to score. Third, the dominance of summative assessment culture may lead teachers to prioritise final scores rather than formative, reflective, and

authentic processes. These factors show that the problem of assessment alignment is structural and pedagogical, not merely technical.

These findings imply that the development of PAI assessment should not only focus on improving item-writing skills, but also on strengthening teachers' understanding of curriculum alignment, content validity, and cognitive demand. *Aqidah Akhlak* assessment needs to be developed as an authentic and value-oriented assessment that can measure moral reasoning, ethical reflection, and the application of Islamic values in real-life contexts. Studies on assessment validity and constructive alignment show that assessment quality depends on the coherence between learning outcomes, instructional design, assessment criteria, and evidence of student learning (Burns et al., 2023; Chan & Lee, 2021; Divjak et al., 2023). Therefore, the findings of this study provide an important basis for improving assessment literacy among PAI teachers and for redesigning assessment instruments that are more consistent with the objectives of religious education.

Overall, the findings show that the assessment problem in *Aqidah Akhlak* is located at the intersection of curriculum alignment, content validity, cognitive demand, and teacher assessment literacy. The alignment percentage of 64.4% indicates that a substantial portion of the instrument has already corresponded with KD. However, the low representation of several competencies and the dominance of LOTS reveal that the assessment has not fully captured the intended quality of *Aqidah Akhlak* learning. Therefore, improving assessment instruments requires three key priorities: developing KD-mapped assessment blueprints, constructing items that accurately assess analytical and reflective competencies, and strengthening teacher capacity in designing HOTS-oriented and value-oriented assessments.

#### **IV. CONCLUSION**

This study concludes that the Grade X *Aqidah Akhlak* assessment instruments show a moderate level of alignment with the Basic Competencies, reaching 64.4%. However, the findings reveal uneven competency representation, with some Basic Competencies adequately covered and others weakly represented or excluded from measurement. The cognitive profile of the assessment instruments was also dominated by Lower Order Thinking Skills (LOTS), accounting for 70%, whereas Higher Order Thinking Skills (HOTS) reached only 13%. These findings indicate that teacher-made assessments still tend to emphasise factual recall and conceptual understanding rather than evaluative and reflective abilities required by the curriculum. This condition has implications for the content validity of the *Aqidah Akhlak* assessment because students' scores may not fully represent the expected competencies, particularly in moral reasoning and value internalisation. However, this study is limited to 180 Year-End Assessments of four Madrasah Aliyah and relies solely on document-based item mapping, without examining teachers' perspectives or classroom assessment practices. Therefore, future research should combine item-to-competency mapping with qualitative inquiry involving teachers, curriculum teams, and classroom observations to explore why certain competencies are underrepresented and how teachers construct assessment instruments in real educational contexts.

## V. REFERENCES

- [1] Agama, A. I., & Agboro, A. A. (2025). Assessing Computer Science Teachers' Competence in Constructing Essay Test Items: A Comparative Study of Lower and Higher Order Thinking Skills in Junior Secondary Schools in Benue State. *International Journal of Emerging Research in Engineering, Science, and Management*, 4(2), 68–76. <https://doi.org/10.58482/ijeresm.v4i2.9>
- [2] Ahmad, I. F. (2019). Analisis Higher Order Thinking Skills (hots) pada Soal Ujian Akhir Siswa Kelas 6 Kmi dalam Kelompok Mata Pelajaran Dirasah Islamiyah di Pondok Modern Tazakka Batang. *Jurnal Pendidikan Agama Islam*, 16(2), 137–164. <https://doi.org/10.14421/jpai.2019.162-02>
- [3] Ahmad, I. F., & Sukiman, S. (2019). Analisis Higher Order Thinking Skills (HOTS) Pada Soal Ujian Akhir Siswa Kelas 6 KMI dalam Kelompok Mata Pelajaran Dirosah Islamiyah di Pondok Modern Tazakka Batang. *Jurnal Pendidikan Agama Islam*, 16(2), 137–164. <https://doi.org/10.14421/jpai.2019.162-02>
- [4] Ahmadizadeh, M., Khodabakhshzadeh, H., Rostamian, M., & Hamid, A. (2026). The State of Assessment Training in EFL Teachers' Courses: Aligning Policies, Standards, and Language Assessment Course Syllabi. *Language Testing in Asia*, 16(1). <https://doi.org/10.1186/s40468-026-00448-2>
- [5] Al-Aarifin Ismail, M., Pa, M. N. M., Al-Muhammady Mohammad, J., & Yusoff, M. S. B. (2020). Seven Steps to Construct an Assessment Blueprint: A Practical Guide. *Education in Medicine Journal*, 12(1), 71–80. <https://doi.org/10.21315/eimj2020.12.1.8>
- [6] Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (Eds.). (1956). *Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. David McKay Company, Inc.
- [7] Ansyari, M. F. (2023). Design of Islamic Religious Education: Purposes, alignment of curriculum components and contexts. *British Journal of Religious Education*, 45(4), 382–393. <https://doi.org/10.1080/01416200.2023.2220940>
- [8] Awali, S. (2020). Strategi Guru Pendidikan Agama Islam dalam Membentuk Karakter Religius Siswa MIN Pirak Kuta Makmur Kabupaten Aceh Utara. *Jurnal At-Tarbiyyah: Jurnal Pendidikan Islam*, 5(2). <https://doi.org/10.54621/jiat.v6i2.381>
- [9] Biggs, J. (1996). Enhancing Teaching Through Constructive Alignment. *Higher Education*, 32(3), 347–364. <https://doi.org/10.1007/BF00138871>
- [10] Brookhart, S. M. (2018). Appropriate Criteria: Key to Effective Rubrics. *Frontiers in Education*, 3(22), 22. <https://doi.org/10.3389/feduc.2018.00022>

- [11] Burns, A., de Bruyn, L. L., & Wilson, S. C. (2023). A rubric approach to assessing information literacy competency in tertiary curricula. *Journal of University Teaching and Learning Practice*, 20(1). <https://doi.org/10.53761/1.20.01.10>
- [12] Chan, C. K. Y., & Lee, K. K. W. (2021). Constructive alignment between holistic competency development and assessment in Hong Kong engineering education. *Journal of Engineering Education*, 110(2), 437–457. <https://doi.org/10.1002/jee.20392>
- [13] Chavda, M., Patel, H., & Bhatt, H. (2024). Quality Education Through Writing: Aligning Learning Objectives in Learning Materials and Question Papers Using Bloom’s Taxonomy. *Quality Assurance in Education*, 32(1), 96–110. <https://doi.org/10.1108/QAE-03-2023-0045>
- [14] Creswell, J. W., & Guetterman, T. C. (2019). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (Sixth edition). Pearson.
- [15] Damayanti, N., Subali, B., Nugroho, S. E., & Sureporn, K. (2020). Items Analysis of Physics Assessment Based on Cognitive Level of High Order Thinking Skills in Bloom Taxonomy. *J. Phys. Conf. Ser.*, 1521(2). <https://doi.org/10.1088/1742-6596/1521/2/022022>
- [16] Divjak, B., Svetec, B., Horvat, D., & Kadoić, N. (2023). Assessment validity and learning analytics as prerequisites for ensuring student-centred learning design. *British Journal of Educational Technology*, 54(1), 313–334. <https://doi.org/10.1111/bjet.13290>
- [17] Epinasti, G., Slamet, S. T. Y., & Yamtinah, S. (2021). Teacher’s Perception towards Assessing of Higher Order Thinking Skills (HOTS) in Elementary Schools. In *Proceedings of the 5th International Conference on Learning Innovation and Quality Education*. <https://doi.org/10.1145/3516875.3516980>
- [18] Firdaus, Ra. T., Kuswandi, I., & Ridwan, M. (2026). Enhancing Islamic Moral Understanding and Critical Thinking through an Augmented Reality Qur’anic Storytelling Platform in Primary Education. *AL-ISHLAH: Jurnal Pendidikan*, 17(4). <https://doi.org/10.35445/alishlah.v17i4.9116>
- [19] Frost, D., & Ackrill, R. (2025). The Multiple Dimensions of Curriculum Mapping: Designing a Comprehensive Outcomes-Based Framework. *London Review of Education*, 23(1). <https://doi.org/10.14324/LRE.23.1.17>
- [20] Gaikwad, S., Wadegaonkar, A., Mitra, G., & Chakravarty, D. (2023). Assessment Literacy, Current Assessment Practices and Future Training: Reflections of Teachers in Higher Education. *International Journal of Learning, Teaching and Educational Research*, 22(7), 1–29. <https://doi.org/10.26803/ijlter.22.7.1>
- [21] García-Beltrán, E., & Cantero-Sandoval, M. A. (2025). Assessment Based on the Revised Bloom’s Taxonomy: Theoretical Framework and Alignment Matrix for the

- Design and Validation of Multiple-Choice Items in Online Higher Education. *EduTec*, 2025-December(94), 313–329. <https://doi.org/10.21556/edutec.2025.94.4079>
- [22] Geshere, G. L., Tarekegne, W. M., & Woyessa, D. D. (2026a). An Analysis of Assessment Tasks and Learning Outcomes in Grade 10 History Curricular Materials in the Era of Competency-Based Curriculum Reform. *Cogent Education*, 13(1). <https://doi.org/10.1080/2331186X.2026.2633000>
- [23] Guci, R. I. (2023). HOTS-Based Language Assessment Literacy: Challenges and Prospects in English Language Teaching. *Indonesian Journal of Applied Linguistics*, 12(3), 818–827. <https://doi.org/10.17509/ijal.v12i3.44261>
- [24] Gunartha, I. W. (2024). Pengembangan Penilaian Berorientasi Hots: Upaya Peningkatan Kemampuan Berpikir Kritis Siswa di Era Global Abad Ke-21. *Widyadari*, 25(1), 133–147. <https://doi.org/10.59672/widyadari.v25i1.3660>
- [25] Hays, R. B., Wilkinson, T., Green-Thompson, L., McCrorie, P., Bollela, V., Nadarajah, V. D., Anderson, M. B., Norcini, J., Samarasekera, D. D., Boursicot, K., Malau-Aduli, B. S., Mandache, M. E., & Nadkar, A. A. (2024). Managing assessment during curriculum change: Ottawa Consensus Statement. *Medical Teacher*, 46(7), 874–884. <https://doi.org/10.1080/0142159X.2024.2350522>
- [26] Khan, H. F., Qayyum, S., Beenish, H., Khan, R. A., Iltaf, S., & Faysal, L. R. (2025a). Determining the Alignment of Assessment Items with Curriculum Goals Through Document Analysis by Addressing Identified Item Flaws. *BMC Medical Education*, 25(1). <https://doi.org/10.1186/s12909-025-06736-4>
- [27] Khoy, B. (2025). Unlocking Cognitive Learning Objectives: A Comprehensive Evaluation of How Textbooks and Syllabi Align with Revised Bloom's Taxonomy Across Disciplines. *Curriculum Perspectives*, 45(2), 189–202. <https://doi.org/10.1007/s41297-024-00295-2>
- [28] Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Longman Inc.
- [29] Lemu, G., Woyessa, D. D., & Tarekegne, W. M. (2026). Cognitive Demands and Curriculum Alignment in Ethiopia's Grade 10 History Textbook: Bridging Policy and Practice. *Curriculum Journal*. <https://doi.org/10.1002/curj.70046>
- [30] Maqi, L. H. A. (2025). Contextual Learning and Teaching to Improve Students Understanding of Aqidah and Akhlak. *Ats-Tsaqofi: Jurnal Pendidikan Dan Manajemen Islam*, 6(2), 83–95. <https://doi.org/10.61181/ats-tsaqofi.v6i2.499>
- [31] Marfuah, A., & Febriza, F. (2019). Penilaian Autentik pada Pembelajaran Pendidikan Agama Islam (PAI) di Sekolah dan Perguruan Tinggi. *FONDATIA*, 3(2), 35–58. <https://doi.org/10.36088/fondatia.v3i2.301>
- [32] McHugh, M. L. (2012). Interrater Reliability: The Kappa Statistic. *Biochemia Medica*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>

- [33] Muhayimana, T., Kwizera, L., & Nyirahabimana, M. R. (2022). Using Bloom's Taxonomy to Evaluate the Cognitive Levels of Primary Leaving English Exam Questions in Rwandan Schools. *Curriculum Perspectives*, 42(1), 51–63. Scopus. <https://doi.org/10.1007/s41297-021-00156-2>
- [34] Newton, P. M., Da Silva, A., & Peters, L. G. (2020). A Pragmatic Master List of Action Verbs for Bloom's Taxonomy. *Frontiers in Education*, 5. Scopus. <https://doi.org/10.3389/feduc.2020.00107>
- [35] Nst, T. M., Remiswal, R., & Khadijah, K. (2025). Implementasi Penilaian Autentik pada Materi Pembelajaran Pendidikan Agama Islam di SDN 329 Suka Damai. *YASIN*, 5(4), 3184–3197. <https://doi.org/10.58578/yasin.v5i4.5953>
- [36] Prakash, R., & Litoriya, R. (2022). Pedagogical Transformation of Bloom Taxonomy's LOTs into HOTs: An Investigation in Context with IT Education. *Wireless Personal Communications*, 122(1), 725–736. Scopus. <https://doi.org/10.1007/s11277-021-08921-2>
- [37] Rahayu, W., & Hariwibowo, H. (2021). Analysis Of the Characteristics of HOTS-Based Assessment for Learning Items Using the Rasch Model. *Review of International Geographical Education Online*, 11(5), 2511–2525. Scopus. <https://doi.org/10.48047/rigeo.11.05.149>
- [38] Saad, M. F. M., Mohd Ataa, N. S., Mohamad, S., Ishak, H., & Nur, A. (2025). Minds Illumination: Exploring Higher Order Thinking Skills (HOTS) in the Tahfiz Integrated Curriculum. *Islamiyyat*, 47(1), 96–106. Scopus. <https://doi.org/10.17576/islamiyyat-2025-4701-08>
- [39] Samsudin, M. A., & Iffah, U. (2019). Penilaian Autentik pada Matapelajaran Pendidikan Agama Islam. *Edupeedia*, 4(1), 77–85. <https://doi.org/10.35316/edupedia.v4i1.528>
- [40] Sarikah, Andrizal, & A Mualif. (2025). Analisis Implementasi Taksonomi Bloom pada Instrumen Tes Kognitif (Mata Pelajaran Al-Quran Hadits Kelas XI MA Plus Ulil Albab Kec. Inuman). *JOM FTK UNIKS*, 5(1). <https://ejournal.uniks.ac.id/index.php/JOM/article/view/4329>
- [41] Stemler, S. E. (2004). A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability. *Practical Assessment, Research & Evaluation*, 9(1). <https://doi.org/10.7275/96jp-xz07>
- [42] Susantini, E., Sari, Y. M., Marzuqi, M. I., & Asteria, P. V. (2025). EduQuestioning Mobile Learning Application: A Catalyst for Developing HOTS-Based Assessment Questions Referring to Revised Bloom's Taxonomy. *Research and Practice in Technology Enhanced Learning*, 20. Scopus. <https://doi.org/10.58459/rptel.2025.20022>
- [43] Webb, N. L. (1997). Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. *Council of Chief State School Officers*, (6).

- [44] Yildirim, A., Stjernkvist, M. P., Hilden, R., Fröjdendahl, B., & Oscarson, A. D. (2024). Developing Summative Assessment Literacy: Novice Language Teachers' Perceptions in Sweden and Finland. *Review of Education*, 12(2). Scopus. <https://doi.org/10.1002/rev3.3487>
- [45] Yusoff, S. M., Marzaini, A. F. M., Zakaria, N., & Lijie, H. (2025). Exploring the Implementation of Classroom-Based Assessment in Malaysian Secondary Schools: Alignment with Policy and Teacher Practices. *Malaysian Online Journal of Educational Management*, 13(1), 81–101. <https://doi.org/10.22452/mojem.vol13no1.5>
- [46] Zahro', A. (2022). Kontribusi Penilaian Autentik dalam Pembelajaran Pendidikan Agama Islam. *Rabbani: Jurnal Pendidikan Agama Islam*, 3(2), 208–221. <https://doi.org/10.19105/rjpai.v3i2.6358>