



Analysis of the Quality of Islamic Religious Education Test Items Based on the SOLO Taxonomy for Developing Summative Assessment in Schools

Nur Fadilah Ramadhany^{1*}, Suwadi², Amri Saputra³

^{1,2,3} Universitas Islam Negeri Sunan Kalijaga

Correspondence e-mail *: 25204011008@student.uin-suka.ac.id

Abstract: Islamic Religious Education (IRE) assessment is still frequently oriented toward the ability to recall information, recognize terms, understand definitions, and determine legal rulings textually. Consequently, IRE test items may measure only basic knowledge rather than depth of understanding, interconceptual relationships, legal reasoning, reflection, and the application of Islamic values in everyday life. This study examines the extent to which IRE test items represent the complexity of students' understanding based on the SOLO Taxonomy framework. Employing a descriptive qualitative approach and content analysis, the study treats IRE test items as assessment texts that represent the cognitive demands placed on students. The data comprised 50 senior high school IRE test items obtained from a school in Batam City. The data were analyzed using Krippendorff's content analysis framework through a critical reading of each item, identification of the structure of its cognitive demands, classification according to SOLO Taxonomy levels, tabulation of frequencies and percentages, and interpretation of assessment-quality patterns. The results show that, of the 50 items analyzed, 35 items (70%) were unistructural, 5 items (10%) were multistructural, 9 items (18%) were relational, and 1 item (2%) was extended abstract, while no prestructural items were identified. These findings indicate that IRE test items remain dominated by basic cognitive demands and therefore do not yet fully represent assessment that promotes interconceptual connections, legal reasoning, reflection, and the application of Islamic values in real-life contexts. This study recommends restructuring the weighting and composition of IRE test items in assessment blueprints by increasing the proportion of relational and extended abstract items. This improvement can be achieved through case-based stimuli, reasoned questions, analyses of interconceptual relationships, attitudinal reflection, and the application of Islamic values to authentic situations, thereby aligning assessment more closely with learning outcomes and principles of meaningful assessment.

Keyword : Islamic Religious Education; question quality; SOLO Taxonomy; summative assessment

Article info: Submitted : 2026-07-03 | Accepted : 2026-08-13 | Published : 2026-08-18

Copyright © 2026, Author.

This is an open-access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)



How to Cite :

Introduction

Assessment is an essential component of Islamic Religious Education because it serves to determine the attainment of learning objectives, the quality of students' understanding, and the effectiveness of the learning process. The quality of learning is strongly influenced by the quality of assessment (Fauzi & Inayati, 2023). In the learning and assessment guidelines for the Merdeka Curriculum, assessment is understood as

an activity for obtaining evidence of the attainment of learning objectives and as part of a learning cycle that provides information for improving the learning process (Mujiburrahman et al., 2023). In the context of Islamic Religious Education, assessment measures not only students' mastery of religious knowledge but also their ability to understand, analyze, present, believe in, and habitually practice Islamic teachings in daily life. Thus, assessment functions not merely as a tool for measuring final outcomes but also as a means of identifying learning needs, monitoring competency development, and evaluating the effectiveness of instructional strategies.

The learning outcomes of Islamic Religious Education emphasize that learning should be directed toward understanding meaning, connecting concepts, developing attitudes, and applying Islamic values in everyday life (Warsah & Habibullah, 2022). This orientation indicates that IRE assessment should measure the depth of students' understanding rather than merely their ability to recall religious information textually. Students need not only to know definitions, scriptural evidence, terminology, or examples of religious behavior, but also to explain relationships among concepts, understand the meaning of teachings, formulate reasons, reflect on values, and apply Islamic teachings in authentic situations. The quality of IRE test items should therefore be evaluated according to the extent to which they present cognitive demands that correspond to IRE learning objectives. In practice, however, IRE assessment items may still be dominated by memorization, term recognition, citation of scriptural evidence, and textual responses (Hidayat & Rohmawati, 2025). Although such items are important for establishing foundational mastery, they are insufficient for representing the depth of students' thinking. Test-item quality is determined not only by the alignment of content with the curriculum but also by the structure of the cognitive demands constructed through the questions. Overly factual assessment may encourage mechanical learning, whereas assessment that requires reasoning, interconceptual connections, argumentation, reflection, and value application is more likely to support meaningful learning (Dhesita & Pinilih, 2025). In this context, the quality of test items is directly related to the quality of learning that emphasizes attitudes and skills.

The quality of test items needs to be analyzed to determine whether they encourage students to progress gradually from simple to more complex levels of thinking. Such analysis is important because the forms of questions posed by teachers influence how students understand the subject matter. Questions that merely ask for definitions, identification, or examples tend to measure surface understanding (Savika & Zuhriyah, 2024). By contrast, questions requiring students to explain causal relationships, compare concepts, analyze cases, construct arguments, or apply Islamic values in new situations are better able to measure conceptual and reflective understanding (Laela & Sa'adah, 2026). Therefore, assessment for learning should be

examined in terms of the structure of the cognitive demands embedded in meaningful IRE test items.

One framework that can be used to analyze test-item quality according to the complexity of understanding is the Structure of the Observed Learning Outcome (SOLO) Taxonomy. The taxonomy comprises five levels: prestructural, unistructural, multistructural, relational, and extended abstract. Through this taxonomy, IRE test items can be mapped according to whether they require a single piece of information, several unconnected pieces of information, relationships among concepts, or the ability to generalize and reflect in a new context. The SOLO Taxonomy is also relevant to the principle of constructive alignment, namely the alignment among learning objectives, learning activities, and assessment. Accordingly, the items developed should accurately reflect the intended level of understanding (Biggs & Collis, 1982).

Studies of test-item quality commonly employ the revised Bloom's Taxonomy or the categories of higher-order thinking skills (HOTS) and lower-order thinking skills (LOTS). The revised Bloom's Taxonomy is useful for identifying dimensions of cognitive processes, including remembering, understanding, applying, analyzing, evaluating, and creating (Krathwohl et al., 1964). HOTS-oriented studies, meanwhile, emphasize analysis, evaluation, creation, reasoning, problem solving, and decision making. Nevertheless, these approaches do not fully reveal the structure and depth of understanding required by a test item. The SOLO Taxonomy offers a different perspective because it considers not only operational verbs but also the complexity of the response structure expected from students (Fernandez & Guzon, 2025). Thus, the SOLO Taxonomy can provide a more detailed account of the depth of thinking required by IRE test items.

Based on this rationale, analyzing the quality of Islamic Religious Education test items according to the levels of the SOLO Taxonomy is important. This study aims not only to classify items into particular levels but also to map the dominant levels of thinking represented in IRE test items. The analysis is expected to generate recommendations for developing IRE assessment that measures not only memorization and basic understanding but also the ability to connect concepts, construct arguments, engage in reflection, and apply Islamic values in daily life.

Methodology

This study employed a descriptive qualitative approach using content analysis. The Islamic Religious Education test items were described and interpreted according to the levels of the SOLO Taxonomy rather than used to test relationships among variables. Each item was analyzed in terms of its cognitive demands and the structure of understanding embedded within it (Krippendorff, 2019). The research data consisted of 50 Islamic Religious Education (PAI) test items drawn from a Grade XI summative assessment document at a senior high school in Batam City. The school

was purposively selected because it serves as a model school, making it a relevant setting for examining the cognitive complexity and quality of assessment items used in Islamic Religious Education. The selection of Grade XI summative assessment items was intended to provide a focused analysis of the extent to which the questions represent different levels of cognitive understanding based on the SOLO Taxonomy. Documents are relevant data sources in qualitative research because they can be systematically analyzed to understand particular content, contexts, and patterns.

Data were collected through document analysis. The document was selected purposively according to several criteria: it contained Islamic Religious Education test items, corresponded to the designated educational level, had a traceable source, identified the document or institution responsible for its preparation, and was relevant to the IRE content under analysis. The researcher located PAI assessment documents, selected the document according to the predetermined criteria, compiled the items, assigned a code to each item, and entered them into a classification sheet. Coding was used to facilitate the analytical process. In this study, the items were coded PAI-S01 through PAI-S50, indicating the subject and item number. For example, PAI-S01 refers to the first PAI item, and the same convention applies to the remaining items.

Data analysis was conducted in several stages. First, all test items were read comprehensively. Second, each item was identified according to its question format, operational verb, content context, and required ability. Third, the items were classified into SOLO Taxonomy levels based on the established indicators. Fourth, the classification results were calculated as frequencies and percentages. Fifth, the distribution of SOLO levels was interpreted to determine whether the test items were dominated by simple cognitive demands or had progressed toward more complex levels of thinking. The results were then used as the basis for formulating recommendations for developing PAI assessment. Data trustworthiness was established through analytical rigor, an audit trail, and expert judgment. Analytical rigor involved repeatedly reading and reviewing each item before assigning its level. The audit trail preserved records of the document source, item codes, classification indicators, rationales for level assignment, and categorization results. Expert judgment involved soliciting feedback from assessment specialists, PAI lecturers, or PAI teachers to evaluate the appropriateness of item classification according to the SOLO Taxonomy. These strategies were employed to reduce subjectivity and strengthen the dependability of the findings.

Result and Discussion

Finding

1. Description of the Research Data

The data consisted of a senior high school Islamic Religious Education assessment document obtained from a school in Batam City. The document was

selected because it contained IRE test items relevant to the focus of the study, namely the analysis of test-item quality according to SOLO Taxonomy levels. A total of 50 items were analyzed and coded PAI-S01 through PAI-S50 to facilitate systematic identification and classification. The content represented in the items covered several areas of Islamic Religious Education, particularly fiqh al-mu'amalah, including sale and purchase, riba, gharar, ijarah, rahn or pawning, and lending and borrowing, as well as fiqh al-munakahat, including marriage, marriage guardianship, divorce, reconciliation after divorce, and mut'ah. Each item was analyzed according to the cognitive demand it imposed rather than merely its sentence form or operational verb. The analysis referred to the five levels of the SOLO Taxonomy: prestructural, unistructural, multistructural, relational, and extended abstract.

Before the 50 PAI test items were classified, operational categories for the SOLO Taxonomy were established as the analytical basis. These categories were used to identify the complexity of the cognitive demands in each item, ranging from items that were conceptually irrelevant to those requiring the application of concepts in new contexts. The operationalization of the five SOLO Taxonomy levels used in this study is presented in the following table.

Table 1. Operationalization of SOLO Taxonomy Levels

SOLO Taxonomy Level	Operational Meaning	Indicators of Cognitive Demand	Characteristic s of IRE Test Items	Classification Decision
Prestructural	The item does not present a relevant demand for understanding or remains conceptually incorrect	The item is ambiguous, inconsistent with the content, illogical, or cannot be answered accurately	The question is unclear, the answer options are confusing, or the Islamic concept used is inaccurate	An item is classified as prestructural when it cannot validly measure students' understanding
Unistructural	The item requires only one simple piece of information	Recognizing a term, stating a definition, determining a ruling, selecting one fact, or recalling one concept	The item asks about the meaning of riba, the legal status of sale and purchase, the definition of marriage, or a particular fiqh term	An item is classified as unistructural when students need only one piece of information to answer it
Multistructural	The item requires	Stating several	The item asks for several	An item is classified as

	several pieces of information but does not require relationships among them	examples, characteristics, types, pillars, conditions, or elements	conditions of sale and purchase, types of riba, pillars of marriage, or forms of mu'amalah	multistructural when the answer consists of several pieces of information that remain separate
Relational	The item requires students to connect several concepts or pieces of information	Explaining reasons, causal relationships, comparisons, meanings, or relationships between concepts and behavior	The item asks why riba is prohibited, how honesty relates to transactions, or how valid and invalid sale and purchase differ	An item is classified as relational when students must integrate concepts, reasons, and context into a coherent whole
Extended Abstract	The item requires the reflective or argumentative application of concepts to a new context	Analyzing a case, proposing a solution, making a judgment, constructing an argument, or applying Islamic values in real life	The item presents a case involving commercial practices, online lending, family conflict, or a contemporary mu'amalah issue that must be addressed using Islamic reasoning	An item is classified as extended abstract when students must transfer Islamic concepts to a new situation and provide a judgment or solution

2. Distribution of SOLO Taxonomy Levels in IRE Test Items

The analysis of the 50 items showed that the PAI test items were distributed across four SOLO Taxonomy levels: unistructural, multistructural, relational, and extended abstract. No items were identified at the prestructural level. The distribution is presented in the following table.

Table 2. Distribution of Test Items by SOLO Taxonomy Level

SOLO Taxonomy Level	Number of Items	Percentage
Prestructural	0	0%
Unistructural	35	70%
Multistructural	5	10%
Relational	9	18%
Extended Abstract	1	2%

As shown in the table, the unistructural level was the most dominant category, comprising 35 items or 70%. The multistructural level comprised 5 items or 10%, while the relational level comprised 9 items or 18%. Only 1 item or 2% was classified at the extended abstract level. The absence of prestructural items indicates that all analyzed items retained content, contexts, and answer expectations relevant to IRE learning.

3. Classification of Test Items Based on the SOLO Taxonomy

The classification of test items based on the SOLO Taxonomy was conducted to determine the complexity of the cognitive demands embedded in each PAI test item. This process is important because item quality cannot be adequately evaluated solely through answer formats or operational verbs; it must also be examined according to the thinking students are required to demonstrate. Each item was therefore coded PAI-S01 through PAI-S50 to make data identification, tracking, and grouping more systematic. After coding, each item was analyzed and categorized according to its cognitive demand as unistructural, multistructural, relational, or extended abstract. These levels were assigned to determine whether the items predominantly measured basic knowledge or had progressed toward the ability to connect concepts, analyze cases, and apply Islamic values in real-life contexts.

Table 3. Classification of Test Item

SOLO Taxonomy Level	Item Codes	Cognitive Demand	Classification Rationale
Prestructural	0	0%	0
Unistructural	PAI-S01-03, S05, S07-08, S11-12, S14, S16-18, S20-29, S31-32, S34, S36-38, S40, S42, S44-47, S49-50	Recognizing one piece of basic information	The items require only a definition, legal ruling, scriptural basis, or one particular wisdom

Multistructural	PAI-S15, S30, S39, S43, S48	Recognizing several pieces of information	The items require students to identify several conditions or pillars
Relational	PAI-S04, S06, S10, S13, S19, S33-35, S41	Connecting a case with a concept	The items require students to connect a case with a provision of Islamic law
Extended Abstract	PAI-S09	Proposing a solution in a new context	The item requires the application of Islamic principles to online transactions

The table shows that most PAI test items were classified at the unistructural level. These items primarily required students to recognize a single piece of basic information, such as a definition, legal ruling, scriptural basis, or particular wisdom. A smaller number were classified as multistructural because they required students to recognize several pieces of information, such as multiple conditions or pillars. Relational items began to require students to connect a case with a concept or a provision of Islamic law. The extended abstract level was represented by only one item, PAI-S09, because it required students to propose a solution in a new context, specifically by applying Islamic principles to online transactions. These findings demonstrate that the weighting of PAI test items remains concentrated on basic knowledge, while items requiring reasoning, interconceptual relationships, and the application of Islamic values remain limited.

Discussion

1. The Dominance of Unistructural PAI Test Items

The findings show that the quality of Islamic Religious Education test items analyzed using the SOLO Taxonomy was dominated by the unistructural level. Of the 50 items analyzed, 35 items or 70% were unistructural, 5 items or 10% were multistructural, 9 items or 18% were relational, and 1 item or 2% was extended abstract; no items were classified as prestructural. This distribution indicates that most PAI items still required students to recognize a single principal piece of information, such as a definition, ruling, term, pillar, condition, scriptural basis, or particular example. Thus, the overall pattern of the items did not yet fully support the measurement of complex structures of understanding.

The dominance of unistructural items is consistent with patterns identified in previous studies. PAI assessments have tended to emphasize mastery of basic knowledge, particularly remembering and understanding, while applying, analyzing, evaluating, and creating remain infrequently represented (Nurhaliza & Fiqri Mustopa, 2023; Savin et al., 2023; Winget & Persky, 2022). Other findings similarly indicate that PAI assessment frequently focuses on content mastery and the memorization of Qur'anic verses, hadith, and religious concepts, while the internalization and practice of values receive less balanced attention. The present findings therefore reinforce the broader pattern that PAI assessment remains oriented toward mastery of religious information rather than depth of understanding and the application of values in real-life contexts (Amaly et al., 2023; Aripin et al., 2025).

Nevertheless, unistructural items retain an important function because students need to master foundational concepts before engaging in more complex thinking (Agustinsa et al., 2021; Mohanna et al., 2023). Items concerning the meaning of sale and purchase, the pillars of marriage, types of *riba*, the legal status of a contract, or the definition of *ijarah*, for example, can help ensure that students understand basic terms and concepts in *fiqh*. However, when the item composition is excessively dominated by this level, assessment risks measuring only the ability to recognize information rather than the ability to understand conceptual relationships, analyze cases, or make decisions based on Islamic principles. Within the SOLO Taxonomy, the unistructural level directs students to only one aspect of information, and therefore requires a limited depth of understanding (Adeniji et al., 2022; Biggs & Collis, 1982; Rahmat & Jamil, 2024). A systematic review of the SOLO model also confirms its usefulness for distinguishing surface understanding from more meaningful forms of understanding (Adeniji et al., 2022). These findings also reveal an imbalance between assessment demands and the characteristics of PAI learning. The learning outcomes for Islamic Religious Education and Character Education are directed not only toward knowing Islamic teachings but also toward analyzing, presenting, believing in, habituating, and applying Islamic values in life (Hidayati & Nurdi, 2023). At Phase F, corresponding to Grade 12, for example, students are expected to analyze Qur'anic and hadith teachings related to critical thinking, science, tolerance, human life, patriotism, and religious moderation, while developing critical, creative, adaptive, and responsible attitudes. IRE assessment should therefore extend beyond factual questions and examine students' ability to interpret problems, connect Islamic concepts with reality, and provide appropriate ethical or legal considerations.

The strong concentration of items at the unistructural level may also be understood in relation to structural conditions surrounding test-item development. Since the present study analyzed assessment documents rather than interviewing the teachers who developed the questions, these factors should be treated as plausible explanations rather than empirically verified causes. One possible factor is the limited

time available for preparing summative assessments. Recall-oriented multiple-choice items are relatively easier to construct, review, score, and use to cover a wide range of curriculum content within a single examination. Another possible factor is the continued reliance on conventional examination patterns, textbooks, previous examination questions, and existing item banks. When such materials predominantly contain questions on definitions, terminology, legal rulings, and factual knowledge, the same pattern can be reproduced across subsequent assessments. A further structural consideration concerns teachers' assessment literacy. Developing relational and extended abstract items requires more than replacing operational verbs. Teachers need to construct meaningful stimuli, formulate authentic problems, establish relationships among several concepts, and ensure that answer options genuinely discriminate between different structures of understanding. Limited exposure to specific training in SOLO Taxonomy-based item development may therefore make conventional recall questions more familiar and practical to produce. These possible conditions help explain why curricular expectations emphasizing analysis and application do not automatically result in assessments with equally complex cognitive demands. This imbalance is particularly important because IRE and Character Education learning outcomes are not limited to knowing Islamic concepts. They also include analyzing teachings, presenting arguments, developing attitudes, habituating values, and applying Islamic principles in everyday life (Hidayati & Nurdi, 2023). In Phase F, which includes Grade XI, students are expected to engage with religious teachings through increasingly critical, contextual, adaptive, and responsible forms of understanding. Assessment should therefore provide opportunities not only to recall religious information but also to interpret problems, relate Islamic principles to contemporary situations, and formulate appropriate ethical and legal judgments.

2. Assessment for Learning and the Promotion of Interconceptual Connections

The fact that multistructural items accounted for only 10% indicates that a small proportion of the items contained more than one element of information but did not require students to integrate these elements into a coherent understanding. At this level, students are generally asked to identify several characteristics, conditions, or categories (Biggs & Collis, 1982), such as the requirements for pledged property, types of divorce, or several provisions of a contract. Such items are more complex than unistructural items because they involve multiple pieces of information. However, their cognitive demand does not yet require students to connect the information in order to solve a problem. In other words, multistructural items represent collections of information rather than analyses of relationships. The presence of this level therefore demonstrates some variation in cognitive demands, but the variation is not sufficiently strong to characterize assessment that promotes deep understanding.

Meanwhile, the presence of 9 relational items or 18% indicates that some IRE test items had begun to measure more complex thinking. Relational items are generally characterized by a case, context, or particular situation that requires students to connect fiqh concepts with principles of Islamic law (Supriadi et al., 2025). Examples include items concerning the sale of stolen goods, concealment of product defects, an ijarah contract with an unclear object of benefit, or an additional amount imposed on a loan. Such items cannot be answered merely by recalling terminology. They require an understanding of the relationships among the case, relevant concepts, the conditions for a valid contract, justice, mutual consent, and prohibitions against gharar, tadlis, or riba.

A comparison of the subject-matter domains represented in the analyzed items also reveals an important pattern. Higher-level questions were more clearly visible in fiqh al-mu'amalah than in fiqh al-munakahat. Several relational items were based on transaction-related problems, including stolen goods, hidden product defects, unclear benefits in ijarah, and additional charges on loans. More importantly, the only extended abstract item in the entire assessment also emerged from the mu'amalah domain, asking students to explain how Islamic principles can prevent fraud in online transactions.

In contrast, the munakahat items represented in the assessment were more frequently structured around the identification of established concepts, such as the pillars of marriage, elements of marriage law, or classifications related to divorce. Consequently, these questions tended to remain at the unistructural or multistructural levels. This comparison suggests that, within the assessment analyzed, mu'amalah contributed more visibly to higher-level cognitive demands than munakahat.

The difference should not be interpreted as evidence that the munakahat domain is inherently less suitable for relational or extended abstract assessment. Rather, it appears to reflect differences in item construction. Mu'amalah is readily connected to contemporary situations familiar to students, such as e-commerce, online lending, defective products, digital transactions, consumer protection, fraud, riba, and contractual uncertainty. These topics naturally provide contextual stimuli that require students to connect Islamic legal concepts with authentic problems.

Munakahat has similar potential when questions are constructed around cases rather than factual identification. Topics such as consent in marriage, the role of guardians, marital rights and responsibilities, family conflict, divorce, reconciliation, and ethical decision-making can require students to integrate normative principles with social circumstances. Therefore, the comparatively lower structural complexity of munakahat questions in the analyzed assessment should be understood as a consequence of how the questions were formulated rather than as a limitation inherent in the content itself.

This finding has an important implication for item development. Increasing the number of higher-level questions does not necessarily require introducing new subject matter. Existing topics can be transformed by changing the structure of the task. A question asking students to identify the pillars of marriage, for example, primarily measures the recall of several components. The same topic can become relational when students are asked to analyze why the presence or absence of a particular element affects the validity of a marriage. It can move toward extended abstract reasoning when students are required to apply relevant principles to a new social context and formulate a reasoned ethical or legal response.

From the perspective of the SOLO Taxonomy, relational items better represent the ability to understand a conceptual structure as an integrated whole because students must connect several pieces of information to determine an appropriate answer (Ariani et al., 2026; Biggs & Collis, 1982; Yuanita et al., 2023). Research on the development of SOLO-based instruments indicates that the relational level requires students to connect information within a question to produce more complete and meaningful responses (Yuanita et al., 2023). However, the proportion of relational items, at only 18%, indicates that the assessment orientation was not yet primarily directed toward case analysis. PAI content, particularly *fiqh al-mu'amalah* and *fiqh al-munakahat*, has considerable potential for the development of relational items because it is directly connected to social life (Wahyuni, 2023). Topics such as sale and purchase, *riba*, *ijarah*, *rahn*, lending and borrowing, marriage, divorce, and reconciliation after divorce contain not only normative concepts but also require students to consider context, understand legal rationales, distinguish valid from invalid contracts, and evaluate the moral and social consequences of actions. The proportion of relational items should therefore be increased so that PAI assessment not only tests the recall of legal rulings but also develops students' ability to reason about their foundations and apply them to authentic cases (Chae & Alexander, 2021).

3. Proportion of Reflective and Applied Items in Real-Life Contexts

The most notable finding is the extremely limited number of extended abstract items, with only 1 item or 2%. This is the highest level of the SOLO Taxonomy because it requires students to generalize, reflect, construct arguments, or apply principles in new contexts (Cao et al., 2023). In this study, the item classified at the extended abstract level asked students to explain how Islam prevents fraud in online transactions. Such an item requires students not only to know the legal principles of sale and purchase but also to connect them with contemporary commercial practices and the principles of honesty, transparency, mutual consent, justice, and the prevention of harm. The small number of extended abstract items indicates that PAI assessment provides limited opportunities for students to engage reflectively and practically with contemporary issues (Widada et al., 2020).

The limited number of extended abstract items requires attention because PAI learning is fundamentally value-oriented rather than merely concerned with factual cognition (Batubara & Matsum, 2025). Ideally, PAI assessment should ask not only “What is the legal ruling of a contract?” but also “Why was the ruling established?”, “How is the principle applied?”, and “What are its implications for justice, public benefit, and social responsibility?” Recent studies of HOTS-oriented assessment in Islamic Religious Education indicate that assessment emphasizing higher-order thinking can foster contextual reasoning, analytical judgment, evaluative reflection, and creative synthesis in the domains of creed, fiqh, and Qur’an and Hadith studies (Amalina et al., 2026; Rahmah et al., 2026).

The findings also strengthen the argument that PAI assessment requires an analytical framework that is not based solely on operational verbs. Test-item quality has commonly been analyzed using Bloom’s Taxonomy or LOTS and HOTS classifications (Syawal et al., 2025). Although these frameworks are important, they are not always sufficient for identifying the structure of understanding required by an item. The SOLO Taxonomy offers a different perspective by evaluating item quality according to the complexity of the thinking structure, ranging from one piece of information and several pieces of information to interconceptual relationships and generalization in new contexts (Malasari et al., 2023). The use of SOLO in this study demonstrates that an item containing contextual information is not necessarily at a high level when students are required only to recognize one concept. Conversely, a multiple-choice item may be relational or extended abstract when its stimulus and answer options require conceptual connections, reasoning, and the application of principles (Zahra & At-Taqiyyah, 2024).

Methodologically, the use of frequencies and percentages in this study does not transform it into quantitative research. Percentages were used as descriptive aids to identify data patterns, while the core analysis remained focused on interpreting item content, cognitive demands, and classification rationales. This approach is consistent with qualitative content analysis, in which simple tabulation may be used to clarify patterns while the principal conclusions remain grounded in interpretations of meaning and content categories. Presenting percentages therefore strengthens the transparency of the findings by enabling readers to observe the dominance of particular levels among the analyzed items.

Based on these findings, the PAI test items were relevant in terms of content but imbalanced in terms of cognitive complexity. The absence of prestructural items indicates that the items generally remained within the scope of PAI content and contained demands that could be classified. However, the dominance of unistructural items indicates that the assessment measured basic understanding more frequently than deep understanding. The principal weakness therefore lies not in content

irrelevance but in the limited variation of cognitive levels requiring conceptual connections, case analysis, reflection, and value application in real-life contexts.

4. Implications of the Findings for the Quality of PAI Assessment

The findings suggest that improving PAI assessment requires more than increasing the numerical proportion of relational and extended abstract items. The more fundamental task is to redesign the alignment among learning objectives, assessment indicators, stimuli, question stems, and expected responses. The SOLO Taxonomy can support this process because each level represents a different structure of understanding rather than merely a different operational verb. Thus, teachers should first determine the structure of understanding expected from students before constructing the item itself (Ariani et al., 2026; Biggs & Collis, 1982). Research on SOLO-based assessment similarly demonstrates that translating the SOLO hierarchy into domain-specific criteria aligned with intended learning outcomes can facilitate progression from fragmented knowledge toward relational and extended abstract reasoning (Ariani et al., 2026).

At the planning stage, teachers can begin by reformulating competency indicators or IPK, or, within the Kurikulum Merdeka framework, the relevant learning objectives (TP), so that they explicitly indicate the expected cognitive relationship. An indicator such as “students can identify the meaning of *riba*” naturally leads to a unistructural item because it requires only one piece of information. To target the relational level, the indicator can be reformulated as “students can analyze the relationship between the characteristics of a loan transaction and the Islamic prohibition of *riba*.” For the extended abstract level, it can be further developed into “students can formulate an alternative solution to a contemporary financial transaction based on Islamic principles of justice, mutual consent, and the prohibition of *riba*.” This alignment is consistent with curriculum guidance that learning objectives should specify the competence students are expected to demonstrate and the level of thinking involved. The current direction of PAI curriculum policy also emphasizes a better balance between knowledge and the application of Islamic teachings in everyday life.

Operationally, the progression can be mapped as follows.

Table 4.
Developing Islamic Religious Education Questions Based on the SOLO Taxonomy

Component	Unistructural	Relational	Extended Abstract
Intended understanding	Recognize one relevant concept	Connect several concepts to explain or solve a case	Transfer, generalize, evaluate, or formulate a

			solution in a new context
Indicator/ TP	Identify the meaning of riba	Analyze why a particular loan arrangement constitutes riba by relating its characteristics to Islamic principles	Formulate a Sharia-compliant alternative to a contemporary loan practice and justify the proposal
Stimulus	Definition or direct factual information	Short vignette containing a realistic problem and several relevant facts	Novel or more complex real-life scenario requiring transfer of principles
Question stem	"What is...?" or "Which of the following is...?"	"Based on the case, why...?", "Which conclusion is most appropriate...?", or "How are ... related?"	"What solution should be proposed...?", "How can these principles be applied to...?", or "Evaluate and justify..."
Expected response	One correct fact or concept	Integrated reasoning based on several relevant concepts	Generalization, reasoned judgment, or application to a new situation

The second operational step concerns restructuring the stimulus. Merely adding a long paragraph, photograph, Qur'anic verse, or contemporary terminology does not automatically transform an item into a higher-level question. A stimulus functions cognitively only when students must use the information it contains to arrive at the answer. For relational items, teachers can construct a short vignette containing an actor, a realistic situation, and two or more relevant facts that must be connected with Islamic principles. For example, rather than asking students to define *tadlis*, the item may describe an online seller who knows that a product is defective but deliberately conceals the defect from the buyer. Students are then asked to determine the problem in the transaction by connecting honesty, disclosure, mutual consent, and the prohibition of deception. Authentic scenarios are particularly useful because they require students to apply knowledge to situations resembling problems encountered outside the classroom (Vlachopoulos & Makri, 2024). In PAI, contextual learning

involving socio-religious cases, digital ethics, and contemporary social problems has also been shown to provide opportunities for analytical and reflective reasoning rather than normative recall (Nashir et al., 2025).

The third step is to restructure the question stem. Moving from unistructural to relational thinking requires changing what students are required to do with information, not simply replacing an easy verb with a more difficult one. For example, the question “What is the definition of gharar?” asks students to retrieve one concept. A relational version could present an online transaction in which the specifications and delivery of a product are unclear and ask, “Why can the transaction be categorized as containing gharar?” Students must then connect the characteristics of the case with the underlying legal principle. To reach the extended abstract level, the stem can require transfer to a new context, for example: “How should an online marketplace redesign this transaction to avoid gharar while protecting both sellers and buyers? Justify your proposal using relevant Islamic principles.” The cognitive progression therefore moves from recognizing a concept, to connecting the concept with the facts of a case, and finally to using the concept to generate a reasoned response in a new situation.

The same procedure can be applied to *fiqh al-munakahat*. A conventional question such as “What are the pillars of marriage?” primarily elicits several factual elements. A relational item can present a vignette in which one required element is absent and ask students to analyze its implications for the validity of the marriage by relating the facts of the case to relevant legal principles. An extended abstract item may present a contemporary family problem involving consent, marital responsibilities, conflict, or reconciliation and require students to propose an ethically and legally reasoned response based on Islamic principles. Therefore, higher-level assessment does not depend on selecting inherently more complex subject matter; it depends on restructuring the relationship among the stimulus, the question, and the knowledge students must mobilize.

For multiple-choice items, cognitive complexity should also be reflected in the answer options. Relational questions should not contain one obviously correct answer that can be identified through a keyword in the stimulus. Instead, distractors can represent partial understanding, incorrect connections between concepts, or plausible but incomplete legal reasoning. The correct option should require students to integrate all relevant information in the vignette. Extended abstract demands may be more appropriately assessed through reasoned multiple-choice items, short constructed responses, case analyses, or written reflections when students need to justify, generalize, or propose solutions. The development and validation of HOTS-oriented multiple-choice and essay items demonstrates that both formats can be designed to

assess higher-order thinking when the cognitive demand is intentionally embedded in the task (Maryani et al., 2021)

A practical item-development sequence for PAI teachers can therefore consist of five checks: (1) identify the learning outcome or TP and determine the intended SOLO level; (2) select a real-life or socio-religious context relevant to the concept; (3) construct a stimulus containing information that must actually be processed; (4) formulate a stem requiring connection, judgment, transfer, or solution rather than simple retrieval; and (5) review the response options or scoring criteria to ensure that students cannot answer correctly through factual recognition alone. A SOLO-based blueprint can document these components before test administration so that teachers can examine whether the overall assessment provides an intentional progression from foundational to complex understanding (Ariani et al., 2026).

These procedures also provide a more concrete direction for professional development through MGMP PAI. Workshops should not be limited to introducing the definitions of the five SOLO levels. Teachers can collaboratively select existing unistructural items from their assessment banks and conduct an item transformation **exercise** in which each item is redesigned into relational and extended abstract versions. Peer-review sessions can then examine whether the revised stimulus is functional, whether the stem genuinely requires interconceptual reasoning, whether the distractors represent meaningful misconceptions, and whether the item remains aligned with the intended learning objective. Strengthening this form of assessment literacy is important because teachers require not only theoretical knowledge of assessment but also the capacity to translate assessment principles into actual classroom tasks (Pastore, 2023).

Thus, the practical shift from lower to higher SOLO levels can be understood as a transformation of the structure of the task rather than simply an increase in question difficulty. A unistructural item asks students to retrieve relevant knowledge; a relational item requires them to organize and connect knowledge to explain a particular situation; and an extended abstract item requires them to transfer that integrated understanding to evaluate, generalize, or formulate a response to a new problem. For PAI assessment, this progression is especially important because the subject is intended not only to develop knowledge of Islamic teachings but also to support students in reasoning about and applying those teachings within changing social realities. Accordingly, a systematic alignment of learning objectives, contextual stimuli, question stems, and SOLO levels can make PAI assessment more meaningful, reflective, and applicable to real-life contexts.

Conclusion

This study demonstrates that the quality of Islamic Religious Education test items analyzed using the SOLO Taxonomy does not yet fully reflect a balance between

mastery of basic knowledge and complex thinking skills. The dominance of unistructural items indicates that IRE assessment more frequently requires students to recognize a single piece of information, while the limited number of relational and extended abstract items indicates insufficient emphasis on connecting concepts, analyzing cases, constructing arguments, and applying Islamic values in real-life contexts. The principal contribution of this study lies in the use of the SOLO Taxonomy as a framework for evaluating the structural complexity of understanding required by each test item, offering an alternative lens for distinguishing items that measure recall, multiple pieces of information, conceptual relationships, reflection, and application. Practically, these findings highlight the need for MGMP PAI to organize workshops or training on developing SOLO Taxonomy-based assessment items so that teachers can design more balanced, contextual, and higher-level questions. Accordingly, the development of IRE test items should increase the proportion of relational and extended abstract items while maintaining appropriate items for foundational knowledge. Future research should broaden the data sources across schools, educational levels, and assessment types, and may also analyze students' responses to examine the correspondence between item demands and the level of understanding actually demonstrated.

References

- Adeniji, S. M., Baker, P., & Schmude, M. (2022). Structure of the Observed Learning Outcomes (SOLO) model: A Mixed- Method Systematic Review of Research in Mathematics Education. *Eurasia Journal of Mathematics, Science and Technology Education*, 18(6), 1–17. <https://doi.org/10.29333/ejmste/12087>
- Agustinsa, R., Muchlis, E. E., & Maizora, S. (2021). Level of Student Understanding in Solving Geometry Problems based on Taxonomy of SOLO (Structure of Observed Learning Outcomes). *Journal of Physics: Conference Series*, 1731(1), 012055. <https://doi.org/10.1088/1742-6596/1731/1/012055>
- Amalina, N., Rahman, A., Djollong, A. F., Kharis, M., Rondhiyah, S., Wahyuni, M., Parepare, U. M., Info, S., Strategies, C. L., & Education, I. R. (2026). *Pengembangan Strategi Pembelajaran Kognitif Berbasis Hots (Higher Order Thinking Skills) pada Mata Pelajaran Pendidikan Agama Islam*. 11(1), 1174–1182.
- Amaly, A. M., Herdiana, Y., Ruswandi, U., & Arifin, B. S. (2023). The Necessity and Reality of Islamic Religious Education in Schools. *Jurnal Ilmiah Islam Futura*, 23(1), 1. <https://doi.org/10.22373/jiif.v23i1.13190>
- Ariani, F., Suyanto, S., Yuliana, L., Asriyah, A., & Rahmah, L. (2026). From Taxonomies to Rubrics: SOLO-based Formative Assessment to Improve Understanding of Population Dynamics. *International Journal of Learning, Teaching and Educational Research*, 25(1), 518–535. <https://doi.org/10.26803/ijlter.25.1.25>
- Aripin, S., Nurdiansyah, N. M., Arief, A., & Yusof, N. binti. (2025). Effectiveness and Journal of Education and Teacher Training Innovation, 2026, 4(2), 887 - 907

- Reformulation of Islamic Religious Education in Schools in the Modern Era. *Jurnal Pendidikan Agama Islam*, 22(1), 215–234. <https://doi.org/10.14421/jpai.v22i1.7432>
- Batubara, A. U., & Matsum, H. (2025). Assessment Patterns in Islamic Education Learning on Fiqh Material. *Halaqa: Islamic Education Journal*, 9(1), 165–184. <https://doi.org/10.21070/halaqa.v9i1.1733>
- Biggs, Jo. B., & Collis, K. F. (1982). *Evaluating The Quality of Learning The SOLO Taxonomy (Structure of The Observed Learning Outcome)* (Vol. 2). A Subsidiary of Harcourt Brace Jovanovich London.
- Cao, J., Fang, Z., & Yu, S. (2023). Automatic Evaluation of Chinese Short-Answer Questions Based on Structure of Observed Learning Outcomes Taxonomy. 2023 *IEEE International Conference on Advanced Learning Technologies (ICALT)*, 315–317. <https://doi.org/10.1109/ICALT58122.2023.00098>
- Chae, S. E., & Alexander, P. A. (2021). The Development of Relational Reasoning in South Korean Elementary and Middle-School Students: A Cross-Sectional Investigation. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.630609>
- Dhesita, S. J., & Pinilih, F. W. (2025). Asesmen Bermakna Humanis untuk Mendorong Otonomi Belajar Siswa Madrasah Aliyah. 1(2), 91–105.
- Fauzi, A., & Inayati, N. L. (2023). Implementasi Evaluasi Pembelajaran Pendidikan Al Islam di Sekolah Menengah Atas Muhammadiyah. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, 4(2), 272–283. <https://doi.org/10.31538/munaddhomah.v4i2.438>
- Fernandez, P. J. M., & Guzon, A. F. H. (2025). A SOLO Taxonomy-based Rubric for Assessing Conceptual Understanding in Applied Calculus. *Journal on Mathematics Education*, 16(2), 559–580. <https://doi.org/10.22342/jme.v16i2.pp559-580>
- Hidayat, R., & Rohmawati, B. (2025). Pengembangan Butir Soal HOTS: Tantangan dan Solusi untuk Meningkatkan Kemampuan Analisis Siswa Pada Pembelajaran PAI. *JUMPER: Journal of Educational Multidisciplinary Research*, 4(1), 1–20. <https://doi.org/10.56921/jumper.v4i1.255>
- Hidayati, Z., & Nurdi, N. (2023). Persepsi Guru Terhadap Implementasi Kurikulum Merdeka dalam Pembelajaran PAI dan Budi Pekerti di Sekolah Dasar. *Jurnal Cendekia Media Komunikasi Penelitian Dan Pengembangan Pendidikan Islam*, 15(01), 30–41. <https://doi.org/10.37850/cendekia.v15i01.442>
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of Educational Objectives Affective Domain. In *Educational Objectives* (Vol. 93, Number 1979). Longman.
- Krippendorff, K. (2019). *Content Analysis an Introduction to Its Methodology*. SAGE Publications Ltd.

- Laela, N., & Sa'adah, K. (2026). Eksplorasi Pengembangan Dimensi Konseptual untuk Meningkatkan Efektivitas Pembelajaran SKI Kelas VIII di MTs Syi'ar Islam Maibit. *Jurnal Manajemen Dan Pendidikan Agama Islam*, 4(1), 181–193. <https://doi.org/10.61132/jmpai.v4i1.1758>
- Malasari, P. N., Sari, A., Haqiqi, A. K., Masrukhin, & Qomariyah, S. (2023). *Using the SOLO Taxonomy to Investigate Students' Mathematical Creative Thinking Proficiency*. 060009. <https://doi.org/10.1063/5.0123686>
- Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitriawanawati, M. (2021). HOTs Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-order Thinking Skills of Prospective Teachers. *Journal of Turkish Science Education*, 18(4), 674–690. <https://doi.org/10.36681/tused.2021.97>
- Mohanna, K., Wall, D., Cottrell, E., & Chambers, R. (2023). The Theory behind the Practical Aspects of Teaching and Learning. In *Teaching Made Easy* (pp. 87–102). CRC Press. <https://doi.org/10.1201/9781003352532-9>
- Mujiburrahman, M., Kartiani, B. S., & Parhanuddin, L. (2023). Asesmen Pembelajaran Sekolah Dasar dalam Kurikulum Merdeka. *Pena Anda: Jurnal Pendidikan Sekolah Dasar*, 1(1), 39–48. <https://doi.org/10.33830/penaanda.v1i1.5019>
- Nashir, M. J., Wardani, A., Intan Apriliani, E., & Zahiri, M. (2025). HOTS-Based Learning Methodology Innovation in Islamic Religious Education: A Case Study at State Senior High School 1 Karanganyar. *Jurnal Tarbiyatuna*, 16(2), 235–256. <https://doi.org/10.31603/tarbiyatuna.v16i2.14987>
- Nurhaliza, A. S., & Fiqri Mustopa, M. (2023). Analysis of the Cognitive Level of School Exam Questions in Islamic Religious Education Based on the 2001 Revised Bloom'S Taxonomy. *Journal of Research and Thought on Islamic Education (JRTIE)*, 6(1), 71–84.
- Pastore, S. (2023). Teacher Assessment Literacy: a Systematic Review. *Frontiers in Education*, 8. <https://doi.org/10.3389/feduc.2023.1217167>
- Rahmah, S., Ramlan, Wardhani, N., Alwiyah, S., & Syarboini. (2026). Implementing Higher-Order Thinking Skills Oriented Assessment in Islamic Religious Education. *Al-Hayat: Journal of Islamic Education*, 10(1), 237–253. <https://doi.org/10.35723/ajie.v10i1.152>
- Rahmat, A., & Jamil, M. W. (2024). Retention-Based Learning: an Approach to Maximizing Student Learning Outcomes in High School Plant Anatomy Lesson. *Malaysian Journal of Learning and Instruction*, 21(2), 49–78. <https://doi.org/10.32890/mjli2024.21.2.3>
- Savika, H. I., & Zuhriyah, I. A. (2024). Peran Analisis Butir Soal Terhadap Kualitas Soal, Kompetensi Guru, dan Prestasi Belajar Peserta Didik di Sekolah Dasar. *PANDU: Jurnal Pendidikan Anak Dan Pendidikan Umum*, 2(2), 43–51. <https://doi.org/10.59966/pandu.v2i2.856>
- Savin, M., Bordelon, C., Moss, C., & Baker, L. (2023). Using Active Learning to Evaluate

- Student Competency Beyond Clinical Skills. *The Journal for Nurse Practitioners*, 19(5), 104596. <https://doi.org/10.1016/j.nurpra.2023.104596>
- Supriadi, Sahri, & Hasmy, A. (2025). Pengembangan Instrumen Penilaian Pembelajaran Pai Dengan Pendekatan Higher Order Thinking Skills (Hots) Pada Siswa Tingkat Dasar. *Jurnal Pendidikan Dan Pengajaran*, 6(3), 119–135. <https://ijurnal.com/1/index.php/jpp>
- Syawal, M. K.-K., Maulana, I., Nasution, S. K., & Nurmawati. (2025). Analisis Butir Soal Sumatif Akhir Semester Genap PAI Berdasarkan Taksonomi Bloom di Kelas VI SD Islam Al-Abid. *BADA'A: Jurnal Ilmiah Pendidikan Dasar*, 7(1), 153–163. <https://doi.org/10.37216/badaa.v7i1.2161>
- 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>
- Wahyuni, A. S. (2023). Implementasi Autentik Assessment pada Pembelajaran Fikih di Madrasah Tsanawiyah. *Tarqiyatuna: Jurnal Pendidikan Agama Islam Dan Madrasah Ibtidaiyah*, 2(1), 48–58. <https://doi.org/10.36769/tarqiyatuna.v2i1.323>
- Warsah, I., & Habibullah, H. (2022). Implementasi Evaluasi Hasil Belajar Pendidikan Agama Islam di Madrasah. *Journal of Education and Instruction (JOEAI)*, 5(1), 213–225. <https://doi.org/10.31539/joeai.v5i1.3595>
- Widada, W., Herawaty, D., Jumri, R., Wulandari, H., Aliza, F., & Anggoro, A. F. D. (2020). Students of the Extended Abstract in Proving Lobachevsky's Parallel Lines Theorem. *Journal of Physics: Conference Series*, 1470(1), 012098. <https://doi.org/10.1088/1742-6596/1470/1/012098>
- Winget, M., & Persky, A. M. (2022). A Practical Review of Mastery Learning. *American Journal of Pharmaceutical Education*, 86(10), ajpe8906. <https://doi.org/10.5688/ajpe8906>
- Yuanita, P., Maimunah, M., Kartini, K., & Arif, N. (2023). Analysis of SOLO taxonomy-based Test Instruments in Approaching the Merdeka Curriculum. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, 12–29. <https://doi.org/10.23917/jramathedu.v8i1.1270>
- Zahra, A. A., & At-Taqiyyah, A. K. (2024). Mengukur Kemampuan Kognitif Tingkat Tinggi dengan Pertanyaan Pilihan Ganda. *Jurnal Pelita Nusantara*, 1(4), 462–470. <https://doi.org/10.59996/jurnalpelitanusantara.v1i4.336>