

**THE QUALITY OF ISLAMIC EDUCATION LEARNING  
ASSESSMENT: AN EVALUATIVE STUDY AT THE ELEMENTARY,  
JUNIOR HIGH, AND HIGHER EDUCATION LEVELS**

**Ahwil Lutan Hidayah<sup>1</sup>, Sofwatun Nida<sup>2</sup>, Arif Zefrizen<sup>3</sup>**

<sup>123</sup>Universitas Islam Negeri Sunan Kalijaga Yogyakarta  
[23204012003@student.uin-suka.ac.id](mailto:23204012003@student.uin-suka.ac.id)<sup>1</sup>, [23204012014@student.uin-suka.ac.id](mailto:23204012014@student.uin-suka.ac.id)<sup>2</sup>,  
[23204011044@student.uin-suka.ac.id](mailto:23204011044@student.uin-suka.ac.id)<sup>3</sup>

**ABSTRACT**

A quality assessment instrument is an important component in the learning evaluation process to measure the achievement of student learning outcomes. This article aims to describe how to analyze the quality of Islamic Religious Education (PAI) learning assessment instruments at the primary, secondary, and tertiary education levels. The method used is a literature study with a descriptive-analytic approach to the aspects of validity, reliability, difficulty level, differentiating power, and effectiveness of distractors in assessment instruments. The results of the analysis show that the quality of assessment instruments is often not optimal. At the primary level, aspects of suitability with learning indicators and materials are the main focus. At the intermediate level, the distribution of question difficulty levels is not balanced, with the dominance of questions that are too easy. At the high level, there is a need to clarify the substance of the material and avoid redundancy of questions. An ideal assessment instrument must meet the principles of validity, reliability and fairness and be prepared according to learning needs. This article emphasizes the importance of systematic development and analysis of assessment instruments to ensure the quality of accurate and representative learning evaluation results. Thus, efforts are needed to improve educators' competence in developing and analyzing assessment instruments according to educational evaluation standards.

**Keyword:** Instrument Quality, Learning Assessment, Islamic Religious Education

**INTRODUCTION**

Assessment is a fundamental component of education because it can measure the success of the learning process and outcomes (Alkhairi et al., 2024). In addition, assessment can also be used as a preliminary evaluation to determine future strategies for improving student learning outcomes. Assessment is defined as a mechanism for collecting data and information on student learning outcomes through observation, interviews, recording, and other methods.

The evaluation includes the preparation of instruments, which cannot be done carelessly or haphazardly. If the evaluation instruments are wrong, the evaluation results will automatically be wrong as well (Nida et al., 2025). Jannah quotes in her book Russel and

Airasian that an important component in learning is assessment (Jannah, 2024). In Law Number 20 of 2003, Article 57, paragraph 1, concerning the National Education System, it is explained that evaluation is carried out in the context of national education quality control as a form of accountability of education providers to interested parties. One form of learning evaluation is testing. Therefore, in conducting assessments, teachers need tools in the form of tests. In order to produce good assessment results (Rizky Ananda Setiyawan & Palupi Sri Wijayanti, 2020).

High-quality assessment instruments can provide an accurate picture of students' abilities and competencies and their learning outcomes (Herliana & Kurniawan, 2021). The quality of assessment instruments is often suboptimal, so that the assessment results do not reflect the students' actual abilities. The quality of the learning process can be seen from the learning outcomes of the students. For this reason, in every learning activity, a test is conducted to determine whether the learning is going well or not (Nida et al., 2025). Learning outcome tests are generally created by educators. To obtain high-quality test results, educators must determine the quality of the tests to be used. Educators must be able to ensure that the testing tools to be used are of a quality that matches the expected competencies (Awaluddin et al., 2024).

Tests are tools for measuring students' knowledge and skills in solving problems. Test instruments must be developed in accordance with standards so that they can be used as a measure of student ability (Iskandar & Rizal, 2018). To create good and accurate testing instruments to improve the quality of learning, it is necessary to analyze the testing instruments. Ministry of Education and Culture Regulation No. 23 of 2016 states that the principles of assessment are: valid, objective, fair, integrated, transparent, comprehensive, systematic, criterion-referenced, and accountable. The quality of assessment is greatly influenced by the assessment instruments used (Indah et al., 2021).

Assessment in the 2013 curriculum focuses on students' ability to apply attitudes (affective), knowledge (cognitive), and skills (psychomotor) acquired through assignments tailored to real life (Setyaningrum et al., 2024). There are three domains for assessing student learning outcomes: knowledge (cognitive), attitude (affective), and skills (psychomotor). These three domains are interrelated because cognitive learning outcomes can be transformed into psychomotor learning outcomes if students act or behave in accordance with the values contained in the cognitive and affective domains (Forisma et al., 2023). In this case, instrument quality analysis can be conducted through theoretical (qualitative) analysis as well as empirical (quantitative) analysis (Sukiman, 2012)

Test instrument quality analysis is a stage that must be undergone by someone with the intention of determining the degree of quality of a test instrument for each item that is part of the test instrument as well as for the instrument as a whole (Rizky Ananda Setiyawan & Palupi Sri Wijayanti, 2020). It should be noted that the quality of test instruments can be analyzed by examining their validity, reliability, level of difficulty, and discriminating power. Assessment instruments must meet the following requirements: (1) substance that represents the competencies being assessed; (2) construction that meets technical

requirements in accordance with the form of the instrument used; (3) use of appropriate and correct language and communication in accordance with the developmental level of the students.

The writing of this article is inseparable from previous research. One of them is the research conducted by Lisda Fitriana Masitoh and Weni Gurita Aedi (2020) entitled "Development of Higher Order Thinking Skills (HOTS) Assessment Instruments in Mathematics for Grade VII Junior High School Students" which aims to develop HOTS-based assessment instruments in mathematics learning. This study falls under the category of research and development (R&D) and adopts the Borg & Gall model. The developed instrument consists of 14 open-ended questions that have undergone expert validation and field testing. Analysis was conducted quantitatively and qualitatively, covering validity, reliability (coefficient 0.733), discriminative power (average 0.33), and difficulty level (average 0.5, moderate category). The results indicate that the instrument is suitable for use and capable of effectively measuring students' higher-order thinking skills.

Lisda Fitriana Masitoh's research focuses on the empirical development of HOTS assessment instruments through field testing, while this study is more theoretical and descriptive in nature, analyzing the quality of PAI assessment instruments at various levels of education. The main difference lies in the research methods (R&D vs. literature review) and the outcomes (instrument development vs. literature review). Despite differing approaches, both emphasize the importance of validity, reliability, and the measurement capabilities of assessment instruments in education (Masitoh & Aedi, 2020).

Based on the above description, this article aims to describe the results of a literature search on the quality analysis of learning assessment instruments at the primary, secondary, and higher education levels. More specifically, this article discusses Islamic religious education.

## **LITERATURE REVIEW**

### **Definition of Assessment**

Common terms often used in assessment include methods that can be used to evaluate the learning outcomes of individual students and groups (Meylani, 2021). Assessment is a procedure used to obtain information related to learning outcomes in the form of observations and written tests, as well as a means of evaluating learning progress. Many people assume that assessment is a form of evaluation, while evaluation itself is an integral component of assessment. The evaluation process involves both quantitative and qualitative measurements, which are integrated and inseparable from one another (Mandar et al., 2025).

When assessing student learning outcomes, educators are often confronted with three important terms that are frequently used but often misunderstood: measurement, assessment, and testing. Tests are part of a series of tasks used to obtain systematic observations that represent the characteristics of students or their psychological profiles ((Agustina, 2022)

Learning is a system consisting of various components that interact with each other to achieve predetermined learning objectives (Nida et al., 2024). In every learning process, it is very important for teachers and students to know whether the objectives have been achieved. This can only be determined if teachers conduct evaluations, both of the learning process and the learning outcomes. In other words, evaluation includes assessment (Albab et al., 2024). From this statement, we conclude that tests are an important and commonly used tool for assessing learning outcomes. In general, assessment is a form of evaluation, while evaluation itself is an integral part of the evaluation process.

### **Analysis of Learning Instrument Quality**

In evaluation activities, assessment tools or techniques are needed so that the implementation will be more focused. Evaluation tools in education used to collect data can be in the form of tests or non-tests. Regarding our understanding of tests, a test is one of the evaluation or assessment tools used to measure students' learning outcomes throughout the learning process. Tests are typically used to measure how far students have progressed or advanced after undergoing a certain period of teaching and learning (H. A. S. Saputri & Larasati, 2023).

Tests are a complement to a teacher's assessment or evaluation to identify the level of a student's ability after exploring basic competency education conducted at the end of learning, such as daily tests. Before conducting a test, it is advisable to first analyze the test items or the quality of the test. In general, the quality of the questions used as assessment tools is not yet known (H. A. Saputri et al., 2023). In other words, we process the test results in such a way that we can identify which components of the learning process are still weak (Khaerudin, 2015). High-quality questions are those that can provide data on students' understanding of the material that has been presented or taught by the teacher.

There are two ways to analyze test items: 1) theoretical or qualitative analysis, and 2) empirical or quantitative analysis (Jihanna Amalia, 2023). Theoretical analysis or qualitative analysis can be conducted before or after the trial is carried out. The analysis method involves examining the items of the questions that have been compiled in terms of: Their alignment with the core competencies and indicators being measured, as well as compliance with requirements in terms of content, construction, and language for multiple-choice questions. There are also some differences in requirements for essay-type questions. Meanwhile, empirical (quantitative) analysis is aimed at examining the validity of the questions, reliability, discriminative power, difficulty level, and specifically for multiple-choice question models, the effectiveness of the distractor function must also be examined (Sukiman, 2012).

### **RESEARCH METHODOLOGY**

The research method used was a literature study. In literature research, literature search involves utilizing library resources to obtain research data without the need for field research (Khatibah, 2011). This article was written by reading, reviewing, and analyzing

existing literature such as books as the main source, research results, journals, articles, and other documents. The data in this article was collected from several sources relevant to the writing theme, such as educational assessment books, Ministry of Education and Culture regulations, and other literature. The collected data was analyzed using descriptive analytical analysis to draw conclusions about the object or topic being discussed in a systematic, factual manner, and in relation to the phenomenon being analyzed (Jonathan Saswono, 2006)

## **RESULTS AND DISCUSSION**

### **Qualitative Analysis of Assessment Instrument Quality**

Theoretical analysis is an examination of questions that focuses on material, construction, and language aspects. The material aspect relates to the scientific substance of the question and the level of thinking involved, the construction aspect relates to the technique of writing questions, and the language aspect relates to the clarity of the question (Sukiman, 2012). Theoretical analysis is an examination of questions that focuses on material, construction, and language aspects. The examination of the quality of objective questions in terms of material is intended to determine whether the material being tested is in line with the specified competencies or learning outcomes, whether the material in the questions is appropriate for the level or grade of the test takers' thinking abilities, and whether the answer keys are in line with the main content of the questions. The quality analysis of questions in terms of construction aims to determine whether the writing techniques used in the question items adhere to the principles of good question writing. In terms of language, the analysis of questions aims to determine whether the language used is clear and easy to understand, does not lead to multiple interpretations, and aligns with the established rules of language usage (SHELEMO, 2023).

Theoretical analysis or qualitative analysis can be conducted before or after the trial is carried out. The analysis method involves examining the questions that have been compiled in terms of their fulfillment of substantive, structural, and linguistic requirements. This examination can be conducted using the following analysis sheet instrument (Sukiman, 2012) :

**Table 1.** Multiple Choice Item Analysis Sheet

| TYPES OF REQUIREMENTS                                                                                | QUESTION NUMBER |   |   |   |   |
|------------------------------------------------------------------------------------------------------|-----------------|---|---|---|---|
|                                                                                                      | 1               | 2 | 3 | 4 | 5 |
| A. Substance Domain                                                                                  |                 |   |   |   |   |
| 1. Questions are in line with basic competencies and indicators                                      |                 |   |   |   |   |
| 2. Questions are in line with the material being tested.                                             |                 |   |   |   |   |
| 3. The content of the questions should be appropriate for the level, type of school, and grade level |                 |   |   |   |   |

|                                                                                                                     |  |  |  |  |  |
|---------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|
| 4. The content of the material asked is in line with the development of the students                                |  |  |  |  |  |
| B. Construction Domain                                                                                              |  |  |  |  |  |
| 1. The main issue (stem) is clearly formulated                                                                      |  |  |  |  |  |
| 2. The formulation of the main question should be limited to the necessary sentences only                           |  |  |  |  |  |
| 3. Homogeneous answer choices in terms of content                                                                   |  |  |  |  |  |
| 4. The length of the alternative/answer choices is relatively the same                                              |  |  |  |  |  |
| 5. The question does not provide clues/point to the correct answer                                                  |  |  |  |  |  |
| 6. Avoided alternative answers such as "all of the above are correct" or "none of the above are correct" or similar |  |  |  |  |  |
| 7. Answer choices in the form of numbers/times are sorted                                                           |  |  |  |  |  |
| 8. All answers must be logical and all distractors must function                                                    |  |  |  |  |  |
| 9. The location of the correct answer is determined randomly                                                        |  |  |  |  |  |
| 10. The main point does not contain negative statements                                                             |  |  |  |  |  |
| 11. If you have to use negative words, underline or italicize them                                                  |  |  |  |  |  |
| 12. Discourse, images, or graphics really work                                                                      |  |  |  |  |  |
| 13. The questions are not interdependent                                                                            |  |  |  |  |  |
| C. Language Domain                                                                                                  |  |  |  |  |  |
| 1. Formulation of communicative sentences                                                                           |  |  |  |  |  |
| 2. Sentences use proper and correct language, in accordance with the language variety                               |  |  |  |  |  |
| 3. The wording of the sentence does not give rise to multiple interpretations or misunderstandings                  |  |  |  |  |  |
| 4. Use common language/words (not local language)                                                                   |  |  |  |  |  |
| 5. The wording of the questions does not contain words that could offend the feelings of the students               |  |  |  |  |  |

**Table 2.** Question Analysis Sheet Short Answer Format

| TYPES OF REQUIREMENTS | QUESTION NUMBER |   |   |   |   |
|-----------------------|-----------------|---|---|---|---|
|                       | 1               | 2 | 3 | 4 | 5 |
|                       |                 |   |   |   |   |

|                                                                                                                                           |  |  |  |  |  |
|-------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|
| A. Substance Domain                                                                                                                       |  |  |  |  |  |
| 1. Questions are in line with basic competencies and indicators                                                                           |  |  |  |  |  |
| 2. Questions are in line with the material being tested.                                                                                  |  |  |  |  |  |
| 3. The content of the questions should be appropriate for the level, type of school, and grade level                                      |  |  |  |  |  |
| 4. The content of the material asked is in line with the development of the students                                                      |  |  |  |  |  |
| B. Construction Domain                                                                                                                    |  |  |  |  |  |
| 1. Formulate sentences in the form of open sentences (incomplete sentences) that only require the addition of words that are answers/keys |  |  |  |  |  |
| 2. The wording of the questions is not taken directly from a textbook                                                                     |  |  |  |  |  |
| 3. The questions are not dependent on the previous questions                                                                              |  |  |  |  |  |
| 4. Blank spaces for answers are placed at the end or near the end of a sentence rather than at the beginning of a sentence                |  |  |  |  |  |
| C. Language Domain                                                                                                                        |  |  |  |  |  |
| 1. Formulation of communicative sentences                                                                                                 |  |  |  |  |  |
| 2. Sentences use good and correct language, in accordance with language varieties                                                         |  |  |  |  |  |
| 3. Sentence formulation does not give rise to multiple interpretations or misunderstandings                                               |  |  |  |  |  |
| 4. Use common language/words (not local language)                                                                                         |  |  |  |  |  |
| 5. The question formulation does not contain words that can offend learners                                                               |  |  |  |  |  |

**Table 3.** Item Analysis Sheet for Essay Form

| TYPE OF REQUIREMENT | TEST NUMBER |   |   |   |   |
|---------------------|-------------|---|---|---|---|
|                     | 1           | 2 | 3 | 4 | 5 |
| A. Substance Domain |             |   |   |   |   |

|                                                                                                                             |  |  |  |  |  |
|-----------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|
| 1. Question items are in accordance with basic competencies and indicators                                                  |  |  |  |  |  |
| 2. Question items are in accordance with the material tested                                                                |  |  |  |  |  |
| 3. The content is appropriate for the level, school type and grade level                                                    |  |  |  |  |  |
| 4. The content of the material asked is developmentally appropriate for the learner                                         |  |  |  |  |  |
| B. Construction Domain                                                                                                      |  |  |  |  |  |
| 1. The scope of the questions and answers must be clear and firm                                                            |  |  |  |  |  |
| 2. Sentence formulation in the form of an interrogative sentence or command that demands a decomposed answer                |  |  |  |  |  |
| 3. There are clear instructions on how to do the questions                                                                  |  |  |  |  |  |
| 4. There is a scoring guideline                                                                                             |  |  |  |  |  |
| 5. Tables, graphs, diagrams, cases, or the like are meaningful (clear information or connection to the problem being asked) |  |  |  |  |  |
| 6. The item does not depend on the previous item                                                                            |  |  |  |  |  |
| C. Language Domain                                                                                                          |  |  |  |  |  |
| 1. Communicative sentence formulation                                                                                       |  |  |  |  |  |
| 2. Sentences use good and correct language, in accordance with language varieties                                           |  |  |  |  |  |
| 3. Sentence formulation does not give rise to multiple interpretations or misunderstandings                                 |  |  |  |  |  |
| 4. Use common language/words (not local language)                                                                           |  |  |  |  |  |
| 5. The question formulation does not contain words that can offend students                                                 |  |  |  |  |  |

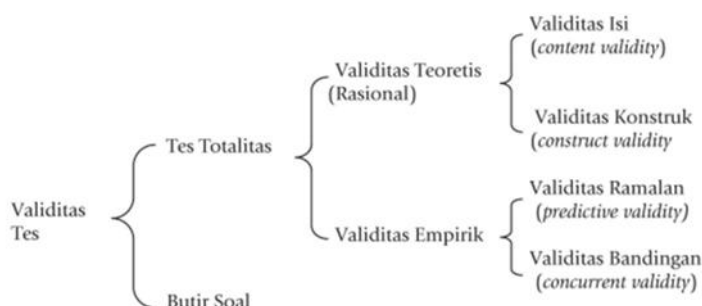
### **Quantitative Analysis of Assessment Instrument Quality**

Quantitative test analysis is directed at examining the level of question validity, reliability, differentiating power, difficulty level and specifically for multiple choice question models it is also necessary to examine the effectiveness of the distractor function (Sukiman, 2012).

*First, Test Validation.* A measuring device can be said to be a valid measuring device if the measuring device can measure what is to be measured precisely. For example, a barometer is a valid tool for measuring air pressure. But this tool is not valid for measuring temperature. Likewise in evaluation tools. A learning outcome test can be said to be a valid test if the test can actually measure learning outcomes. So it is not just measuring memory or language skills, for example (Annisa Prihartini ;Setri Yusma Sari ; Lisa Sefrita, 2021). Validity can be interpreted as accuracy, truth, validity or validity. If the word valid is associated with the function of the test as a measuring tool, then a test is said to be valid if the test accurately, correctly, validly, or validly can measure what should be measured (Khaerudin, 2015). Validity is a testing step carried out on the content of an instrument, with the aim of measuring the accuracy of the instrument used in a study (H. A. Saputri et al., 2023)

Test validity can be viewed from several aspects, namely: Predictive validity, Concurrent validity, Content validity, d. Construct Validity (Khaerudin, 2015). Analysis of test validity can be done from two aspects, namely: in terms of the test as a totality and in terms of its items, as an integral part of the test in total. The various test validity analyses can be described as follows: (Sukiman, 2012)

**Figure 1.** Test Validation



*Second, Reliability.* A reliable test provides a consistent measure of a student's ability to question achievement regarding a goal. Reliability shows consistent values. The reliability of a test instrument refers to the level of stability or consistency of the test instrument, where when given to the same subject, either by different people, at different times, or in different places, the results tend not to have significant differences (Fernando et al., 2023) A learning outcome test can be declared reliable if the results of measurements made using the test repeatedly on the same subject, always show the same results or are steady and stable.

There are two types of reliability, namely external reliability and internal reliability. External reliability is obtained when the measures or criteria are outside the instrument, while internal reliability is calculated based on data derived from the instrument itself.

- a. **External reliability.** There are two methods to test the external reliability of an instrument, namely using the parallel technique and the repeat technique. If the researcher chooses the parallel technique, then two sets of instruments need to be developed. The two instruments are then tested on one group of respondents (respondents do both), and the results of the two trials are correlated through correlation techniques, such as product-moment correlation or Pearson correlation. In the analysis, the data from the two trials are considered as X and Y values. The resulting correlation level indicates the extent of the instrument's reliability.
- b. **Internal reliability.** Internal reliability can be obtained by analyzing data from a single testing session. There are different ways to assess internal reliability, and the choice of technique depends on the form of the instrument as well as the researcher's preference. Sometimes, the use of different techniques can result in varying reliability indices, which may be influenced by the nature or characteristics of the data, as well as factors such as rounding. Some commonly used techniques for measuring reliability involve specific formulas, such as 1) Spearman-Brown; 2) Flanagan; 3) Rulon; 4) K-R 20; 5) K-R 21; 6) Hoyt; 7) Alpha.

Item difficulty level is a statement about how easy and how difficult an item is for students subjected to measurement. Suke Silverius quoted by khaerudin states that the level of difficulty of the item is the percentage of students who can answer correctly the item (Khaerudin, 2015). It can tentatively be said that one of the characteristics of a good item is that it is neither too difficult nor too easy for the particular group to be tested.

From some of the above definitions, it can be concluded that the level of difficulty is how much the level of difficulty / difficulty of a question item is indicated by the percentage of students who answer correctly to the question item. The level of difficulty of the question is a measurement of how much the degree of difficulty of a question. A good question is a question that is not too easy or not too difficult. Questions that are too easy do not stimulate students to increase their efforts to solve them. Conversely, questions that are too difficult will cause students to become discouraged and have no enthusiasm to try again because they are out of reach.

The way that can be taken to find out whether the learning outcomes test items already have an adequate level of difficulty or not, it can be seen from the size of the item difficulty index (difficulty index). The item difficulty index is a number or number that indicates the difficulty and ease of a question item. The amount of the difficulty index is between 0.00 and 1.0. This difficulty index shows the level of difficulty of the question. A question with a difficulty index of 0.0 indicates that the question is too difficult, on the other hand, an index of 1.0 indicates that the question is too easy. The number that shows the difficulty and ease of a question is called the difficulty index. The amount of the difficulty index is between 0.00 and 1.0. This difficulty index shows the level of difficulty of the question. A question with a difficulty index of 0.0 indicates that the question is too difficult, on the contrary, an index of 1.0 indicates that the question is too.

The formula used to determine the level / difficulty index of each item is:

**Figure 2.** Formula for measuring item difficulty

$$P = \frac{B}{Js}$$

With:

P = Difficulty index

B = The number of students who answered the question correctly

Js = Total number of students taking the test

This is the interpretation of test difficulty

**Figure 3.** Test Difficulty Interpretation

| No | Interval    | Klasifikasi |
|----|-------------|-------------|
| 1. | 0,00 – 0,30 | Soal Sukar  |
| 2. | 0,31 – 0,70 | Soal Sedang |
| 3. | 0,71 – 1,00 | Soal Mudah  |

*Third*, distinguishing power. Differentiating power analysis is to examine items in terms of the ability of a test to distinguish students who fall into weak categories and strong categories (Maulana, 2023). Meanwhile, Anas Sudijono explained that, the distinguishing power of an item is the ability of a learning outcome test item to be able to distinguish between testees with high ability (clever) and testees with low ability (stupid) in such a way that most of the testees who have high ability to answer the item are more likely to answer correctly, while testees with low ability to answer the item, most of them cannot answer the item correctly (Ramadhani et al., 2025).

The index of discriminating power is calculated on the basis of dividing the group into two parts, namely the upper group of high-ability test takers and the lower group of low-ability test takers. The index of discriminating power is defined as the difference between the proportion of correct answers in the upper group and the proportion of correct answers in the lower group. This group division can be done with the most widely used method is to determine 27% of the upper group and 27% of the lower group.

The distinguishing power of the item can be known by looking at the size of the discrimination index number. The item discrimination index is generally given the symbol "D" (discriminatory power). Like the difficulty index, this discrimination index ranges from 0.00 to 1.00. In the discrimination index, a negative sign is used if an inverted question shows the quality of the testee, that is, smart children are called stupid and stupid children are called smart. To calculate the differential power, the following formula is used:

**Figure 4.** Formula for Calculating Differentiating Power

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

**Description:**

D = magnitude of differential power

JA = number of participants in the upper group

JB = number of participants in the lower group

BA = number of upper group participants who answered correctly

BB = number of lower group participants who answered correctly

**Figure 5.** Interpretation of Calculating Distinguishing Power

| No | Indeks daya beda | Klasifikasi |
|----|------------------|-------------|
| 1. | 0,71 - 1,00      | baik sekali |
| 2. | 0,41 - 0,70      | baik        |
| 3. | 0,21 - 0,40      | cukupjelek  |
| 4. | 0,00 - 0,20      |             |

*Fourth, Distractor Function.* Distractors are said to function when at least one test taker from the low group chooses them. There should be more voters from the lower group than the upper group). Distractors can also be said to function when test takers from the upper group can distinguish between distractors and answer keys so that those who choose the answer key are more than those who choose distractors (Arfah, 2021)

Distractor function analysis is carried out specifically for objective form multiple choice items. In the multiple choice model question, it is equipped with several alternative answers called options. These options usually range from 3 to 5 pieces. Of the 3, 4 or 5 options, there is one that is the correct answer and that is called the answer key while the rest are wrong answers and are called distractors (Sukiman, 2012).

Analyzing the function of distractors is known as analyzing the pattern of distribution of item answers on multiple choice questions. From the pattern of distribution of item answers, it can be determined whether the distractor is functioning properly or not. A distractor can be said to function properly if at least 5% of test takers choose it (Solichin, 2017).

### Quality Analysis of Elementary Level Islamic Learning Instruments

The following is an example of analyzing the quality of the instrument in analyzing the quality of each item in class V of SD Budi Mulia, in the cognitive domain.

The name of Surah At Tin is taken from the word At Tin found in the first verse of Surah At-Tin. The meaning of the word At-Tin is...

- |                      |                |
|----------------------|----------------|
| A. Grape fruit       | C. Tin Fruit   |
| B. pomegranate fruit | D. Olive Fruit |

Answer Key: C

Based on the analysis of the quality of the items in the material feasibility aspect of question number 1, from the four aspects analyzed, including; 1) the questions are in accordance with the indicators, 2) the material asked is in accordance with the composition (urgency, relevance, continuity, high daily use), 3) homogeneous and logical answer choices and 4) there is only one answer key.

Based on the analysis of the quality of the question items in the aspect of the feasibility of the construction of question number 1, from the ten aspects analyzed, including; 1) the subject matter is formulated briefly, clearly, and firmly, 2) the formulation of the subject matter and answer choices is a statement that is only needed, 3) the subject matter does not provide clues to the answer key, 4) the subject matter is free from double negative statements, 5) the answer choices are homogeneous and logical in terms of material, 6) images, graphs, tables, diagrams, or the like are clear and functional, 7) the length of the answer choices is relatively the same. 8) answer options do not use the statement "all answers above are wrong/right" and the like, 9) answer options in the form of numbers/time are arranged in order of the size of the numbers or chronologically, and 10) question items depend on the answers to the previous questions (Nugrohaji & Muhsin, 2024).

Based on the analysis of the quality of question items in the aspect of language feasibility of question number 1, from the six aspects analyzed, including; 1) using language that is in accordance with Indonesian language rules, 2) using communicative language, 3) not using local / taboo language, 3) answer choices do not repeat the same word / group of words, unless it is a single understanding. 5) the question sentence does not copy/plagiarize exactly a reading text. 6) the sentence in the main question does not offend a person's personality, ethnicity, race, and religion, it is found that there are aspects that have not been fulfilled such as the existence of answer choices that repeat words such as answer choices C and D.

### **Quality Analysis of Intermediate Level Islamic Learning Instruments**

The following is an example of item analysis on the level of difficulty. The results of the Islamic religious education UAS in Class XI SMA 1 Bekasi. taken as an example of 30 students with 10 questions with manual calculations. The data below is only an example, the actual data is contained in the attachment, which is 30 students with 30 questions or items.

**Tebel 1.** Analyze each item using Microsoft Excel

| Nomor Butir | Tingkat Kesukaran |            |
|-------------|-------------------|------------|
|             | Koefisien         | Keterangan |
| 1           | 0,667             | Sedang     |
| 2           | 0,967             | Mudah      |
| 3           | 0,967             | Mudah      |
| 4           | 0,133             | Sulit      |
| 5           | 0,933             | Mudah      |
| 6           | 0,900             | Mudah      |
| 7           | 0,767             | Mudah      |
| 8           | 0,867             | Mudah      |
| 9           | 0,433             | Sedang     |
| 10          | 0,633             | Sedang     |
| 11          | 0,767             | Mudah      |
| 12          | 0,967             | Mudah      |
| 13          | 0,567             | Sedang     |
| 14          | 0,500             | Sedang     |
| 15          | 0,967             | Mudah      |
| 16          | 0,967             | Mudah      |
| 17          | 0,900             | Mudah      |
| 18          | 0,933             | Mudah      |
| 19          | 0,967             | Mudah      |
| 20          | 0,767             | Mudah      |
| 21          | 0,533             | Sedang     |
| 22          | 0,267             | Sulit      |
| 23          | 0,867             | Mudah      |
| 24          | 0,867             | Mudah      |
| 25          | 0,900             | Mudah      |
| 26          | 0,867             | Mudah      |
| 27          | 0,700             | Sedang     |
| 28          | 0,867             | Mudah      |
| 29          | 0,867             | Mudah      |
| 30          | 0,933             | Mudah      |

From the results of the analysis of the questions above, the distribution of the 30 items based on the level of difficulty can be categorized in the following table:

**Tabel 2.** Question Item Level of Difficulty

| No | Tingkat Kesukaran      | No. Butir Soal                                                                 | Jumlah | Persentase |
|----|------------------------|--------------------------------------------------------------------------------|--------|------------|
| 1  | 0,00 – 0,30 ( Sukar )  | (4, 22)                                                                        | 2      | 6,7 %      |
| 2  | 0,31 – 0,70 ( Sedang ) | (1, 9, 10, 13, 14, 21, 27)                                                     | 7      | 23,3 %     |
| 3  | 0,71 – 1,00 ( Mudah )  | (2, 3, 5, 6, 7, 8, 11, 12, 15, 16, 17, 18, 19, 20, 23, 24, 25, 26, 28, 29, 30) | 21     | 70 %       |

From the percentage of the distribution table of the items above, it can be seen that the Final Exam Question of the Gasal Semester of the subject of Islamic Religious Education and Ethics of class XI in terms of the level of difficulty there are 21 items in the category of easy questions, 7 items in the category of medium questions and 2 items in the difficult category. It is further stated that to compile an exam script, it is better to use items with a balanced level of difficulty, namely 25% = Easy, 50% = medium, and 25% = Difficult.<sup>19</sup> So

based on the explanation above, it can be concluded that the questions used in the odd semester final exam for Islamic Religious Education subjects at SMA 1 Bekasi can be said to be unbalanced.

### **Quality Analysis of Higher Level PAI Learning Instrument**

Bagian

The following is an analysis of the quality of learning outcomes instruments in the cognitive domain, the cognitive domain can be used by analyzing qualitatively. The following is an analysis of the quality of assessment instruments in the module of Islamic religious education courses at Yogyakarta State University (UNY) using the type of qualitative analysis requirements in the form of multiple choice and description:

#### **Module 1 “Man and Religion”**

- a. The material domain is in accordance with the measurement objectives and in accordance with the level. However, there is still one indicator that has not been discussed so it is necessary to add material related to the values of plurality, diversity, unity and integrity.
- b. The construction domain is clearly structured, avoids alternative answers, items are not dependent, and are not double-meaning.
- c. The language domain is communicative, in accordance with a refined spelling order, and does not contain language that offends learners.

#### **Module 2 “Islamic religion”**

- a. The material domain is in accordance with the measurement objectives and is appropriate for the level. However, the discussion indicators related to religion as a religion of rahmatan lil 'alamin need to be sharpened.
- b. The construction domain is clear, avoiding alternative answers. However, there is repetition of questions in the 2nd two and 6th numbers.
- c. The language domain is communicative, in accordance with the improved spelling order, and does not contain language that offends learners.

#### **Module 3 “Sources of Islamic Teaching Law”.**

- a. The material domain is in accordance with the measurement objectives and in accordance with the level. However, indicators and discussions related to the source of Islamic teachings in the form of qiyas need to be added to expand students' knowledge. The discussion related to qiyas as a source of Islamic law after the Qur'an, hadith, and ijma' is misinterpreted in various methods of determination in ijihad, so that the discussion becomes ambiguous.
- b. The construction domain is clearly organized, avoids alternative answers, items are not dependent, and are not double-meaningful.
- c. The language domain is communicative, in accordance with the improved spelling order, and does not contain language that offends learners.

## CONCLUSION

This article discusses the importance of the quality of assessment instruments in education, especially in learning Islamic Religious Education (PAI) at primary, secondary, and tertiary levels. Quality assessment instruments play an important role in accurately measuring students' abilities, competencies, and learning outcomes. However, many current assessment instruments are not optimal, so the evaluation results do not always reflect students' actual abilities.

Instrument quality analysis can be done qualitatively (theoretical) and quantitatively (empirical). Qualitative analysis focuses on the substance, construction, and language aspects. While quantitative analysis includes validity, reliability, difficulty level, differentiating power, and the effectiveness of the distractor function. Each of these aspects is important to ensure that the assessment instrument is able to measure learning outcomes fairly, objectively and thoroughly.

The study also showed differences in the quality of instruments at different levels of education. At the primary level, the appropriateness of the indicators and materials is often the main concern. At the secondary level, the analysis of the level of difficulty shows an imbalance in the categories of easy, medium and difficult questions. While at the higher levels, attention is paid to the substance and clarity of the material being taught.

Overall, this article emphasizes the importance of understanding and implementing assessment instrument analysis to improve the quality of learning evaluation. This is in line with the principles of assessment in education, such as validity, fairness and sustainability. This research suggests the need to increase educators' competence in compiling and analyzing assessment instruments in accordance with applicable standards.

## REFERENCES

- Agustina, R. (2022). *Pengembangan Instrumen Assessment Test*. 10(2), 695–707.
- Albab, U., Jawad, A., M, F. A., Sukman, S., Heremba, R., Alhamid, A. R., & Hasyim, M. F. (2024). Integrasi Pengukuran Kuantitatif dan Penilaian Kualitatif dalam Evaluasi Pendidikan Integration of Quantitative Measurement and Qualitative Assessment in Educational Evaluation. *JICN: Jurnal Intelek Dan Cendikiawan Nusantara*, (November), 6435–6443.
- Alkhairi, A. A., Nida, S., & Dewanti, S. S. (2024). The influence of PAI learning outcomes and learning motivation on students' honesty character. *Jurnal Penelitian Pendidikan Islam [Journal of Islamic Education Research]*, 12(2), 145–158.
- Annisa Prihartini ;Setri Yusma Sari ; Lisa Sefrita. (2021). Kualitas Instrumen Evaluasi. *IQRA': Jurnal Ilmiah Keislaman*, 03(01).
- Arfah, M. (2021). Evaluasi Hasil Belajar Pendidikan Agama Islam (PAI). *Jurnal Literasiologi*, 7(2), juli-desember.

- Awaluddin, R. Z. S., Sabililhaq, I., & Nida, S. (2024). Reimplementation of madrasah ramah anak in building discipline character of students. *ATTARBIYAH: Journal of Islamic Culture and Education*, 9(2), 193–212.
- Fernando, D. A., Hartatiana, & Ismail, F. (2023). Pentingnya Validitas Dan Reliabilitas Instrument Evaluasi Pada Pembelajaran Pendidikan Agama Islam. *Jurnal Tarbiyah Islamiyah*, 8(3), 64–73.
- Forisma, A., Ni'mah, Z., & Sukiman. (2023). Teknik Dan Instrumen Asesmen Keterampilan Pendidikan Agama Islam Di Dikdasmen Dan Pendidikan Tinggi. *Jurnal Evaluasi Pendidikan*, 14(1), 17–24. <https://doi.org/10.21009/jep.v14i1.36741>
- Herliana, O., & Kurniawan, R. Y. (2021). Pengembangan Instrumen Penilaian Berbasis HOTS pada Mata Pelajaran Ekonomi. *Jurnal Pendidikan Ekonomi*, 6(1), 47–51.
- Indah, M., Karoma, K., & Rusdi, A. (2021). Analisis Tes Butir Soal Guru dalam Mata Pelajaran Pendidikan Agama Islam (PAI) di Sekolah Menengah Pertama Negeri 8 Palembang. In *Muaddib: Islamic Education Journal* (Vol. 4, Issue 1). <https://doi.org/10.19109/muaddib.v4i1.8860>
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12–23. <https://doi.org/10.21831/pep.v22i1.15609>
- Jannah, R. (2024). Analisis Kualitas Butir Soal dan Abilitas Siswa pada Asesmen Madrasah Ibtidaiyah dengan Rasch Model. *Jurnal Didaktika Pendidikan Dasar*, 8(1), 331–342. <https://doi.org/10.26811/didaktika.v8i1.1234>
- Jihanna Amalia, S. (2023). Analisis Kualitas Instrumen Asesmen Ranah Kognitif Pendidikan Agama Islam Di Universitas Negeri Yogyakarta. *Dar El Ilmi: Jurnal Keagamaan, Pendidikan Dan Humaniora*, 10(1).
- Jonathan Saswono. (2006). *Metode Penelitian Kualitatif dan Kuantitatif* (Issue January).
- Khaerudin, K. (2015). Kualitas instrumen hasil belajar. *Jurnal Madaniyah*, 2(IX), 212–235.
- Khatibah, K. (2011). Penelitian kepustakaan. *Iqra': Jurnal Perpustakaan Dan Informasi*, 5(01), 36–39.
- Mandar, Y., Putri, M., Hidayat, Y., Lutan, A., & Nida, S. (2025). Innovative Approaches In Islamic Education: Imre Lakatos ' Methodological Paradigm And Its Implications. *Geneologi PAI: Jurnal Pendidikan Agama Islam*, 12(01), 17–28.
- Masitoh, L. F., & Aedi, W. G. (2020). Pengembangan Instrumen Asesmen Higher Order Thinking Skills ( Hots ) Matematika. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 04(02), 886–897.
- Maulana, H. (2023). Analisis Kualitas Instrumen Evaluasi Pembelajaran Menggunakan Media Digitalisasi Untuk Memotivasi Hasil Belajar Peserta Didik. *Jurnal Pendidikan Bhinneka Tunggal Ika*, 1(4), 1–12.
- Meylani, Y. (2021). Pengaruh Penerapan Model Pembelajaran Discovery Learning Berbasis Hots Terhadap Hasil Belajar Afektif pada Mata Pelajaran IPA Siswa Kelas V SDN 171 Seluma. *Almufi Jurnal Pendidikan*, 1(3), 125–132.

- Nida, S., Rafiqie, M., Mashuri, S., Setyaningrum, P. N., & Sagra, A. A. (2025). Dynamic Transformation of Islamic Religious Education Curriculum In Indonesia. *Edupeia: Jurnal Studi Pendidikan Dan Pedagogi Islam*, 9(2), 192–199.
- Nida, S., Rahmawati, N. W., Isbah, M. F., & Putri, M. (2024). Pendidikan Islam Di Era Globalisasi 5.0: Strategi Inovatif Untuk Tantangan Masa Depan. *At-Tajdid: Jurnal Pendidikan Dan Pemikiran Islam*, 8(2), 512–525.
- Nugrohaji, A. S., & Muhsin, A. M. (2024). Persepsi Mahasiswa Pai Uinsa Tentang Relevansi Pembelajaran Agama Islam Dan Kebudayaan Jawa. *Azkia: Jurnal Aktualisasi Pendidikan Islam*, 19(2), 83–96.
- Ramadhani, P. W., Mardian, R., & Rofiqon, M. H. (2025). Peran Nilai-Nilai Islam Dalam Mengerahkan Etika Ilmu Pengetahuan Dan Teknologi Modern. *Azkia: Jurnal Aktualisasi Pendidikan Islam*, 20(1), 77–89.
- Rizky Ananda Setiyawan, & Palupi Sri Wijayanti. (2020). Analisis Kualitas Instrumen Untuk Mengukur Kemampuan Pemecahan Masalah Siswa Selama Pembelajaran Daring Di Masa Pandemi. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(2), 130–139. <https://doi.org/10.46306/lb.v1i2.26>
- Saputri, H. A. S., & Larasati, N. J. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(5), 2986–2995.
- Saputri, H. A., Zulhijrah, Larasati, N. J., & Shaleh. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran, dan Daya Beda Butir Soal. *Didaktik : Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 09(05), 2986–2995.
- Setyaningrum, P. N., Nida, S., Rahmawati, N. W., & Isbah, M. F. (2024). Inovasi Pendidikan Islam Berbasis Kisah Al-Qur'an: Menelaah Inspirasi Dan Implementasi Dalam Pengembangan Pembelajaran. *Paramurobi: Jurnal Pendidikan Agama Islam*, 7(2), 140–152.
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Dirāsāt: Jurnal Manajemen & Pendidikan Islam*, 2(2), 192–213.
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Insan Madani.