

Characteristics of Islamic religious education assessment instruments

Rini Mulyani ^{a,1,*}, Karimullah ^{b,2}, Wahab ^{c,3}, Sahrul Sobirind, ^{d,4}

^{a,b,c,d} Pascasarjana, Institut Agama Islam Pontianak, Indonesia

¹ adindarini30@gmail.com ; ² karimfuad20@gmail.com ; ³ abdulwahabassambasi@gmail.com ; ⁴

sahrulsobirin82@gmail.com

*Correspondent Author

Received: 14 Februari 2024

Revised: 12 Maret 2024

Accepted: 10 April 2024

KEYWORDS

Characteristics;
Assessment Instrument;
Islamic Religious Education

ABSTRACT

Validity, reliability, differentiation, and level of difficulty are essential aspects of educational instruments, ensuring the accuracy and usability of data for decision-making. Validity is categorized into content validity and construct validity. Content validity ensures that the instrument aligns with the curriculum, material, and educational objectives, providing an accurate representation of what students are expected to learn. Meanwhile, construct validity focuses on how well the instrument measures the intended concept, ensuring that each test item effectively captures the learning outcomes being assessed.

Reliability refers to the consistency of the instrument's results over time, often determined by Cronbach's alpha coefficient. A Cronbach's alpha value of 0.5 or higher is generally considered acceptable for research purposes, indicating that the instrument produces reliable data. This metric helps ensure that variations in scores reflect actual differences in student performance rather than measurement errors or inconsistencies in the instrument itself.

In addition, the differentiation ability of a test indicates how well the instrument distinguishes between high-performing and low-performing students. This aspect ensures that the test can effectively identify differences in students' understanding. The level of difficulty measures how challenging each question is, ensuring a balanced distribution of easy, moderate, and difficult items. Together, these components contribute to the development of robust assessment tools that support accurate and meaningful educational evaluation.

A well-constructed instrument enhances the quality of education by guiding teachers and researchers in identifying student competencies and areas needing improvement. It also provides objective data that can inform curriculum development and teaching strategies. By combining validity, reliability, differentiation, and appropriate difficulty levels, educational assessments can promote fair and equitable learning environments, ensuring that every student's progress is accurately measured and addressed.

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



Introduction

The national education system listed in law number 20 of 2003 explains that education not only forms intelligent students but forms the character or personality of students so that a generation with morals and character is born that is in line with national and religious values. According to (Ali, 2022) as professionals, teachers are obliged to plan learning, carry out a quality learning process, and assess and evaluate assessment results. Meanwhile (Afifatun, 2018) argues that evaluation is an important and inseparable part of the current education system. Improving the quality of education can be seen from the values obtained by students. Of course for that a good and unbiased assessment system is needed. The existence of a relationship between lessons and student engagement is needed so that the learning process is more effective (Sunawan, 2017).

In the Qur'an surah al-ankabut verse 2-3 explains that the need to conduct an evaluation:

أَحْسِبِ النَّاسُ أَنْ يُتْرَكُوا أَنْ يَقُولُوا آمَنَّا وَهُمْ لَا يُفْتَنُونَ وَلَقَدْ فَتَنَّا الَّذِينَ مِنْ قَبْلِهِمْ فَلَيَعْلَمَنَّ اللَّهُ الَّذِينَ صَدَقُوا وَلَيَعْلَمَنَّ الْكَاذِبِينَ

"Do people think once they say, "We believe," that they will be left without being put to the test. We certainly tested those before them. And 'in this way' Allah will clearly distinguish between those who are truthful and those who are liars".

Based on the verse of the Qur'an above, the creator starts a test, exercise or evaluation to all his creatures. Measuring their faith potential in the khazanah knowledge, the evaluation is held to measure the extent of the achievements that students have achieved in learning and see the compound. (Daryanto, 2013) provides a comment on TP (Purpose of Assessment): (a) Describe the learning abilities of students so that they can be known positively and negatively. (2) Seeing the success of the education and teaching process at school, namely how far its effectiveness in changing the behaviour of students. A measuring instrument called an instrument is a measuring instrument that has been tested for validity and reliability of data. Several factors that affect the validity and reliability of data depend on whether or not the data collection instrument or object measure of a research variable. Whether an instrument is good or bad is determined by its validity and reliability (Zogara, 2021).

In line with what is expressed (Zogara, 2021) the eligibility of validity and reliability status has certain data and conditions. In reality in the field, the assessment instruments used so far have not been tested for validity and reliability, so that the validity of the instrument is very important (Hariyati T, 2019). By construct or construct validation and by criteria with criterion validation (Yusup, 2018). In education and learning procedures, the quality system of teaching materials prioritises educator professionalism and student achievement. The education and learning process requires a combination of learning components that work together to form a complexity of disciplines both in terms of cognition, emotional and psychomotor. The application of PAI learning focuses on specialisation and on aspects of faith, morals and worship.

All areas of PAI research are based on the Qur'an and hadith. Furthermore, educators need to have a good interpretation of appropriate PAI learning methods to ensure that educators do not make mistakes in implementing them (Sulaiman, 2017). The learning sequence is connected according to the learning context and based on the final result. Then, the students' results are evaluated to integrate the students' interpretation of the taught material. Real, numerical education is student learning outcomes. Then the answer is by means of a validity test because it produces a measurement, in contrast to the reliability test which measures the extent to which the instrument can be trusted.

Method

The method in this article uses a library study (library research) according to Sutrisno Hadi (2002) the Libery Research method is a study whose study uses library data such as books, journals, and other documents related to the research title and how to method with data collection by understanding and studying the theory of literature related to the research. The data collection uses a way of finding sources and constructing from various sources such as books, journals, research that has been done.

Result and Discussion

1. Validity

a. Content validity

Content validity is determined using expert agreement. Expert agreement in the field of study or often used in determining content validity (Retnawati, 2014) is used or psychological constructs are measured.

Table I. Criteria for Rating Instrument Items by Validators

Rate	Description
1	Not Relevant
2	Less Relevant
3	Fair
4	Relevant
5	Very Relevant

Table 2: Content validity criteria

Content validity coefficient	Criteria
0.00-0.19	Very Low
0.20-0.39	Low
0.40-0.59	Medium
0.60-0.79	High
0.80-1.10	Very High

From the two theories above, the researcher concludes that content validity is an instrument that can be said to be valid if it is in accordance with the curriculum, material and objectives.

b. Construct validity

According to (Sugiyono, 2019) it takes about 30 sample members used, for data collection, analysing and correlating item scores to carry out construct validity using the moment product formula: (Sugiyono, 2019).

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}}$$

That is:

r = correlation coefficient

n = total instrument item score

$\sum X_i$ = overall answer score

$\sum Y_i$ = the square result of the item score

$\sum Y_i^2$ = the result of the square of the total answer score

$\sum X_i Y_i$ = the number of times the item answer score is multiplied by the total score

if $r \geq 0.3$, the minimum requirement is considered acceptable if the correlation is less than 0.3, the item in the device is said to be invalid (Sugiono, 2016). From the two theories above, the researcher concludes that construct validity is the validity that questions how far the test items are able to measure what they really want to measure in accordance with the specific concept or conceptual definition that has been set.

2. Reliability

Reliability refers to an understanding that the instrument is reliable enough to be used as a data collection tool because the instrument is good. Reliability also refers to the level of credibility of something and can be trusted (Arikunto, 2013).

Table 3. Degree of Reliability

Reliability coefficient	Interpretation
$0,90 < \leq 1,00$	Very high
$0,70 < \leq 0,90$	High
$0,40 < \leq 0,70$	Medium
$0,20 < \leq 0,40$	Low
$r \leq 0,20$	Very low

According to (Basuki, 2014) the reliability of the instrument is if the Cronbach's alpha value is 0.60 and less than 1, then the instrument has a high correlation or is reliable, whereas if the Cronbach's alpha value is below 0.5 and below, then the instrument has a low correlation or is not reliable. And this is also supported by (Surapranata, 2004) that a reliability coefficient of 0.5 can be used for research purposes. From the two theories above, the researcher concludes that the reliability of the instrument is determined by the Cronbach's alpha value, if the Cronbach's value.

3. Distinguishing power

Item differentiation is an index that shows the level of ability of the item to distinguish the upper group (high-achieving test takers) from the lower group (low-achieving test takers). (Elviana, 2020) From the theory above, the researcher concludes that the differentiating power of the question is the ability of the question to distinguish between groups of test takers who understand and groups of test takers who do not understand. According to (Arikunto, 2013) in

the book Learning Evaluation (Widiyanto, 2018) Classification of Instrument Distinguishing Power.

Table 4. Classification of Differential Power

SCORE	CLASSIFICATION
0,00-0,20	(poor)
0,20-0,40	(satisfactory)
0,40 – 0,70	(good)
0,70 – 1,00	(excellent)
0,00	Problem cannot be used

In line with the opinion (Elviana, 2020) The coefficient of differential power ranges from -1.00 to +1.00, a positive value only means that the greater the index of differential power, the closer to 1, the item is able to distinguish the upper group and the lower group. Items that are considered to have good differentiating power if the item is > 0.5 .

4. Level of difficulty

The level of item difficulty is the percentage or proportion of test takers to answer an item correctly. The magnitude of the level of difficulty ranges from 0.00 - 1.00. The greater the level of difficulty obtained from the results of the calculation, the easier the question is and the question must be revised. A question has $p = 0.00$ means that no student answers correctly and if it has $p = 1.00$ means that all students answer correctly (Arifin, 2017). A difficult question will discourage students because it is out of their reach, while a question that is too easy will discourage students from working on it. According to the provisions that are often followed, the difficulty index is often classified as follows:

Table 5. Instrument Difficulty Level Predicate

SCORE	PREDICATE
0,00 – 0,30	Difficult
0,30 – 0,70	Medium
0,71 – 1,00	Easy

However, some argue that questions that are considered good, namely moderate questions, namely questions that have a difficulty index of 0.30 - 0.70. It should be noted that questions that are too easy or too difficult do not mean that they should not be used. This depends on its usefulness.

Characteristics of Islamic Religious Education Assessment Instruments

Islamic religious education (IRE) plays a crucial role in the holistic development of a Muslim individual, encompassing spiritual, moral, and intellectual growth. The assessment instruments used in IRE must therefore align with these comprehensive educational goals. The primary objective of these assessments is not only to gauge students' knowledge but also to nurture their spiritual and ethical development in accordance with Islamic principles.

One of the fundamental characteristics of IRE assessment instruments is their alignment with Islamic values and principles. These assessments are designed to reinforce Tawhid (the Oneness of God), the central tenet of Islam, and to cultivate a deep sense of faith and understanding of Islamic teachings. Additionally, they emphasize Akhlaq (ethics and morality), ensuring that students not only learn Islamic laws and jurisprudence but also embody the ethical conduct prescribed by Islam.

The holistic approach of IRE assessments is another key characteristic. Islamic education aims to develop the whole person, integrating spiritual, moral, and intellectual dimensions. Therefore, assessment instruments must evaluate not just cognitive knowledge but also students' spiritual growth and moral behavior. This can be achieved through various forms of assessment, such as reflective essays, character-based evaluations, and participation in community service projects.

Formative assessments are an integral part of IRE, allowing for continuous monitoring of students' progress. These assessments provide ongoing feedback that helps students improve and deepen their understanding of Islamic teachings. Methods such as regular quizzes, class discussions, and peer assessments are commonly used to keep track of students' learning and to provide immediate feedback for improvement.

Summative assessments, including traditional exams, oral tests, and project-based evaluations, are also prevalent in IRE. These assessments are used to measure students' comprehension at the end of a learning period. Oral exams, in particular, are significant in IRE as they allow students to demonstrate their memorization and recitation skills, which are essential components of Islamic education, especially in the study of the Quran.

Authentic assessment is another characteristic feature of IRE assessment instruments. This approach focuses on real-world applications of knowledge and skills, encouraging students to apply what they have learned in practical contexts. For example, students might be assessed on their ability to lead a prayer, deliver a sermon, or engage in community service, thus bridging the gap between theoretical knowledge and practical implementation.

The validity and reliability of assessment instruments are crucial in ensuring their effectiveness. Validity ensures that the assessments accurately measure what they are intended to measure, such as students' knowledge of Islamic teachings and their ability to apply them. Reliability ensures that the assessment results are consistent and dependable across different instances and contexts. High-quality IRE assessment instruments are meticulously designed to maintain these standards.

Cultural and contextual relevance is also vital for the effectiveness of IRE assessments. These instruments must reflect the cultural backgrounds and contexts of the learners, making the assessments more meaningful and relatable. For instance, incorporating local traditions and examples in assessments can enhance students' engagement and understanding.

Flexibility and adaptability are important characteristics of IRE assessment instruments. They should be able to cater to the diverse learning needs and paces of students. This flexibility allows for differentiated instruction, where assessments can be tailored to suit individual students' abilities and learning styles. Additionally, these instruments should be adaptable to changing educational contexts, such as the integration of technology in the learning process.

Despite the importance of these characteristics, there are several challenges in developing and implementing effective IRE assessment instruments. One significant challenge is balancing traditional Islamic educational methods with modern pedagogical practices. Educators must find ways to integrate classical Islamic teaching methods with contemporary educational tools and strategies. Additionally, ensuring that teachers are adequately trained and competent in both Islamic knowledge and modern assessment techniques is crucial for the success of IRE assessments.

In conclusion, the characteristics of Islamic religious education assessment instruments are deeply rooted in the principles and goals of Islamic education. They emphasize a holistic approach, cultural relevance, and adaptability, while ensuring validity and reliability. Despite the challenges, these assessment instruments play a vital role in fostering the spiritual, moral, and intellectual development of students, aligning with the overarching objectives of Islamic religious education.

Islamic religious education (IRE) plays a crucial role in the holistic development of a Muslim individual, encompassing spiritual, moral, and intellectual growth. The assessment instruments used in IRE must therefore align with these comprehensive educational goals. The primary objective of these assessments is not only to gauge students' knowledge but also to

nurture their spiritual and ethical development in accordance with Islamic principles. This discussion aims to explore the characteristics of effective assessment instruments in IRE, focusing on their alignment with Islamic values, holistic approach, inclusivity, types of assessments, and the challenges faced in their implementation.

Fundamental Principles of Islamic Religious Education Assessment

One of the fundamental characteristics of IRE assessment instruments is their alignment with Islamic values and principles. These assessments are designed to reinforce Tawhid (the Oneness of God), the central tenet of Islam, and to cultivate a deep sense of faith and understanding of Islamic teachings. Additionally, they emphasize Akhlaq (ethics and morality), ensuring that students not only learn Islamic laws and jurisprudence but also embody the ethical conduct prescribed by Islam. This alignment ensures that the assessments do not merely test rote memorization but also encourage students to internalize and practice Islamic values.

Types of Assessment Instruments in Islamic Religious Education

IRE assessments utilize a variety of instruments, each serving a specific purpose. Formative assessments are an integral part of IRE, allowing for continuous monitoring of students' progress. These assessments provide ongoing feedback that helps students improve and deepen their understanding of Islamic teachings. Methods such as regular quizzes, class discussions, and peer assessments are commonly used to keep track of students' learning and to provide immediate feedback for improvement.

Summative assessments, including traditional exams, oral tests, and project-based evaluations, are also prevalent in IRE. These assessments are used to measure students' comprehension at the end of a learning period. Oral exams, in particular, are significant in IRE as they allow students to demonstrate their memorization and recitation skills, which are essential components of Islamic education, especially in the study of the Quran.

Characteristics of Effective Islamic Religious Education Assessment Instruments

The effectiveness of IRE assessment instruments is determined by several key characteristics. Validity and reliability are crucial in ensuring their effectiveness. Validity ensures that the assessments accurately measure what they are intended to measure, such as students' knowledge of Islamic teachings and their ability to apply them. Reliability ensures that the assessment results are consistent and dependable across different instances and contexts. High-quality IRE assessment instruments are meticulously designed to maintain these standards.

Comprehensiveness is another important characteristic. Effective IRE assessments cover a wide range of knowledge and skills, including cognitive, affective, and psychomotor domains. This ensures that students are evaluated on their overall development rather than just their academic knowledge. For instance, assessments might include evaluating students' understanding of Islamic concepts, their ability to apply ethical principles in real-life situations, and their participation in community service activities.

Cultural and Contextual Relevance

Cultural and contextual relevance is also vital for the effectiveness of IRE assessments. These instruments must reflect the cultural backgrounds and contexts of the learners, making the assessments more meaningful and relatable. For instance, incorporating local traditions and examples in assessments can enhance students' engagement and understanding. This approach ensures that students see the relevance of their learning in their daily lives and cultural settings, thereby fostering a deeper connection with the material.

Flexibility and Adaptability

Flexibility and adaptability are important characteristics of IRE assessment instruments. They should be able to cater to the diverse learning needs and paces of students.

This flexibility allows for differentiated instruction, where assessments can be tailored to suit individual students' abilities and learning styles. Additionally, these instruments should be adaptable to changing educational contexts, such as the integration of technology in the learning process. This adaptability ensures that the assessments remain relevant and effective in various educational settings.

Challenges in Developing and Implementing Assessment Instruments

Despite the importance of these characteristics, there are several challenges in developing and implementing effective IRE assessment instruments. One significant challenge is balancing traditional Islamic educational methods with modern pedagogical practices. Educators must find ways to integrate classical Islamic teaching methods with contemporary educational tools and strategies. Additionally, ensuring that teachers are adequately trained and competent in both Islamic knowledge and modern assessment techniques is crucial for the success of IRE assessments.

Resource availability is another challenge. Access to quality materials and tools, as well as financial and infrastructural constraints, can hinder the development and implementation of effective assessment instruments. Schools and educators need adequate resources to create and administer comprehensive assessments that meet the educational goals of IRE.

Conclusion

In this case (validity, reliability, differentiation, level of difficulty) is an instrument that is very important in education so that the data generated can be used properly.

1. Validation is divided into content and construct validation. Content validation is an instrument that can be said to be valid if it is in accordance with the curriculum, material and objectives. While construct validity is validity that thinks
2. how much the test items can measure something that will be measured and adjusted to the specific concepts that have been determined.
3. The reliability of the instrument is determined by the Cronbach's alpha value, if the Cronbach's alpha value of 0.5 can be used for research purposes.
4. The distinguishing power of the question's ability to distinguish groups of test takers who understand and groups of test takers who do not understand.
5. The level of difficulty concludes that a difficult question will make students not enthusiastic because it is out of reach, while if the question is too easy it will make students not interested in doing it. So if the assessment instrument from the four concepts in the good criteria, so the instrument analysed by the reviewer has good capacity.

References

- Afifatun. (2018). *Perencanaan Sistem Evaluasi Pendidikan*. CV. LADUNY ALIFATAMA.
- Ali, M. (2022). Optimalisasi Kompetensi Kepribadian Dan Profesionalisme Guru Pendidikan Agama Islam dalam Mengajar. *Ar-Rusy*, 100-121.
- Allen. (1979). *Introduction to measurement theory*. Mexico: Cole Publishing Company.
- Arikunto. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Basuki, H. (2014). *Asesmen Pembelajaran*. Bandung: PT.Remaja Rosdakarya.
- Daryanto, A. (2013). *Evaluasi dan Penilaian Pembelajaran Kurikulum*. Yogyakarta: Gava Media.
- Depag. (2007). *Mushaf Alquran dan Terjemahan*. Jakarta: Alquran Pentashihan Departemen Agama RI.
- Elviana. (2020). ANALISIS BUTIR SOAL EVALUASI PEMBELAJARAN PAI PROGRAM ANATES. *Jurnal Mudarrisuna*, 58-74.
- Erlinawati, M. (2021). Test Validity and Reliability In Learning Evaluation. *Bulletin of Community Engagment*, 1-7.
- Hariyati T, Z. (2019). VALIDITY AND RELIABILITY OF PERFORMANCE ASSESSMENT. *JURNAL*

LITERASIOLOGI, 100-107.

- Hikmah, M. (2021). Validitas dan Realibitas Tes dalam menunjang Hasil Belajar PAI. Vol I.
- Khafidin. (2014). Analisis Validitas Dan Reliabilitas Tes Mata Pelajaran Pendidikan Agama Islam Tingkat SMA. *Jurnal Penelitian Pendidikan Islam*, 253–266.
- Mardapi. (2016). *Pengukuran Penilaian Dan Evaluasi Pendidikan*. Yogyakarta: UIN Yogyakarta.
- Retnawati. (2014). *Membuktikan validitas Instrumen Pengukuran*. Jakarta: UI Press.
- Sutrisno Hadi, *Metodelogi Research*, Andi Offset, Yogyakarta, 2002
- Sugiyono. (2019). *Statistika Untuk Penelitian*. Bandung: Alfabete.
- Sulaiman. (2017). *Metodologi Pembelajaran PAI*. Banda Aceh: Yayasan Pena Banda Aceh.
- Sunawan, D. H. (2017). Classroom Engagement. *Atlantis Press*, 50-82.
- Sundayana. (2016). *Statistika Penelitian Pendidikan*. Bandung: Alfabeta.
- Surapranata. (2004). *Analisis, validitas, reliabilitas, dan interpretasi hasil tes implementasi kurikulum 2004*. Bandung: PT. Remaja Rosdakarya.
- Suryadi, A. (2020). *Evaluasi Pembelajaran*. Bandung: CV. Jejak.
- Widiyanto. (2018). *Evaluasi Pembelajaran*. Madiun: UNIPMA PRESS.
- Yusup. (2018). Uji Validitas Dan Reliabilitas Instrumen Penelitian. *Jurnal Ilmiah Kependidikan*, 1-7.
- Zogara, A. u. (2021). *MEMAHAMI DASAR-DASAR PENELITIAN KESEHATAN*. In Haura Publishing.