E-ISSN : 2774-8472
P-ISSN: 2774-8480
Jurnal Hurriah: Jurnal Evaluasi Pendidikan dan Penelitian
Vol. 4, No. 1, Mar 2023, PP 117-124

ANALYSIS OF THE QUALITY OF ISLAMIC RELIGIOUS EDUCATION SUBJECT TEST APPLYING THE ITEM RESPONSE THEORY APPROACH TO THE RASCH MODEL

Alfi Faroh Kamaliya[1], Sri Sumarni[2], Himawan Putranta[3], Apriliana Indah Lestari[4], Fatma Nurhayati[5], Nely Anggraeni Ayuningtiyas[6]

UIN Sunan Kalijaga Yogyakarta
20104010049@student.uin-suka.ac.id

Abstract

This research aims to analyze and describe multiple choice tests and AKM Islamic Religious Education Subjects made by students through the Rasch Item Response Theory (IRT) approach. This type of research is descriptive quantitative research with research subjects totaling 30 student responses to multiple choice test instruments and AKM with five alternative answers. Quantitative data analysis was carried out through the Rasch Model Item Response Theory (IRT) approach with the help of the QUEST program, and using the content validity instrument by formulating the Aiken's V formula. The results of the analysis carried out showed that 13 items fit and 2 items did not fit and all items were valid, The reliability coefficient on the item summary shows the "enough" criteria and the estimation case summary reliability coefficient shows the "weak" criteria, the item difficulty level shows 13 items in the good category and 2 items in the poor category, as well as the results of validation by experts processed by experts which uses the Aiken's V formula shows that all items are classified as valid.

Keywords: *test quality, validity, reliability*

Abstrak

Penelitian ini bertujuan untuk menganalisis dan mendeskripsikan tes pilihan ganda dan AKM Mata Pelajaran Pendidikan Agama Islam (PAI) buatan mahasiswa melalui pendekatan Item Response Theory (IRT) model Rasch. Jenis penelitian ini adalah penelitian kuantitatif deskriptif dengan subjek penelitian sejumlah 30 respon mahasiswa terhadap instrument tes pilihan ganda dan AKM dengan lima alternative jawaban. Analisis data kuantitatif dilakukan melalui pendekatan Item Response Theory (IRT) Model Rasch dengan bantuan program QUEST, dan menggunakan instrument validitas konten dengan merumuskan formula Aiken's V. Hasil analisis yang dilakukan menunjukan 13 nomor butir fit dan 2 butir tidak fit dan keseluruhan butir soal valid, Koefisien reliabilitas pada Summary of item menunjukkan pada kriteria "cukup" dan koefisien reliabilitas summary of case estimate menunjukkan kriteria "lemah", tingkat kesukaran butir soal menunjukkan 13 butir soal kategori baik dan 2 butir soal kategori kurang, serta hasil validasi oleh para ahli yang diolah menggunakan rumus Aiken's V menunjukkan semua butir soal tergolong valid.

Kata Kunci: *kualitas tes, validitas, reliabilitas*

INTRODUCTION

A teacher must have several competencies in order to carry out his or her duties and obligations as an educator, one of which is pedagogical competence. In accordance with (Regulation No. 16 of 2007 of the Minister of National Education of the Republic of Indonesia Concerning Academic Qualification Standards and Teacher Competency, 2007.), Assessment and evaluation of learning processes and outcomes is one of the core competencies in pedagogic competence. Teachers must master these competencies, as evidenced by their ability to create assessment and evaluation tools for learning processes and outcomes. An excellent assessment is one that is capable of meeting the assessment principles. Some of the assessment principles mentioned are contained in Minister of

Manuscript received December 2022, revised February 2023, and published March 2023

Jurnal Hurriah: Jurnal Evaluasi Pendidikan dan Penelitian
Vol. 3, No. 3, Sep 2022, hal. 117-124

Education and Culture Regulation No. 23 of 2016, which includes the following principles: valid, objective, fair, integrated, open, comprehensive, and sustainable, systematic, criteria-based, and accountable. The assessment instrument used has a fairly strong influence on the quality of the assessment. Analyzing the instrument's quality is one of the procedures in the process and learning outcomes assessment activities. This instrument must meet the substance, construction, and language requirements, as well as demonstrate validity and reliability. (Indah & Rusdi, 2021).

As prospective educators, particularly those enrolled in Islamic Religious Education study programs, should be aware of the competencies they must possess in order to become professional educators in the future. Among the necessary teacher competencies is pedagogical competence, which includes the implementation of assessment and evaluation of learning processes and outcomes. This implies that students must develop assessment and evaluation instruments for learning processes and outcomes through practice.

As an educator, the learning outcomes test instrument created by the teacher can provide numerous descriptions or information about the abilities of their students. However, among the many issues that arise when the teacher prepares the test instrument, it is discovered that there are still flaws in the process of compiling the instrument, resulting in a test that is not measurably valid. Such an instrument cannot provide any description or information about students' abilities. To assess the success of the learning process, which is frequently carried out by educators, learning outcomes instruments are developed. (Pratama, 2020).

As an educator, the learning outcomes test instrument created by the teacher can provide numerous descriptions or information about the abilities of their students. However, among the many issues that arise when the teacher prepares the test instrument, it is discovered that there are still flaws in the process of compiling the instrument, resulting in a test that is not measurably valid. Such an instrument cannot provide any description or information about students' abilities. To assess the success of the learning process, which is frequently carried out by educators, learning outcomes instruments are developed.

The goal of the test item quality analysis is to determine whether the items compiled can serve as an adequate measure of learning outcomes. This activity also provides information about a question's lack of merit and serves as a guide for making improvements. The items can be analyzed from a variety of perspectives, including their content validity, empirical validity, reliability, and level of difficulties.

In the world of educational measurement, two approaches to test analysis are commonly used, namely classical test theory (Classical Test Theory) and item response theory (Item Response Theory).

However, the use of the classical test theory analysis approach has been discontinued because it is thought to have numerous flaws. (Pratama, 2020).

According to (Mardapi, 2012) In the Rasch model, measuring is a direct comparison of two objects, namely individuals and items. The individual here refers to the test taker's ability, while the item is a parameter of difficulty level. Thus, when the skill of test takers grows, for example, the chances of successfully answering the test items appear to increase. As a result, the opportunity to correctly answer the test items leads to two things: the ability of the individual test takers and the level of difficulty of the items. The same fundamental principle as described by the Rasch model is also expressed by (Sumintono & Widhiarso, 2015) which states that the Rasch model is a measurement model that determines the relationship between test-takers' aptitude and the level of test-item difficulty. It's also provides a more detailed overview of this relationship; for instance, if a test-taker is able to answer 85% of the questions correctly, he has a higher ability than those who can only answer 70% of the questions correctly.

This study aims to describe, using Aiken, the characteristics of multiple-choice questions and the General Competency Assessment (AKM) of Islamic Religious Education Subjects (PAI), including estimations of empirical validity and reliability, item difficulty level, and content validity analysis.

METHOD

The objective of this descriptive quantitative research is to obtain an overview of test quality based on the characteristics of the Rasch Model test and the content validity of the Aiken V. test. Students enrolled in the Islamic Religious Education (PAI) 5th semester study program at UIN Sunan Kalijaga Yogyakarta and UIN K.H. Abdurrahman Wahid Pekalongan comprised the demographic of this study. In this study, the sample consisted of 30 students: 15 students enrolled in semester 5 PAI at UIN Sunan Kalijaga Yogyakarta and 15 students enrolled in semester 5 PAI at UIN K.H. Abdurrahman Wahid Pekalongan. 30 student responses to the multiple-choice test instrument and the AKM for the Islamic Religious Education subject KD comprised the subjects of this study. 3.9. Examining the Effectiveness of Hajj, Zakat, and Waqf for Individuals and KD 3.6 Analyzing and Evaluating Islamic Marriage Provisions, with five possible answers.

Documentation is the data gathering approach used in this study. Documentation might take the form of a series of questions and responses from responders. The quantitative data analysis in this study was conducted using the Rasch Model Item Response Theory (IRT) approach with the assistance of the QUEST program, as well as the content validity instrument using Aiken to formulate the Aiken V formula to calculate content validity based on the results of expert assessments of n people to an item regarding the extent to which the item represents a construct, as measured by Aiken V coefficient values ranging from 0-0.99.

Jurnal Hurriah: Jurnal Evaluasi Pendidikan dan Penelitian
Vol. 3, No. 3, Sep 2022, hal. 117-124

RESULT AND DISCUSSION

A. Empirical Validity Estimation

Tabel 1. *Recapitulation of the Empirical Validity of Test Questions*

Nomor Butir Soal	INFIT MNSQ	OUTFIT MNSQ	Status	Keterangan
1	1,07	1,16	BUTIR FIT	VALID
2	0,74	0,55	BUTIR TIDAK FIT	VALID
3	1,23	2,27	BUTIR FIT	VALID
4	0,79	0,72	BUTIR FIT	VALID
5	1,1	1,2	BUTIR FIT	VALID
6	1,11	3,6	BUTIR FIT	VALID
7	1,11	0,98	BUTIR FIT	VALID
8	1,11	1,07	BUTIR FIT	VALID
9	0,89	1,1	BUTIR FIT	VALID
10	0,96	0,45	BUTIR FIT	VALID
11	1,17	1,25	BUTIR FIT	VALID
12	1,02	0,99	BUTIR FIT	VALID
13	0,75	0,68	BUTIR TIDAK FIT	VALID
14	0,98	0,93	BUTIR FIT	VALID
15	0,99	0,85	BUTIR FIT	VALID

The Quest program is used for this empirical validity investigation, which examines at the mean INFIT Mean of Square (INFIT MNSQ) value and its standard deviation. (Adams & Kho, 1996). If the INFIT MNSQ value of an item is between 0.77 and 1.30, it is said to be fit. The acceptance limit of the items implements INFIT MNSQ (between 0.77 and 1.30) and INFIT t with a limit of -2.0 to 0.2, resulting in items that meet the goodness fit. The OUTFIT Mean of Square (OUTFIT MNSQ) can also be used to determine the appropriateness of each item with the model. If 0.5 OUTFIT MNSQ 1.5, an item is said to fit the model. (Boone et al., 2014).

Table 1 shows the findings of the empirical validity of the exam questions utilizing Quest. According to the table, the empirical validity of the 15 items has the status of fit items at 13 item numbers and 2 things are not fit since the INFIT MNSQ is less than 0.77 - 1.30. However, given the OUTFIT MNSQ is between 0.5 and 1.5, all items are eligible.

A. Reliability Estimation

Tabel 2. *Item Case Estimate and Case Estimate Value Criteria*

Nilai Reliabilitas Item Estimate dan Case Estimate	Kriteria
> 0,94	Istimewa
0,91 – 0,94	Bagus Sekali
0,81 – 0,90	Bagus
0,67 – 0,80	Cukup
< 0,67	Lemah

(Sumintono & Widhiarso, 2015)

Tabel 3. *Test Item Reliability Results*

Reliabilitas	Koefisien Reliabilitas	Kategori
Summary of item estimate	0,76	Reliabel
Summary of case estimate	0,48	Tidak Reliabel

On the basis of the item separation index (item estimate) and the person separation index (case estimate), the item's dependability can be ascertained. The greater the separation index value of the test items, the greater the overall accuracy of the test items. And the higher the value of the person separation index, the greater the consistency of each item in measuring a person's capability. (Subali & Suyata, 2013).

As for the results of the items' reliability, they are shown in Table 3. The summary estimate's coefficient of reliability is 0.76. According to Sumintono and Widhiarso's table 3 criteria for the Rasch model's reliability value, 0.76 meets the "sufficient" criterion because it falls within the value range of 0.67 to 0.80. While the summary of case estimate reliability coefficient is 0.48, this indicates a "weak" criterion because the value is less than 0.67. This value indicates that respondents' responses are inconsistent. This can be interpreted as the respondent answering the queries carelessly, resulting in a low reliability value.

Jurnal Hurriah: Jurnal Evaluasi Pendidikan dan Penelitian
Vol. 3, No. 3, Sep 2022, hal. 117-124

B. Estimated Item Difficulty Level

Tabel 4. *Item Difficulty Analysis*

Nomor Butir Soal	*Difficulty*	Kategori
1	-0,97	Baik
2	0,1	Baik
3	-0,36	Baik
4	1,64	Baik
5	0,83	Baik
6	-2,12	Kurang
7	-0,63	Baik
8	1,64	Baik
9	0,48	Baik
10	-2,12	Kurang
11	-0,97	Baik
12	0,83	Baik
13	1,48	Baik
14	-0,36	Baik
15	0,52	Baik

The difficulty analysis of the test items was performed using the Quest software, and the items were deemed satisfactory if the difficulty index (b) fell between -2.0 and 2.0.(Retnawati, 2016). The difficulty level of the items can be discerned based on the value displayed in the.It output data from the program Quest.

The difficulty range of the items is between -2.12 and 1.64, which indicates that thirteen items fall into the decent category. While two items fall into the less category due to their scores falling outside the interval of -2.00 to +2.00, two items fall into the less category. The items are categorized as difficult if their difficulty score is close to +2.00, and they are categorized as simple if their difficulty score is close to -2.00.

C. Results of Content Validity Analysis Using Aiken's V

Tabel 5. *Content Validity Recapitulation*

No. Butir	Hasil Penilaian Ahli				S=r-l0				Sigma s	n(c-1)	Koefisien Aiken	Nilai V Minimal	Keterangan
	A	B	C	D									
1	3,86	4	4	3,93	2,86	3	3	2,93	11,79	12	0,98	0,92	Valid
2	3,86	3,86	3,93	3,93	2,86	2,86	2,93	2,93	11,57	12	0,96	0,92	Valid
3	3,86	4	4	3,93	2,86	3	3	2,93	11,79	12	0,98	0,92	Valid
4	3,86	3,93	4	3,93	2,86	2,93	3	2,93	11,71	12	0,98	0,92	Valid
5	3,86	4	4	3,93	2,86	3	3	2,93	11,79	12	0,98	0,92	Valid
6	3,79	3,93	3,93	3,93	2,79	2,93	2,93	2,93	11,57	12	0,96	0,92	Valid
7	3,79	4	4	3,93	2,79	3	3	2,93	11,71	12	0,98	0,92	Valid
8	3,79	4	3,93	3,93	2,79	3	2,93	2,93	11,64	12	0,97	0,92	Valid
9	3,86	3,93	4	3,93	2,86	3	3	2,93	11,71	12	0,98	0,92	Valid
10	3,86	4	4	3,93	2,86	3	3	2,93	11,79	12	0,98	0,92	Valid
11	3,86	4	4	3,93	2,86	3	3	2,93	11,79	12	0,98	0,92	Valid
12	3,86	4	4	3,93	2,86	3	3	2,93	11,79	12	0,98	0,92	Valid
13	3,86	4	4	3,93	2,86	3	3	2,93	11,79	12	0,98	0,92	Valid
14	3,93	4	4	3,93	2,93	3	3	2,93	11,86	12	0,99	0,92	Valid
15	3,93	4	4	3,93	2,93	3	3	2,93	11,86	12	0,99	0,92	Valid

In the conducted investigation, a total of fifteen questions were evaluated. Based on the evaluations of four validators, the study of multiple-choice questions and AKM for the Islamic Religious Education (PAI) subject was conducted. Examine the item items quantitatively using the validation page provided to the validator. There are multiple aspects, including material aspects, construction aspects, language aspects, compatibility aspects with literacy indicators for AKM questions, and numerical compatibility aspects for AKM questions.

The results of the experts' evaluation were then analyzed using Aiken's V formula so that the extent of the content's validity could be determined. The results of expert evaluations that have been processed using Aiken's V formula can be seen in the table 1 located above. The results of validation by experts using the Aiken's V formula indicate that all multiple-choice questions and AKM for Religious Education Subjects (PAI) created by students are classified as valid, as shown in the table. The validity of the items is demonstrated by the Aiken coefficient, which ranges between 0.96 and 0.99. Overall, the Aiken coefficient of multiple-choice and AKM questions for Religious Education Subjects (PAI) is 0.98. These results demonstrate that the item measuring instrument has a high level of content validity.

Jurnal Hurriah: Jurnal Evaluasi Pendidikan dan Penelitian
Vol. 3, No. 3, Sep 2022, hal. 117-124

CONCLUSION

On the basis of students' multiple-choice tests and AKM for Islamic Religious Education (PAI) subjects, the following characteristics of the tests and respondents as test participants can be identified:

1. The results of the empirical validity of the 15 items show that 13 item numbers are fit, 2 items are not fit, and all items are valid.
2. The reliability coefficient on the Summary of items reflects the "fair" criteria, whereas the reliability coefficient on the Summary of Case Estimate reflects the "weak" criteria.
3. The difficulty level of the items shows that 13 items are included in the good category. While 2 items fall into the less category.
4. The findings of expert validation using the Aiken's V formula reveal that all items are classified as valid.

REFERENCE

Adams, R. ., & Kho, S.-T. (1996). *Acer quest version 2.1*. The Australian Council for Educational Research.

Boone, W. ., Staver, J. ., & Yale. (2014). *Rasch analysis in the human sciences*. Springer.

Indah, M., & Rusdi, A. (2021). Analisis Tes Butir Soal Guru dalam Mata Pelajaran Pendidikan Agama Islam (PAI) di Sekolah Menengah Pertama Negeri 8 Palembang. *Muaddib: Islamic Educational Journal*, *4*(1), 21–28.

Istiqomah, N. A. I., & Akhmad, F. . (2021). problematika pembelajaran daring pai serta upaya kepala sekolah dalam mengatasinya. JURNAL HURRIAH: Jurnal Evaluasi Pendidikan Dan Penelitian, 2(4), 1-9. https://doi.org/10.5806/jh.v2i4.32

Mardapi, D. (2012). *Pengukuran Penilaian dan Evaluasi Pendidikan*. Yuha Medika.

Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 16 Tahun 2007 Tentang Standar Kualifikasi Akademik Dan Kompetensi Guru, 1 (2007).

Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, *7*(1).

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penilaian*. Parama Publishing.

Subali, B., & Suyata, P. (2013). STANDARDISASI PENILAIAN BERBASIS SEKOLAH. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *17*(1).

Sudjiono, A. (2009). *Pengantar Evaluasi Pendidikan*. Rajawali Press.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch Pada Assessment Pendidikan*. Trim Komunikata.