Edification Journal
Pendidikan Agama Islam
Sekolah Tinggi Agama Islam (STAI)
Bumi Silampari Lubuklinggau
ISSN: 2741-7681 (P); 6886-0988 (E)
Volume 7 No. 2, Januari 2025 | 275-291
DOI: https://doi.org/10.37092/ej.v7i2.1007

ANALYSIS OF THE QUALITY OF THE FARAID MATERIAL AKM INSTRUMENT TO MEASURE STUDENTS' REASONING ABILITY WITH ASSISTANCE *MINISTEP*

Nabila Lutfiati, Himawan Putranta
UIN Sunan Kalijaga, Yogyakarta, Indonesia
22104010050@student.uin-suka.ac.id, himawan.putranta@uin-suka.ac.id

Article History
Received : 25-12-2024
Revised : 10-01-2025
Accepted :18-01-2025

Keywords :
AKM
Faraid
Student Reasoning
Ministep

Abstrac

21st century skills in Indonesia are relatively low, especially in the literacy and numeracy fields. The government's solution is to implement a Minimum Competency Assessment (AKM). This research aims to analyze the quality of the AKM instrument on faraid material in measuring students' reasoning abilities. The instruments used are evaluated with the help of the software Ministep-based Quick model to ensure validity, reliability, level of difficulty, and function of distractors. This research uses quantitative research with a descriptive approach. The results of the analysis show that the empirical validity of most of the items meets the standards, although several items are declared invalid and require revision. Question reliability showed good consistency with a reliability score of 0.84, while respondent reliability only reached 0.39, indicating a lack of consistency in student answers. Difficulty level analysis produces a distribution of questions that includes too easy, medium, and too difficult, categories, ensuring the instrument covers a wide range of student ability levels. In addition, 60% of the distractors functioned well to ensure the effectiveness of the wrong answer choices in differentiating students' understanding. Thus, using Ministep as an evaluation tool helps develop a more valid, reliable, and representative AKM instrument for measuring students' reasoning abilities on varied material.

Introduction

Educational evaluations are generally always coupled with learning outcomes or assessment *output* from the educational process that has been carried out. In this educational evaluation process, we can determine the progress of an education and assess whether the educational goals have been achieved. The ultimate goal is perfect education (Hidayat et al., 2019: 128). So, educational evaluation is an assessment of the educational process carried out to measure, examine, and provide feedback on the learning process to assess success and achieve educational goals. One element of the assessment instrument in educational evaluation is a written test. The test itself is a measuring tool in assessment that functions to measure the level of students' abilities,

be it cognitive, affective, psychomotor, interest, motivation, perception, and so on (Murtafiaf, 2018: 25).

Specifically, in the 21st century, evaluation of the education process in Indonesia is related to the low literacy skills of Indonesian students based on the PISA test results (*Programme for International Student Assessment*). Indonesia's ranking in the PISA survey 2018 was 73 out of 78 participating countries (OECD, 2019). The test results show that Indonesian students still experience difficulties working on questions with *high-order thinking skills* (HOTS), which describes a student's critical thinking abilities. This statement is based on the results of the answers to the National Examination (UN) HOTS question section (Sani, 2021: 2). Moreover, the reality that occurs when the National Examination is carried out is only a formality due to the large number of frauds that occur (Hadi, 2020: 815). Not only literacy skills but numeracy skills also need to be improved for students in Indonesia (Sani, 2021: 3).

The government, in this case, the Ministry of Education and Culture (Kemendikbud), has provided a solution to improve literacy and numeracy skills for Indonesian students, namely by implementing a competency assessment for all schools in Indonesia in the form of the AKM (Minimum Competency Assessment) test (Kemendikbud, 2020: 3). Literacy and numeracy skills are important for students at the elementary/MI, SMP/MTs, and SMA/MA levels and university students. Literacy and numeracy skills are defined as the ability to process knowledge and understanding and implement them effectively to answer various problems in everyday life (Nashirulhaq et al., 2022: 118). Students as students at higher school levels, namely universities, should have strong reasoning abilities and critical thinking skills (Suharto et al., 2022: 326). AKM aims to help students solve problems that exist in everyday life in order to provide positive participation in society, so a student must have this ability (Kemendikbud, 2020: 5; Sani, 2021: 3).

Social problems occur in various areas of life, whether religious, social, economic, or cultural. Religious issues often become conflicts in society regarding the distribution of inheritance (Abdillah & Anzaikhan, 2022: 286). Fighting over inheritance between families, which should only be a family conflict, can often escalate into societal conflict. This is caused by the family or community not understanding the concepts in science *Faraidh* (Adibah, 2017: 7).

Mastery of faraid material has significant relevance in everyday life, especially in ensuring justice in inheritance distribution by Islamic law's provisions. Faraidh provides structured and systematic guidance regarding the distribution of inheritance to heirs, considering factors of kinship, gender, and generation (Abdillah & Anzaikhan, 2022: 287). This guide is very important to avoid family disputes that often arise due to unfair division of assets. In a social context, understanding faraidh helps create harmony in society, because justice in the distribution of inheritance can strengthen family relationships and reduce conflict (Basri, 2020: 38). Apart from that, from an educational perspective, mastery of faraidh material also shows how Islamic law provides concrete and relevant solutions to everyday problems, making this knowledge an integral part of Islamic religious education (Adibah, 2017: 3).

Furthermore, studying faraid also contributes to developing critical and logical thinking skills. In dividing an inheritance, one must understand the basic concepts of fiqh such as nisbah and ashabah and apply them in complex cases. This process demands in-depth analysis, problem-solving, and fair decision-making based on data

and legal rules (Asihati et al., 2023: 500). In addition, faraidh calculations often involve mathematical operations such as fractions, proportions, and algebra, which can improve mathematical logic skills. These abilities, which are honed through faraidh learning, are useful in inheritance law and can be applied in various fields that require analytical thinking and systematic problem solving (Kamulia et al., 2022: 48).

One way to measure students' reasoning abilities in faraidh material is by providing AKM regarding faraid material (Azzahra et al., 2024: 86). The development of quality Minimum Competency Assessment (AKM) instruments for faraidh material is very important in increasing students' understanding and skills in Islamic inheritance law (Hadianto, 2024: 91). A good AKM instrument can effectively measure students' literacy and numeracy competencies, which is the main aim of AKM in assessing critical and logical thinking abilities (Kemendikbud, 2020: 4). Developing quality AKM instruments can also help map the education quality in educational units (Andikayana et al., 2021: 84).

Previous research conducted studies that analyzed AKM questions, but there was still little regarding Islamic religious education material, especially faraid material. Research related to this research includes an analysis of 10 AKM questions regarding inheritance and marriage in Islam. The results show that for AKM questions regarding inheritance, the empirical validity of the AKM instrument is valid, but the reliability of case estimates is still weak. In contrast, analysis of marriage material shows that the 10 questions are valid with a spread of 0.77-1.30, but the consistency is still weak (Azzahra et al., 2024: 92; Hadianto, 2024: 101). Yasinta, Nurdiana, and Asmah showed in their research that students' mathematical reasoning abilities were still low. Namely, 70% of students were in the low category, 20% in the medium category, and only 10% in the high category. These results were obtained after research subjects worked on AKM questions, and students with a high level of reasoning could work on AKM questions well (Yasinta et al., 2024: 30).

The difference between this research and previous research is that in this research, the respondents were 43 students from the Islamic studies study program whose aim was to analyze the quality of the AKM instrument for faraid material. In this case the analysis tool is *Ministep* on faraid material to measure the reasoning ability of students, especially students of the Islamic studies group study program. This research also focuses on faraid or heritage material studied by Islamic studies students.

The research is very important for teachers and educators, especially in the context of Islamic religious education. This research aims to ensure that the AKM instrument used in faraid learning has high validity and reliability, so that the assessment results truly reflect students' reasoning abilities. Through Rasch model-based analysis using the Ministep tool, this study evaluates the quality of the items based on difficulty level, differentiating power, and conformity to the model. This is very relevant for teachers, because quality instruments can help them understand students' abilities more deeply and develop targeted learning strategies (Sumintono & Widhiarso, 2015: 4). In addition, this study provides recommendations for improvements to the AKM instrument, so that it can be used as an effective evaluation tool to improve the quality of faraid learning in the classroom. Thus, this research is not only useful to support better assessment practices, but also contributes to teachers' professional development in implementing learning evaluation.

This study aims to analyze the quality of the AKM instrument on faraid material in measuring students' reasoning ability with the help of Ministep. Specifically, this study aims to evaluate the validity and reliability of the instrument to ensure that the measurement tool used can provide precise and consistent assessment results. In addition, this research also focuses on identifying the difficulty level and differentiating power of the items, as well as the suitability of the items with the Rasch model to determine the effectiveness of the instrument as an evaluation tool. The results of this study are expected to provide recommendations for improvements to the faraid material AKM instrument, so that it can be used optimally in supporting the learning process and assessing students' reasoning skills.

Literature Review

Minimum Competency Assessment (AKM)

Minimum Competency Assessment (AKM) is part of the education evaluation system implemented in Indonesia to measure the basic competencies possessed by students, especially in terms of reading literacy and numeracy (Yasinta et al., 2024 : 23). AKM is a basic competency assessment that includes reading and numeracy (mathematics literacy). It is given to students to help them develop self-development and actively and positively participate in society (Kemendikbud, 2020: 3). This AKM is usually given to students in classes V SD/MI, VII SMP/MTs, and XI SMA/MA. In AKM, three skills are measured: logical-systematic thinking skills, reasoning skills based on the concepts of the material being studied, and finally, skills in filtering and processing the information obtained. Reading literacy, as measured in AKM, is defined as skills in understanding, using, evaluating, and reflecting on written texts to implement them in society in order to contribute to Indonesian citizens and citizens of the world (Nashirulhaq et al., 2022: 121). Meanwhile, numeracy is defined as thinking skills with concepts, procedures, facts, and mathematical tools to solve every problem experienced in every area of human life (Andikayana et al., 2021: 83).

The aim of AKM is quite clear: to collect information on students' abilities regarding the expected competencies, literacy, and numeracy skills, to improve the quality of teaching and learning, ultimately improving student learning outcomes (Sani, 2021: 4). There are two components in AKM: reading literacy and numeracy (Andikayana et al., 2021: 84). Reading literacy is usually informational or fictional texts that help improve cognitive processes by finding information, interpreting and integrating, and evaluating and reflecting. Context in reading literacy can be related to personal, socio-cultural, and scientific. Meanwhile, the components of numeracy are numbers, measurement and geometry, data and uncertainty, and algebra. This numeration also helps improve understanding, application, and reasoning related to data. The context of reading literacy and numeracy can be related to personal, socio-cultural, and scientific factors (Kemendikbud, 2020: 7).

Faraid material

Faraid, or the law of inheritance in Islam, is a branch of knowledge that specifically regulates the distribution of inheritance to heirs based on the Qur'an, Hadith, and Ijma' ulama. Faraid knowledge has a main foundation in QS. An-Nisa verses 11, 12, and 176 explicitly mention the number of rights of each heir, such as children, parents, spouses, and siblings (Sulthan et al., 2024: 226). Faraid also considered the status of kinship,

gender, and role in the family (Abdillah & Anzaikhan, 2022: 286). As a structured knowledge, faraid reflects social justice based on sharia, with the main purpose of maintaining harmony in the family and preventing conflicts that could arise due to unclear division of inheritance (Zuleika & Desinthya, 2015: 97).

In studying classical fiqh, faraid developed through the deep interpretation of jurists, such as Imam Abu Hanifah, Imam Malik, Imam Shafi'i, and Imam Ahmad bin Hanbal. Each sect pays attention to variations in cultural context, customary law, and social development. For example, the Syafi'i school emphasizes the literalistic principle in determining the rights of heirs, while the Hanafi school is more flexible in considering distant relatives (Zuleika & Desinthya, 2015: 102).

Ministep as an analysis tool

Ministep is the free version of the software Winsteps, designed for data analysis using measurement models Quickly. Ministep is software used to analyze assessment data in context Item Response Theory (IRT) (Farzad et al., 2017: 86). Item Response Theory (IRT) is a statistical theory about examinee's performance in item and test and how performance relates to the examinee abilities that are measured by the items in the test (Mckinley, 2020: 2). The quality assessment of this AKM instrument can be measured with the help of an application Ministep. In the context of recent research, Ministep has been used to develop and validate an instrument measuring individual participation post-hand rehabilitation.

This study emphasizes the importance of the application model Rasch in ensuring the validity and reliability of the instruments developed (Farzad et al., 2017: 86). Model Rasch was implemented through the software Ministep, which was used to evaluate the quality of the assessment instrument by checking the suitability of the items to the model. This model offers information about item difficulty, respondent competence, and the overall reliability and validity of the instrument (Untary & Risdianto, 2020: 65). Analysis Quickly, it is highly valued for its objectivity and is well suited to improving instruments such as the AKM to make them more fair and accurate.

In measuring students' ability in Faraid material, Ministep was used to analyze the quality of the AKM assessment instrument and how well the questions can measure student competence in reasoning and applying Faraid material. Ministep allows researchers to carry out a more detailed analysis of the questions in the AKM assessment instrument. By using IRT, Ministep can help determine the level of difficulty of the questions, question discrimination (the extent to which the questions can differentiate students' abilities), and how well the questions measure students' abilities in certain topics (Saputra & Sudrajat, 2024: 1969). This is very important to ensure that the questions in the AKM instrument are valid and relevant in measuring student competency.

Research Method

This research uses quantitative research with a descriptive approach. Quantitative research is an approach that tests theory by measuring the relationship between variables using statistical tools. The data collected is in numbers and analyzed using certain statistical methods. This research is usually done with a previously

designed research design, such as a survey, experiment, or secondary data analysis (Creswell & Guetterman, 2018: 185). This quantitative research method is often referred to as a method *of discovery* because, with this method, researchers can discover and develop various new sciences and technologies (Sugiyono, 2020: 8).

Quantitative research is divided into two parts, according to the level of research, namely descriptive research and experimental research. The approach in this research is descriptive research, which is interpreted as research that does not provide treatment to the variables studied (Sugeng, 2022: 29). This descriptive approach explains information about a symptom, event, or incident as it is (Mukhid, 202: 16). In this research, students' reasoning ability on faraid material will be explained.

This approach was chosen because it allows objective numerical data-based analysis to evaluate each question item's validity, reliability, level of difficulty, and distractor function. The instrument analyzed is in the form of multiple choice questions with four answer options designed to measure students' logical and analytical abilities regarding the distribution of inheritance in Islamic law. The data collection technique in this research is a test in the form of AKM questions, given via Google form and distributed to respondents. The respondents who were taken in this research were 42 people from the category of male and female students from 2021 to 2023 with the study program of the Islamic study cluster, namely the study program of Islamic Religious Education and Al Quran Science and Tafsir at UIN Sunan Kalijaga Yogyakarta. The question is related to understanding the concept of faraid or inheritance in Islam. The research instrument is in the form of ten multiple-choice questions with details of five questions related to literacy and five questions related to numeracy.

Instrument analysis is carried out using the software *Ministep*, which is based *on Rasch* (Muslimin & Sunardi, 2019: 173). *Model Rasch* was chosen because of its excellence in providing an in-depth and comprehensive analysis of the questions' quality and the respondents' abilities. The analysis process involves several main steps. First, data on students' answers to the instrument is collected and input into the device *Ministep*. Next, the software processes the data to produce various parameters, such as empirical validity, item and person reliability, item difficulty level, and distractor function. Empirical validity is evaluated based on parameters *he enters* And *outfit mean square* (MNSQ). Question items are considered valid if the MNSQ value is within the recommended range, namely 0.5 to 1.5. Instrument reliability is calculated in two dimensions, namely, item reliability and respondent reliability. Item reliability shows the consistency of the items in measuring the construct in question, while respondent reliability reflects the extent to which students' answers are consistent with the questions given. The reliability score is calculated based on the index generated by *Ministep*, where a score > 0.7 is considered reliable. Each item's difficulty level is analyzed using the logit value generated by *model Rasch* (Oktaviyanthi & Agus, 2020: 1131).

This research, using a quantitative approach, ensures that the AKM instrument developed is not only of high quality but also able to provide a comprehensive picture of students' reasoning abilities on varied material. The Ministep-based approach provides objective and in-depth results, which are very important for evaluating and developing better assessment instruments in the future (Taufiqi & Ellianawati, 2023: 16).

Result and Discussion

Result

1. Empirical Validity Test

The following are the results of the empirical validity test of the Faraid material questions given to respondents. A validity test will be used to measure the validity of each question item. In this research, the validity test uses an application *ministep*. Respondents' answers are processed with the help of the application *software ministep.* In this application, the data is processed in parts *summary statistics.* The following is a presentation of the results of the validity test.

Table 1. Empirical Validity Test Results

No. Item	INFIT		OUTFIT		MEASURE		DESC.
	MNSQ	ZSTD	MNSQ	ZSTD	CORR	EXP	
2	1.58	1.13	2.78	1.50	0.15	0.44	Invalid
3	1.41	2.10	1.57	1.67	0.30	0.53	Valid
9	1.32	1.83	1.43	1.27	0.35	0.53	Valid
10	0.94	-0.40	0.83	-0.24	0.54	0.50	Valid
5	0.80	-0.23	0.49	-0.24	0.52	0.44	Valid
6	0.78	-0.71	0.53	-0.91	0.63	0.51	Valid
7	0.78	-1.04	0.73	-0.78	0.65	0.53	Valid
4	0.70	-0.64	0.39	-0.67	0.60	0.46	Valid
8	0.61	-1.27	0.31	-1.40	0.70	0.50	Valid
1	0.59	-0.53	0.15	-0.53	0.57	0.40	Valid

An item is said to be valid if it has an MNSQ logit value between 0.5 and 1.5, a ZSTD logit value between -2.0 and 2.0, and a CORR value between 0.4 and 0.85. It is finished if the logit value of two of the three columns above (MNSQ, ZSTD, and CORR) reaches the standard good value, namely the INFIT, OUTFIT, and MEASURE columns. Based on the table, we can see that question item number 2 is said to be invalid because the logit values in the MNSQ column, which are 1.58 and 2.78, do not meet the standard, namely 0.5 – 1.5, so the MNSQ column is not achieved. Then we look at the ZSTD column where the values are 1.13 and 1.50, according to the standard, namely -2.0 – 2.0, so the ZSTD column meets the standard. Then, we look at the CORR column, where the value is 0.15, which is not by the standard, namely 0.4 – 0.85. Two columns do not comply with the standards, namely MNSQ and CORR, and only one column, namely ZSTD, meets the standards, so it can be concluded that question item number 2 is invalid.

2. Item and Person Reliability Test

The reliability of a research instrument determines the results of a study. Reliability is the level of reliability or consistency of a measuring instrument. If measuring instruments are used on the same object under the same conditions, the measurement results must be relatively consistent over time (Sugiyono, 2020: 121). In this study, reliability measured the consistency of the AKM instrument for 42 respondents regarding the faraid question. Whether or not a question is reliable can be

measured with help *ministep* in the same section as the question validity test. The following table presents the item's results and person estimate reliability tests in this research.

Table 2. Reliability Test Results *Item* And *Person Estimate*

Reliability Type	Score	Information
Reliability of Items	0.84	Reliable
Person Reliability	0.39	Not Reliable

This table tests the reliability of the instrument based on two dimensions: question reliability and respondent reliability. Question reliability reflects the consistency of the items in measuring certain abilities, while person reliability describes the consistency of students' answer patterns. A variable is reliable if its value *Alpha Cronbach* (r_{11}) ≥ 0.60. Judging from the data above, the value *alpha Cronbach* for the question is 0.84. This shows that the questions in this research are reliable. Meanwhile, the value *of alpha Cronbach* from *persons* or respondents is 0.39, so it is unreliable.

3. Distractor Function Test Results

Distractors as cheats in assessment instruments are very necessary in learning evaluation. Distractors are an important component of multiple-choice questions designed to illustrate common mistakes that test participants may make. A good distractor should be able to attract participants who do not have the correct understanding but not confuse participants who understand the concept. The following are the results of the distractor function test for ten AKM questions heritage.

Table 3. Distractor Function Test Results

Question Items	Answer Keys	Number of Responses for Each Choice				Distractor Function Test Results
		A	B	C	D	
2	B	1	39	1	1	Not Working
3	B	2	26	2	12	Works
9	D	10	1	7	24	Works
10	C	14	9	18	1	Works
5	D	1	0	2	39	Not Working
6	B	5	35	1	1	Works
7	B	4	30	5	3	Works
4	A	38	1	1	2	Not Working
8	C	4	2	36	0	Works
1	B	2	40	0	0	Not Working

The table above shows the number of people or respondents who chose each answer option. If the wrong choice answered by the respondent is more or almost equal to the correct answer key, then the question's distractor function works, and vice versa.

4. Question Difficulty Level Test Results

In determining a question's difficulty level, we can look at the number of respondents who answered correctly and incorrectly. In the evaluation process, this difficulty index is symbolized by the letter P, which stands for "proportion". The difficulty index value for each item can be obtained using the formula introduced by *Du Bois* (Sudijono, 2016: 371), namely:

$$P = \frac{Np}{N}$$

Where:
P = Proportion or item difficulty index number
Np = The number of testees who can answer the item correctly
N = number of testees who took the learning outcomes test.

Meanwhile, the measure used is: the lower the number obtained, the more difficult the question is. On the other hand, the higher the number, the easier the question is. Robert L. Thorndike and Elizabeth Hagen explain the measure of the difficulty level of this question as follows (Sudijono, 2016: 372).

Table 4. Question Difficulty Level Criteria

Large P Value	Information
Less than 0.30	Too Difficult
0,30 – 0,70	Fair (Medium)
More than 0.70	Too Easy

From the table above, we can understand that a problem's difficulty level is based on the quantity P, namely starting from ≤ 30 to ≥ 0.70. So, there are levels in the questions' difficulty; if the question is categorized as too difficult, it must have a P value of more than ≤ 0.30. The second level has a medium level with a P value $\geq 0.30 \leq 0.70$; the third level is categorized as too easy with a P value of ≥ 0.70. Based on the explanation above, the level of difficulty of each question item is obtained, namely:

Table 5. Difficulty Level for Each Question Item

Question Items	P value	Category
2	0,92	Too easy
3	0,61	Currently
9	0,57	Currently
10	0,42	Currently
5	0,92	Too Easy
6	0,83	Too Easy
7	0,71	Too Easy
4	0,90	Too Easy
8	0,85	Too Easy
1	0,95	Too Easy

The table above shows the level of difficulty of each question. For example, question number 2 is said to be too easy because it has a P value of 0.92. Presenting the results of the level of difficulty of the questions above, it can be concluded that there are four criteria for the level of questions: very difficult, difficult, medium, and easy. Furthermore, we can conclude the number of questions in each criterion from Table 5. The summary of the question categories is good. That very difficult, difficult, medium, and easy can be seen from the following table.

Table 6. Summary of Difficulty Levels of Question Items

Criteria	Item
Too Easy	1,2,4,5,6,7,8
Currently	3,9,10
Too difficult	-

It can be concluded that the seven questions, questions 2, 4, 5, 6, 7, and 8. T falls into the too-easy category. Threedium categories, name questions nine and 10, and then questions into the tattoo-difficult

5. Category results of Respondents' Answers Regarding Heritage Numeration

The total number of questions in this research is 10, with details of 5 literacy questions and 5 numeracy questions. Analysis results with the help *ministep* show that the number of respondents who answered correctly was greater than the number of respondents who answered incorrectly. The interval between the number of correct and incorrect respondents is also quite large. However, in question number 10, the interval between the number of correct and incorrect respondents was not too far. The following is one of the results of the respondents' answers, namely, in question number 10 regarding numeration.

Figure 1. Percentage of Respondents' Answers to Number 10

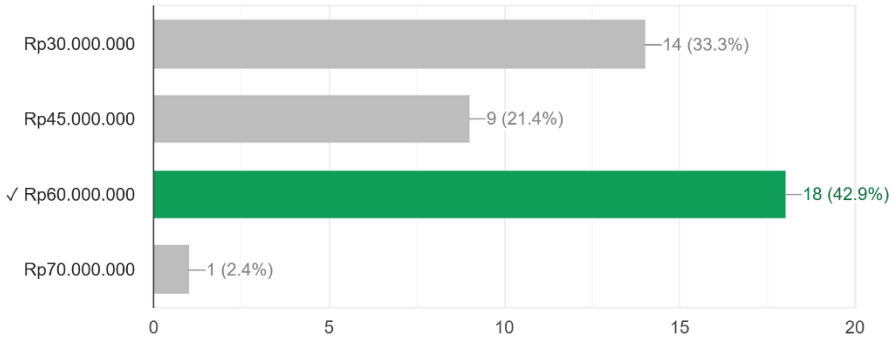

Figure 1 shows the percentage of respondents' answers to question 10 related to numeration in inheritance calculations. This horizontal bar graph depicts the distribution of respondents' answers based on the available answer options, namely IDR 30,000,000, IDR 45,000,000, IDR 60,000,000 and IDR 70,000,000. Each bar represents the number of respondents who chose a particular option, accompanied by the total percentage. The results show that most respondents, 20 people (44.4%), chose the answer IDR 60,000,000, which is the correct answer for this question. On the other hand, several respondents chose the wrong answer. A total of 15 people (33.3%) chose IDR 30,000,000. Furthermore, nine people (20%) chose IDR 45,000,000, while only one person (2.2%) chose IDR 70,000,000.

Discussion

An effective evaluation instrument is an important element in the learning process, especially for measuring students' reasoning abilities (Oktaviyanthi & Agus, 2020: 1125). AKM instrument analysis on Faraid material was carried out using software assistance *from the Minister based on the Rasch* model. This discussion will describe in detail the results of the analysis based on each table to provide an overview of the quality of the instrument in terms of validity, reliability, distractor function, and level of difficulty of the questions (Isnaeni et al., 2018: 110). One of the benchmarks for an assessment instrument is said to be good if proven valid and reliable (Ketaren et al., 2024: 3280).

The validity of the test functions as a measurement tool used to measure. In testing the validity of the test, three aspects are looked at, namely the content, theoretical design and criteria of the tool (Anshari et al., 2024: 969). In this research, the empirical validity test results are used, shown in Table 1. Empirical validity is the validity of an instrument seen from the experience side, namely tested from experience, which is then divided into predictions (*predictive validity)* and validity at the same time *(concurrent validity*) (Riyani et al., 2017: 62). Judging from Table 1 above, the validity of the AKM questions was analyzed through the QUEST program by looking at the average INFIT score *Mean of Square* (INFIT MNSQ) and OUTFIT MNSQ. The INFIT MNSQ value is acceptable if it is between 0.77 – 1.30, with the INFIT t limit being -2.0 to 2.0. Meanwhile, OUTFIT MNSQ is between 0.5 and 1.5 (Azzahra et al., 2024: 89). So, it can be concluded from the presentation in Table 1 that there is only one question, namely question number 2, which is said to be invalid because it does not meet the standards, while the other nine questions are said to be valid.

The reliability test results presented in Table 2 show that the reliability of the question is reliable, with a value of 0.84. For reliability, a *person* (respondents) with a value of 0.39 is in the unreliable category. This is because a variable is said to be reliable if it has a value *Alpha Cronbach* ≥ 0.60 (Rosita et al., 2021: 283). Crocker and Algina (1986), as quoted by Sumarna, stated that test reliability is influenced by factors such as test length, speed of work, homogeneity of test components, and level of difficulty of questions (Setiyawan, 2014: 348). In this study, respondents did not achieve reliability. Respondent reliability refers to the consistency and stability of answers given by individuals to measurement instruments, such as questionnaires or tests (Amalia et al., 2022: 10). If the respondents are unreliable, the answers tend to change or be inconsistent, which can affect the accuracy and reliability of the data collected (Setiyawan, 2014: 346). This is due to a lack of understanding of the responses to the questions given in this research.

Meanwhile, table 3 shows the question distractor function test results in this study. This study had six questions where the distractor worked and four questions where the distractor did not. This is because a distractor functions if it is chosen by a minimum of 5% of all respondents for four answer choices and 3% for three answer choices (Arikunto, 2021: 224; Sudijono, 2016: 413). In the table, for example, for question number 1, which has answer key B, of the 42 respondents, 40 (96%) answered correctly, and 2 (4%) respondents chose option A. No one chose options C and D. Analysis of distractor function This is done to understand how effective an answer choice is in a multiple-choice question. If there is a misleading option that the test taker does not choose at all, as was the case in this study in question number 1, it indicates that the option is not good, perhaps because it is too obvious (Muzayyanah, 2020: 14). A distractor can be treated in three ways: accepted because it is good, rejected because it is not good, and Rewritten because it is not good (Arikunto, 2021: 244).

One factor is whether the questions' difficulty level influences a distractor function. The test results for the difficulty level of the questions in this study can be seen in Tables 4, 5, and 6. The level of difficulty of the questions is important to ensure that there are variations in the level of ability being measured. Questions that are too easy or difficult can reduce the test's ability to differentiate between participants who understand the material well and those who do not (Sugiyono, 2020: 149). The difficulty

index is the number that shows how difficult or easy a question is. Difficulty index values range from 0.00 to 1.00. This index describes the difficulty level of the question, where an index of 0.00 indicates that the question is very difficult, while an index of 1.0 means the question is very easy (Sudijono, 2016: 371). The results of this research show that many questions still fall into the too-easy category. This must be corrected because when creating an assessment instrument, it must have a balanced level of difficulty, namely 25% in the difficult category, 50% in the medium category, and 25% in the easy category (Fatimah & Al Fath, 2019: 46).

One of the questions included in the medium category is question number 10, which is related to AKM numeration. As can be seen in Figure 1, 42.9% answered correctly, and the highest number of respondents who answered the wrong option was option A, with 33.3%. This also shows that in question number 10, the interval between the number of respondents who answered correctly and the number of respondents who answered the wrong option was the smallest among the other questions. Small differences between the number of respondents who answered correctly and incorrectly on a question item can occur due to several factors. Initially, the level of challenge in the questions impacts how the answers are spread out. Questions that are of medium difficulty usually result in an even mix of correct and incorrect answers, showing that they can distinguish between students who grasp the subject and those who do not. (Kustati & Amelia, 2024: 6956). Secondly, the ability of the questions to distinguish between students is quite significant. Questions that do not effectively differentiate skills result in a less varied range of responses among students, making the answers more evenly spread out (Mahendra & Rahayu, 2019: 6). Thirdly, how clear the question is plays a key role. Questions that are vague or not straightforward can lead to misunderstanding, resulting in respondents providing random answers, making the spread of responses nearly uniform (Kustati & Amelia, 2024: 6957). Fourth, the effectiveness of the incorrect options in multiple-choice questions also has an impact. Choices that are weak or too obvious fail to mislead participants who are unfamiliar with the content, which in turn lowers the ability of the questions to differentiate between answers. (Arikunto, 2021: 248).

Conclusion

Based on the analysis done on the AKM instrument of Faraid material to measure students' reasoning ability with the help of *Ministep*, it can be concluded that the use of the AKM instrument in the context of Faraid material is very effective in measuring students' reasoning competence, especially in understanding and applying the concepts of Islamic inheritance law. AKM, which focuses on basic competencies such as reading literacy and numeracy, can be adapted to test students' ability to think logically and critically, which are important skills in solving inheritance distribution problems based on Faraid provisions. Faraid material, which requires an in-depth understanding and application of formulas and principles of Islamic inheritance law, is very relevant to be tested using the AKM instrument. The AKM instrument can help measure how students can analyze and solve inheritance distribution problems involving mathematical calculations and legal considerations. This shows that AKM is an assessment tool that measures factual knowledge and students' reasoning abilities in more complex contexts. Use *Ministep* in AKM instrument analysis has proven to be very helpful in increasing the validity and reliability of assessment instruments. With

the help of *Ministep*, an in-depth analysis can be carried out on the quality of the AKM questions, such as the difficulty level, question discrimination, and the extent to which the questions can accurately measure student abilities.

From the research can also be concluded that empirically from 10 questions there are 9 AKM questions related to Faraid material is said to be valid. The Alpha Cronbach value of these questions is also already reliable where the value is 0.84 but for the reliability of the respondent, it is still below the average of 0.39. Overall, the distractors in this question also functions well. This is indicated by 60% of the distractors work well. However, judging from the level of difficulty of the questions there are still many questions that are categorized as "too easy" than "difficult", namely with a percentage of 70% questions that are still considered too easy and 30% questions considered as currently.

Overall, applying AKM for Faraid material supported by the analysis *Ministep* can increase the effectiveness of educational evaluation in Islamic inheritance law. Using valid and reliable instruments, it is hoped that students will be better trained in critical and logical thinking and be able to apply knowledge of Islamic inheritance law practically. Thus, this research significantly contributes to developing better and more appropriate assessment instruments for measuring students' reasoning abilities in various complex learning contexts.

Research Limitation

Although this research provides valuable insight into the analysis of Faraid's material AKM instrument to measure students' reasoning ability with assistance *Ministep*, several limitations need to be considered, including the following: First, limitations on Faraid material. This research only focuses on Faraid material as an assessment context, which is limited to one field of study of Islamic law. Therefore, the results of this study may not be completely generalizable to other materials or disciplines. The second limitation is usage *Ministep*. Although *Ministep* is an effective tool for analyzing the quality of questions, limitations in access and understanding of this software can be a barrier for some researchers or educators unfamiliar with this analysis technique. Third, the research subjects are limited. This research was conducted on students with a limited sample size. This can affect the validity and representativeness of research results, especially if they are to be applied to a wider group or across various educational institutions.

Fourth, preparing questions is difficult. The writer of AKM questions on Faraid material must have a deep understanding of Faraid principles and the ability to design questions that test students' knowledge and critical reasoning. This process depends on the question developer's ability and can affect research results if the questions' quality is not optimal. The final limitation is the limited aspect of reasoning that is measured. Although the AKM instrument can measure students' reasoning abilities, this research has not fully explored other dimensions of reasoning, such as moral or ethical reasoning, which can also be an important aspect of Faraid learning.

Recommendation

Based on the research results and limitations identified, several recommendations for further development in this research are made. The first

recommendation concerns expanding the material. Future research can expand the scope of material analyzed using the AKM instrument to other fields of study. This could help determine whether the same approach is effective for assessing capabilities in other disciplines. Second, the use of ministeps by question developers. It is recommended that question developers be more intensive in studying and using *Ministep* to improve the quality of the questions prepared. Training and workshops regarding the use *of Ministep* for educators and researchers can increase the effectiveness of the analysis of assessment instruments.

Third, increasing the research sample. Research with a larger and more diverse sample covering a variety of educational institutions would provide more representative and reliable results. In addition, using a longitudinal approach to measure changes in students' reasoning abilities over time could also be a focus of future research. Fourth, focus on developing more varied questions. Researchers can focus more on developing questions that test knowledge and other aspects of reasoning, such as ethical, logical, and critical reasoning. This will enrich the student learning experience and improve the quality of assessments. Fifth, integration with other methods. To make the AKM instrument more holistic in measuring reasoning, integration with other assessment methods, such as project-based or portfolio assessments, is necessary to see how reasoning is applied in real-life contexts.

REFERENCES

Abdillah, M. A., & Anzaikhan, M. 2022 . Inheritance Property Distribution System in Islamic Law. *Al-Qadha : Journal of Islamic Law and Legislation*, *9*(1), 285–305. https://doi.org/10.32505/qadha.v9i1.4134

Adibah, I. S. 2017 . *MAWARIS FIQIH LEARNING AT BONDOWOSO STATE MADRASAH ALIYAH 2016/2017 ACADEMIC YEAR*.

Amalia, R. N., Dianingati, R. S., & Annisaa', E. 2022. The Influence of the Number of Respondents on the Validity and Reliability Test Results of the Self-Medication Knowledge and Behavior Questionnaire. *Generics: Journal of Research in Pharmacy*, *2*(1), 9–15. https://doi.org/10.14710/genres.v2i1.12271

Andikayana, D. M., Dantes, N., & Kertih, I. W. 2021. Development of Minimum Competency Assessment Instrument (Akm) Reading Literacy Level 2 for Grade 4 Elementary School Students. *Journal of Indonesian Educational Research and Evaluation*, *11*(2), 81–92.

Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. 2024 . Analysis of the Validity and Reliability of Final Odd Semester Summative Questions for PAI Subjects. *Educative : Journal of Educational Sciences*, *6*(1), 964–975. https://doi.org/10.31004/edukatif.v6i1.5931

Arikunto, S. 2021 .*Basics of Educational Evaluation 3rd Edition*. Literary Earth. https://books.google.co.id/books?id=j5EmEAAAQBAJ

Asihati, T., Hapsari, R., Madrasah, P., & Selatan, J. 2023 . Progressive of Cognitive and Ability Application of Numeracy Literacy in Islamic Religious Education (PAI) Lessons in Madrasas. *Edu.Pedia*, *xx*, *No.* Xx (xx), 499–504. https://doi.org/10.56855/jpr.v1i4.753

Azzahra, N. S., Sumarni, S., & Putranta, H. 2024 . Analysis of the Validity and Reliability of Quality Multiple Choice Questions for Minimum Competency Assessment (AKM) in Islamic Religious Education Subjects Using... *QuranicEdu: Journal of …*, *4*(1), 85–94. https://jurnalannur.ac.id/index.php/QuranicEdu/article/view/681

Basri, S. 2020 . Islamic Inheritance Law (Fara'Id) and its Application in Islamic Society. *Journal of Legal Certainty and Justice*, *1*(2), 37. https://doi.org/10.32502/khdk.v1i2.2591

Creswell, J., & Guetterman, T. 2018 . *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, 6th Edition*.

Farzad, M., Layeghi, F., Hosseini, A., Whiteneck, G., & Asgari, A. 2017 . Using the Rasch Model to Develop a Measure of Participation Capturing the Full Range of Participation Characteristics for the Patients with Hand Injuries. *Journal of Hand and Microsurgery*, *9*(2), 84–91. https://doi.org/10.1055/s-0037-1604060

Fatimah, L. U., & Al Fath, K. 2019 . Analysis of Question Difficulty, Discriminatory Power, and Distractor Functions. *Journal of Islamic Communication and Education*, *8*(12), 37–64.

Hadi, L. 2020 . Pros and Cons of the Independent Learning Policy. *Wahana Pendidikan Scientific Journal*, *6*(4). https://doi.org/10.5281/zenodo.4302861

Hadianto, F. 2024 . *Analysis of Islamic Religious Education Minimum Competency Assessment Question in Pai B Students Batch 2021 Uin Sunan Kalijaga*

Yogyakarta Introduction One of the evaluations in the education system, especially at the secondary level, is carried out using PISA (P. *5*, 88–102.

Hidayat, R., Ag, S., & Pd, M. 2019 . *Educational Science (Concepts, Theories, and Applications)* (C. Wijaya & Amiruddin (eds.)). Indonesian Educational Development Concern Institute (LPPPI).

Isnaeni, S., Fajriyah, L., Risky, E. S., Purwasih, R., & Hidayat, W. 2018 . Analysis of Mathematical Reasoning Ability and Learning Independence of Middle School Students on Straight Line Equation Material. *Journal of Medives : Journal of Mathematics Education IKIP Veteran Semarang*, *2*(1), 107. https://doi.org/10.31331/medives.v2i1.528

Kamulia, A. A. U., Tohir, M., Safitri, H. D. S. D., & Rosyidah, U. 2022 . Faraid's Problem-Solving Skills Judging from Students' Mathematical Abilities. *Kadikma*, *13*(1), 40–51.

Ketaren, M. A., Girsang, K., Manurung, M., & Riauati, E. 2024 . Validity Test and Differential Power Test of Artificial Multiple Choice Questions with Summative Tests for UPT SD Negeri 065013 Medan Class IV Students. *National Community Service Journal*, *1*(11), 3278–3283.

Kustati, M., & Amelia, R. 2024 . *Pelajaran Pai Di Sma Negeri 10 (Telaah Terhadap Reliabilitas , Daya Kesukaran Menggunakan Software Anates)*. *7*, 6954–6961.

Mahendra, E., & Rahayu, W. 2019 . ANALISIS BUTIR SOAL. *Workshop Peningkatan Kompetensi Evaluasi Pembelajaran GuruSLB N 1 Buleleng, Singaraja 4 - 7 September 2019*. https://doi.org/10.13140/RG.2.2.26498.71360

Mckinley, R. L. 2020 . An Introduction to Item Response Theory. *Measurement and Evaluation in Counseling and Development*, *22*(1), 37–57. https://doi.org/10.1080/07481756.1990.12022910

Ministry of Education and Culture. 2020 . *AKM and its Implications for Learning*. Center for Assessment and Learning, Research and Development Agency, and Ministry of Education and Culture books

Mukhid, A. 2021 . *Quantitative approach research methodology*, Jakad Media Publishing.

Murtafiaf, N. H. 2018. Educational evaluation. In *Jakarta: Rineka Cipta*.

Muslimin, M., & Sunardi, S. 2019 . Analysis of High School Students' Mathematical Reasoning Ability on Spatial Geometry Material. *Kreano, Journal of Creative-Innovative Mathematics*, *10*(2), 171–178. https://doi.org/10.15294/kreano.v10i2.18323

Muzayyanah, I. 2020. Analysis of Teacher-Made Multiple Choice PAI & BP Mid-Semester Assessment Questions (Comparison of SMPN 18 and SMPN 44 Semarang). *Eprints.Walisongo.Ac.Id*. https://eprints.walisongo.ac.id/15352/1/SKRIPSI_1603016019_INFI%27ATIN_MUZAYYANAH.pdf

Nashirulhaq, N., Nurzaelani, M. M., & Raini, Y. 2022. The Importance of Basic Literacy and Numeracy Skills at Middle School Level. *Proceedings…*, 118–122. http://pkm.uika-bogor.ac.id/index.php/PTP/article/download/1313/974

OECD. 2019. Programme for International Student Assessment Survey. In *OECD Data*.

Oktaviyanthi, R., & Agus, R. N. 2020. Evaluation Instrument for Students' Adaptive

Mathematical Reasoning Ability. *AKSIOMA: Journal of the Mathematics Education Study Program*, *9*(4), 1123. https://doi.org/10.24127/ajpm.v9i4.3150

Riyani, R., Maizora, S., & Hanifah, H. 2017. Validity test development of a test to measure relational understanding capability on quadrate equations material for class viii junior high school students. *Journal of School Mathematics Learning Research (JP2MS)*, *1*, 60–65. https://doi.org/10.33369/jp2ms.1.1.60-65

Rosita, E., Hidayat, W., & Yuliani, W. 2021. Validity and Reliability Test of the Prosocial Behavior Questionnaire. *FOKUS (Guidance & Counseling Study in Education)*, *4*(4), 279. https://doi.org/10.22460/fokus.v4i4.7413

Sani, R. A. 2021. *Akm-oriented learning: minimum competency assessment*. Literary Earth.

Saputra, O. A., & Sudrajat, A. 2024. Analysis of Chemistry Learning Outcome Assessment Instruments to Measure High Order Thinking Skill (Hots) Even Semester Class X Based on Rasch Modeling. *Didactics: Educational Journal*, *13*(2), 1967–1984.

Setiyawan, A. 2014. Factors Affecting Test Reliability. *Journal An-Nur*, *WE*(2), 341–354.

Sudijono, A. 2016. *Introduction to educational evaluation*. PT RajaGrafindo. https://books.google.co.id/books?id=6Y-TtQEACAAJ

Sugeng, B. 2022. *Fundamentals of quantitative (explanatory) research methodology*. Deepublish.

Sugiyono. 2020. *Quantitative, Qualitative, and R&D Research Methodologies*.

Suharto, B., Firdaus, A. A., Kristanto, D., Suharno, N. E., Uman Putra, D. F., & Kusuma, V. A. 2022. The influence of Airlangga University students' literacy skills in various fields of science. *Public Service Journal*, *6*(2).

Sulthan, M., Begum, I., Ismail, I., Yaakob, Z., Razick, A. S., Mahroof, M., & Abdullah, A. 2024. *Gender Equity in Muslim Family Law : Modern and Contemporary ' Ulamand . ' 34*(2), 221–256. https://doi.org/10.21580/ahkam.2024.34.2.20773

Sumintono, B., & Widhiarso, W. 2015. *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*.

Taufiqi, F., & Ellianawati. 2023. Developing an Online-Based Evaluation Instrument Using a STREAM Approach on Dynamic Fluid Material to Measure Students' Reasoning Ability. *Unnes Physics Education Journal*, *12*(1), 14–21. http://journal.unnes.ac.id/sju/index.php/upej

Untary, H., & Risdianto, E. 2020. *Analisis data penelitian dengan model Rasch dan Winstep*. Halaman Moeka Publishing. http://ebooks.uinsyahada.ac.id/index.php?p=show_detail&id=72

Yasinta, Nurdiana, R., & Asmah, S. N. 2024. Analysis of students' mathematical reasoning abilities through numeric literacy takes questions. *Journal of Mathematics and Natural Sciences*, *4*(3), 21–32.

Zuleika, A., & Desinthya, N. P. 2015. Islamic Inheritance Law (Faraid) and Its Economic Implication. *Tazkia Islamic Finance and Business Review*, *8*(1), 6.