Open Access
Bulletin of Indonesian Islamic Education Research
Vol. 1, No. 1, June 2025
https://attractivejournal.com/index.php/ier

The Ability of Islamic Education Teachers in Developing Assessment Instruments to Enhance Students' Critical and Creative Thinking Skills at Madrasah Aliyah Negeri Yogyakarta

Moh. Solikul Hadi[1*], Muhammad Hijazi[2], As'ari Latif[3], Mohamad Ardin Suwandi[4], Mega Primaningtyas[5]
[1] *Sekolah Tinggi Agama Islam Masjid Syuhada Yogyakarta, Indonesia*
[2] *Universitas Al-Azhar Cairo, Egypt*
[3] *Stricting Azzam Aliyah, Suriname*
[4] *Tomsk Polytechnic University, Rusia*

Email: mohsolikulhadi@gmail.com

ARTICLE INFO
Article history:
Received
January 28, 2025
Revised
March 19, 2025
Accepted
June 20, 2025

Abstract

This study aims to analyze the ability of Islamic Religious Education (PAI) teachers in developing assessment instruments that can improve students' critical and creative thinking skills at the State Islamic Senior High School (MAN) Yogyakarta. This study uses a quantitative approach with a descriptive survey method. Data were collected through a closed questionnaire with a Likert scale distributed to 35 PAI teachers in the MAN Yogyakarta environment, as well as documentation of examples of assessment instruments that they have used. Data analysis was carried out using descriptive statistics and Pearson correlation tests to see the relationship between teachers' abilities in compiling instruments and the results of improving students' skills. The results of the study showed that quantitatively, PAI teachers have moderate to high abilities in developing assessment instruments, with an average score in the "quite good" category. The instruments developed generally contain indicators of critical and creative thinking, although limitations were still found in the aspects of question innovation and variations in assessment forms. The correlation test showed a significant and positive relationship ($r = 0.63$, $p < 0.05$) between the quality of the instrument and the improvement of students' thinking skills, based on quantitative data from student assessment results and teacher perceptions. In conclusion, the quantitative results of the study support that the ability of Islamic Religious Education teachers in compiling assessment instruments has an important role in improving students' critical and creative thinking skills. The implications of this study underline the importance of ongoing training in the development of assessment instruments based on high-level thinking skills to support more effective and transformative Islamic Religious Education learning in madrasahs

Keywords: Assessment Instruments, Critical and Creative Thinking, Islamic Religious Education Teachers

Published by CV. Creative Tugu Pena
Website https://attractivejournal.com/index.php/ier
ISSN xxxx-xxxx

INTRODUCTION

Islamic Religious Education (PAI) in madrasas has a very important role in forming the character of students, not only in terms of spirituality and morality, but also in terms of critical and creative thinking skills (Budiman et al., 2021). In the modern era marked by rapid developments in science and technology, the challenges faced by students are no longer sufficient to be faced by memorizing religious material alone (Güven, 2010). They need to be equipped with the ability to analyze problems, evaluate various information, and create innovative solutions in accordance with Islamic values. Therefore, Islamic Religious Education learning must be directed to encourage the birth of students who are able to think critically and creatively, in accordance with the demands of the times (Siregar, 2021) .

One important way to achieve this goal is through the preparation of appropriate assessment instruments. Assessment instruments are not just tools to measure learning outcomes, but also function as a means to develop students' ways of thinking (Lubis et al., 2010). Islamic Religious Education teachers have a central role in compiling these instruments, so that the teacher's ability to design assessments becomes a key factor in determining the success of the learning process (Moh. Solikul Hadi, 2024). Unfortunately, in reality, there are still many teachers who do not fully understand how to develop instruments that are able to explore the high-level thinking potential of students, such as critical and creative thinking (Hadi, 2018) . Many assessment instruments are still conventional, tending to only assess low-level cognitive aspects, such as remembering and understanding (Hadi et al., 2024) .

The State Islamic Senior High School (MAN) in Yogyakarta as an Islamic educational institution that has a good reputation and is a reference for other Islamic schools, of course has a responsibility to continue to improve the quality of its education. (Hadi et al., 2020) . The ability of Islamic Religious Education teachers at MAN Yogyakarta in compiling assessment instruments that encourage the development of critical and creative thinking in students is an important aspect that needs further research. (Solikul Hadi, nd) . Thus, this study is relevant to explore the extent of the competence of Islamic Religious Education teachers in designing assessment instruments that are not only valid and reliable, but also capable of being a driver for the birth of a generation that thinks progressively, provides solutions, and is innovative within the framework of Islamic values. (Moh. Solikul Hadi, Muhammad Nuril Anam, 2021) . This study aims to describe in depth the ability of Islamic Religious Education teachers at Yogyakarta State Islamic Senior High School in compiling assessment instruments that support the development of critical and creative thinking skills of students and to identify the forms of assessment instruments that have been used that can contribute to forming the character of students who are able to think logically, analytically, and creatively in dealing with everyday life problems. (Fahmi Husein, Moh Solikul Hadi, 2024) .

Several previous studies have shown that teachers' ability to develop assessment instruments has a significant role in improving the quality of learning and cognitive development of students. Research conducted by Sari (2020) shows that teachers' competence in developing assessment instruments based on Higher Order Thinking Skills (HOTS) has a positive influence on student learning outcomes (Daryanes et al., 2023) . The study concluded that teachers who understand the principles of assessment that require high-level thinking can create a more challenging and meaningful learning atmosphere for students. Furthermore, research by Maulida regarding the implementation of authentic

assessment in Islamic Religious Education learning shows that instruments such as project assessments, portfolios, and case studies encourage students to think more deeply, explore new ideas, and develop their argumentative skills (Nikmah & Toha, 2019) . This type of assessment is considered more effective than conventional forms of assessment that only assess memorization skills. Meanwhile, Nurhadi in his research on the relationship between 21st century skills assessment and student creativity found that the use of assessment instruments that lead to solving real problems, creating original works, and presenting ideas openly, can increase students' creativity and flexibility of thinking (Yasin, 2020) . These studies strengthen the assumption that the preparation of good assessment instruments by Islamic Religious Education teachers is an integral part of efforts to improve students' critical and creative thinking skills (Muhammad Jihan Khopia, Asep Tutun Usman, 2024) .

Although there have been many studies discussing the importance of teachers' ability in compiling assessment instruments and their influence on learning outcomes and the development of students' thinking skills, this study has a different focus and approach. Previous studies such as those conducted by Sari, Maulida, and Nurhadi tend to examine general aspects of HOTS-based or authentic assessments in a broad learning context, without specifically highlighting Islamic Religious Education (PAI) subjects and without directly linking them to critical and creative thinking skills as two adjacent competencies. This study specifically places PAI teachers as the main subject of study and students' critical and creative thinking skills as variables that need to be improved through the role of teachers in compiling assessment instruments (Mega Aris Saputra et al., 2024) . This study not only analyzes the form and quality of the assessment instruments used, but also traces the functional relationship between teachers' abilities in compiling these instruments and improving students' critical and creative thinking skills (Smuts & Smuts, 2022) , especially in the Yogyakarta State Madrasah Aliyah environment which has its own characteristics and academic culture (Aref et al., 2023) .

METHOD

This study uses quantitative research methods. Quantitative research can be interpreted as a research method based on the philosophy of positivism, used to research a particular population or sample. (Sugiono, 2014) . Sampling techniques are generally carried out randomly, data collection uses research instruments, data analysis is quantitative statistical in nature with the aim of testing the established hypothesis. (Nurlan, 2019) . The research subjects used were students at the Yogyakarta State Islamic Senior High School. (Santoso & Madiistriyatno, 2021) . The researcher used *purposive sampling and snowball sampling* techniques . The data collection was carried out using questionnaires, structured observations, structured interviews and tests. After all the data obtained and then analyzed using the validity test instrument quality analysis technique using Product Moment Correlation, Reliability Test, Difficulty Level, discriminatory power, as well as the effectiveness of the deceptive question/instrument (Mackiewicz, 2018)

RESULT AND DISCUSSION

CRITICAL AND CREATIVE THINKING

Ennis said that critical thinking is thinking in a reasonable and reflective manner with an emphasis on making decisions about something that must be believed or done (Afida, 2023) . In contrast to the view expressed by Wijaya, who said that critical thinking is the ability to think that is more directed at analyzing ideas or ideas that are more specific, identifying, reviewing, and developing them into more perfect ideas (Oey-

Gardiner et al., 2007) . The potential for a person to be creative is indeed bestowed on everyone, but this potential still needs to be developed, not just left alone, so that creative potential can develop optimally (Hasnida et al., 2024) . By optimizing creative potential, it is hoped that it will be able to produce something new, which is valuable and useful for many people.

According to Chen, creative thinking is a way of thinking that produces something new, which comes from a previously existing idea, then developed to produce something new. Meanwhile, according to Riyanto, creativity is a process that demands balance and application of the three essential aspects of analytical, creative, and practical intelligence, where these three aspects, if used together, give birth to success, thus giving rise to something that has never existed before. (Isnaen & Albastiah, 2021) . The basis of creativity involves many components that produce creative output. These components include that creative thinking involves aesthetics and practical standards, relies on attention to goals and outcomes, relies more on mobility, and is not only objective but also subjective, and relies more on intrinsic motivation than extrinsic motivation (Munir & Zumrotus Su'ada, 2024) .

HOTS and LOTS Question Assessment Instruments

Questions that measure memory do not provide enough encouragement for students to study harder in preparing themselves to become creative members of society in the future. (Güven, 2010) . Therefore, students need to be given questions that require high-level thinking processes (*higher order thinking skills* or HOTS). (Meltzer, 2002) . The existence of questions that require high-level thinking processes makes students able to apply their knowledge in everyday life. He always questions everything he knows, both intentionally and unintentionally. It should be noted that questions that require students to think highly will make students more prepared to face various problems in the future. (Smuts & Smuts, 2022) .

To construct questions that measure high-level thinking processes, various information is presented, usually in the form of stimuli. (Mega Aris Saputra et al., 2024) . Stimulus can be in the form of text, images, graphs, tables, and so on that contain information from real life. The stimulus used should be interesting, meaning it encourages students to read. (Aref et al., 2023) . Based on this information, students are asked to transferring information from one context to another , processing and applying information , seeing the relationships between different information , using information to solve problems , critically examining/reviewing ideas or concepts and information (De Felice et al., 2021) . In the high-level thinking process, students demonstrate an understanding of information and reason, not just remembering or *recalling.* (D'Angelo et al., 2023) . The subject matter needs to provide the information needed to answer the question and students demonstrate understanding of and information and/or manipulate or use the information. Questions that are higher order thinking in nature do not always have to be more difficult, for example determining the meaning of a very rarely used word is not yet included in HOT (Haak, nd) . Difficult questions do not necessarily mean *higher order thinking skills,* unless they involve reasoning to find the meaning of words from a context or stimulus. (Ajmain et al., 2019) . In principle, *higher order thinking* is a way of thinking logically or a reasoning process. Assessments that focus on *higher order thinking* include questions and answers , exploration and analysis , reasoning when obtaining information, not remembering it , solving, criticizing, and translating , cognitive processes that are measured include analysis, synthesis, and evaluation. (Alhumaid, 2023) .

Writing reasoning questions, the question writer is required to be able to determine the behavior to be measured and the description of the material that will be formulated into a stimulus in a certain context according to the expected behavior. This stimulus will be used as a basis for making questions. (Uyuni & Adnan, 2020) . The description of the material to be asked (which is appropriate for reasoning questions) is

not always available in textbooks. Therefore, in writing reasoning questions, mastery of the material and creativity in writing questions are required. Because the questions are written referring to the indicators contained in the grid, the formulation of the indicators must also lead to questions that require reasoning. (NurArif Farhan, Wahyu Firman Syah, Abdul Khobir, 2020) . According to Bloom, Kratwhwol, & Anderson, that the level of thinking of students in thinking there are six levels, namely remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). The levels of thinking at C1, C2, and C3 are low-level thinking levels (*Low Order Thinking*) and the levels of thinking at C4, C5, and C6 are high-level thinking levels (*Higher Order Thinking*). Thus, *Low Order Thinking* is a way of thinking of students at a low level which is only limited to knowledge of C1-C3. (Oktaviana et al., 2022) .

The Effectiveness of Research Instruments in Developing Students' Critical and Creative Powers

Validity Test

Testing to determine the validity of the statement items is done by comparing the calculated *r product moment value* and then consulting with *the product moment* table *r value* with the criteria that if the calculated *r value* is equal to or greater than the table *r value* , the question is declared valid. To see the table *r value* , it is necessary to first find the degree of freedom (*degree of freedom*) or abbreviated as df with the formula: df = N-nr, where N is the number of test participants (*testees)* and nr is the number of correlated variables. So in this case, the number of questions (N) is 40, while the number of correlated variables (nr). Thus df = 40-2 = 38, while the level of significance for the validity test generally used is 5%. With df = 38 and a significance level of 5%, *the r* table is 0.320. Next, we compare the calculated *r value* with *the r* table. The results of the instrument validity test can be seen in the following table . The results of the validation of the PAS questions for semester 1 class X MANPK Yogyakarta for the Fiqh subject in the 2017/2018 academic year contained 40 questions. Based on the calculation results above, 17 questions were declared valid with a percentage of 42.5%, namely question numbers 3, 6, 10, 12, 14, 16, 18, 19, 23, 24, 26, 29, 31, 33, 36, 37, 38. Then there were 23 questions declared invalid with a percentage of 57.5%, namely 1, 2, 4, 5, 7, 8, 9, 11, 13, 15, 17, 20, 21, 22, 25, 27, 28, 30, 32,34, 35, 39, 40.

Reliability

Reliability testing The validation results of the PAS questions for semester 1 class X MANPK Yogyakarta for the Fiqh subject in the 2017/2018 academic year were carried out manually with the help of the *Excel program* and using the KR 20 formula. The results of the calculation were then interpreted with the criteria if ≥ 0.70 then the question can be said to have high reliability. The following is the calculation of the reliability level:

$$r_{11} = \left(\frac{n}{n-1}\right)\left(\frac{S^2 - \sum pq}{S^2}\right)$$

S^2 is the square of the standard deviation, the following is the calculation of the standard deviation.

$$SD = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2}$$

$$= \sqrt{\frac{48482}{81} - \left(\frac{1944}{81}\right)^2}$$

$$= \sqrt{22{,}5432}$$

$$= 4{,}7476$$

$S2 = SD2 \quad = 4{,}7476^2$

$= 22{,}543$

Then the proportion of students who answered correctly is multiplied by the proportion of students who answered incorrectly (data attached). Results $\sum pq = 7{,}6485$

After all the components of the formula for finding the level of reliability are known, here are the calculation results.

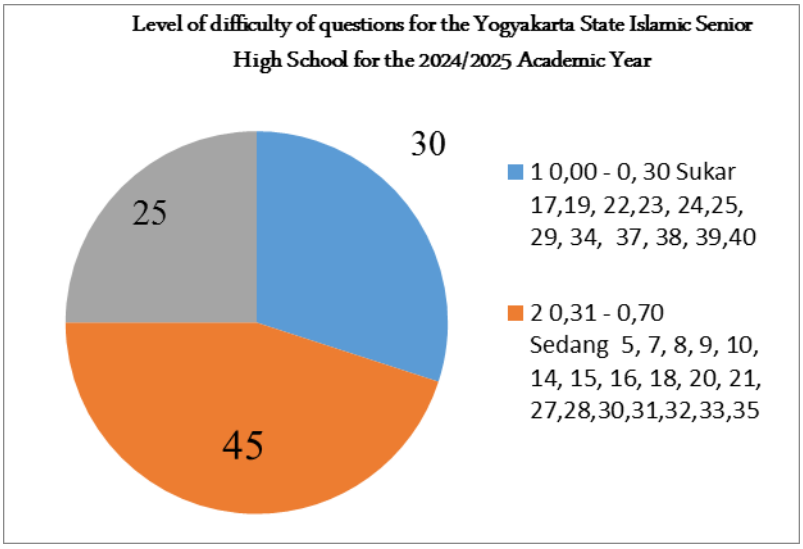
$$r_{11} = \left(\frac{40}{40-1}\right)\left(\frac{22{,}543-7{,}6485}{22{,}543}\right) = 0{,}6776$$

Based on the results of manual calculations, it is known that the question has a reliability of 0.6776 , so it can be concluded that the PAS questions for the first semester of class X MANPK Yogyakarta for the 2017/2018 Academic Year have a low level of reliability.

Analysis of the Difficulty Level of the questions

The calculation of the difficulty index is done by providing an interpretation of the calculation results (Abrori & Hadi, 2020) . The way to provide an interpretation is by consulting the results of the calculation of the difficulty level index with a benchmark or criteria as follows (Nana Sudjana): questions with P 0.00 to 0.30 are questions that are classified as difficult; questions with P 0.31 to 0.70 are questions that are classified as moderate; and questions with P 0.71 to 1.00 are questions that are classified as easy. Based on the calculation results, there are 12 questions or 30% that are included in difficult questions, 18 questions or 45% are classified as moderate questions, and 10 questions or 25% are included in easy questions. The following is a description of the questions based on the level or index of difficulty .

Figure 1.1 Distribution of Questions for Yogyakarta State Islamic Senior High School in the 2024/2025 Academic Year



Distinguishing Power

In the calculation of discriminatory power using the help of the *Excel program* . The number of subjects in this study was 81 students, so it is included in the small group. In the small group to calculate the discriminatory power first divided into 50% upper group and 50% lower group. The results of the calculation of discriminatory power are interpreted into five criteria, namely :
D: 0.00 – 0.20 then the discriminatory power of the question item is poor
D: 0.21 – 0.40 means the discriminatory power of the question is sufficient
D: 0.41 – 0.70 means the discriminatory power of the question is good
D: 0.71 – 1.00 means the discriminatory power of the question item is very good.
D: negative means the discriminatory power of the question is not good and is discarded.
The discriminatory power is calculated using the following formula.

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

The number of the upper group is 24 students and the lower group is also 24 students (data attached). The way to find out the discriminating power is by subtracting the results of the proportion of students who answered correctly in the upper group divided by the number of students in the upper group with the results of the proportion of students who answered correctly in the lower group divided by the number of students in the lower group. The following are the results of the calculation of discriminating power.

Based on the results of the calculation of discriminatory power with the help of the *Excel program* , it shows that 14 questions or 35% have poor discriminatory power, 14 questions or 35% have sufficient discriminatory power, 4 questions or 10% have good discriminatory power, while 8 questions with negative values of 20% have poor discriminatory power and must be discarded.

Table 1. ~~1.~~ Distribution of PAS Questions for Yogyakarta State Islamic Senior High School Academic Year 20 24 /20 25

Based on Distinguishing Power

No	Distinguishing Power	No. Question	Amount	Presentation
1	0.00 – 0.20 Bad	1, 2, 6, 13, 15, 21, 27, 28, 31, 32, 35, 37, 38, 40	14	35%
2	0.21 – 0.40 Enough	4, 5, 8, 11, 16, 17, 18,20, 25, 26, 29, 33, 34, 36	14	35%
3	0.41 – 0.70 Good	12, 19, 22, 39	4	10%
4	0.71 – 1.00 Very well	-	-	-
5	D negative Ugly as hell	3, 7, 9, 10, 14, 23, 24, 30	8	20%

, it can be concluded that overall, the PAS questions for the Yogyakarta State Islamic Senior High School for the 2024/2025 Academic Year has poor distinguishing power, and some questions are even considered to be of poor value and must be discarded because they are not suitable.

Distractor Function Analysis

The calculation of the effectiveness of the distractor was carried out using the Anates version 4 program. The calculation results for each distractor were interpreted into five criteria, namely:

IP = 76% - 125% means very good,
IP = 51% - 75% or 126% -150% means good,
IP = 26% - 50% or 151% - 175% means not good,
IP = 0% - 25% or 176% - 200% means bad, and
IP = more than 200% means very bad.

The assessment of the effectiveness of distractors on each question item uses criteria adapted from the *Likert scale* as follows.

a) If the four distractor answers to the question can function well, then the question can be said to have very good distractor effectiveness.
b) If there are three distractor answers that work, then the question is said to have good distractor effectiveness.
c) If there are two distractor answers that work, then the question is said to have quite good distractor effectiveness.
d) If there is only one distractor answer that works, then the question is said to have poor distractor effectiveness.
e) If all the distractor answers do not work, then the question is said to have poor distractor effectiveness.

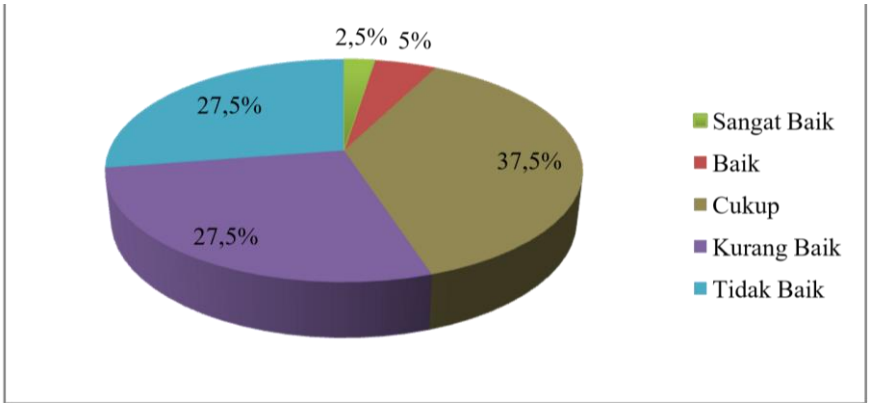
Based on the analysis results, there is 1 question item or 2.5% that has very good distractor effectiveness, 2 questions or 5% have good distractor effectiveness, 15 questions or 37.5% have sufficient distractors, 11 questions or 27.5% have less good and bad distractors. The following is a description of the questions based on the distractor effectiveness.

Table 1. 2. Distribution of PAS Questions for Yogyakarta State Islamic Senior High School Academic Year 20 24 /20 25

Based on the effectiveness of the distractor function

No	Effectiveness of the Deception	Item No.	Amount	Presentation
1	Very good	9	1	2.50%
2	Good	15, 29	2	5%
3	Enough	3, 4, 5, 10, 12, 16, 18, 22, 31, 32, 34, 35, 36, 37, 40	15	37.5%
4	Not good	2, 7, 14, 17, 21, 23, 24, 25, 26, 27, 30	11	27.5%
5	Not good	1, 6, 8, 11, 13, 19, 20,28, 33, 38, 39	11	27.5%

Figure 1.2 Distribution of distractor functions for PAS questions for Yogyakarta State Islamic Senior High School in the 2024/2025 academic year .



Quality of Question Items

Interpretation of the quality of the test items is adapted from the *Likert scale* as follows.

a. If the question meets the four criteria of a good question, namely validity, level of difficulty, discriminatory power, and distractor effectiveness, then the question can be said to be a very good question and can be stored in the question bank.
b. If the question meets three of the four criteria of a good question, namely validity, level of difficulty, discriminatory power, and distractor effectiveness, then the question is said to be a good question and cannot be stored in the question bank. The question needs to be revised until it meets the four criteria.
c. If the question meets two of the four criteria of a good question, namely validity, level of difficulty, discriminatory power, and distractor effectiveness, then the question is said to be a moderate question and cannot be stored in the question bank. The question needs to be revised until it meets the four criteria.
d. If the question meets one of the four criteria of a good question, namely validity, level of difficulty, discriminatory power, and effectiveness of distractors, then the question is said to be a bad question and cannot be stored in the question bank. The question needs to be revised significantly so that it is better to discard the question or not store it in the question bank.
e. If the question does not meet the four criteria of a good question, namely validity, level of difficulty, discriminatory power, and distractor effectiveness, then the question is said to be a very bad question and cannot be stored in the question bank. The question requires significant revision so it is better to be discarded.
f. In addition to item-based requirements, the test as a whole must be reliable with the provisions explained previously.

Based on the analysis results, there were no very good quality questions, 2 good quality questions, 16 questions were included in the questions with moderate quality, 16 poor quality questions, and 6 questions with very poor

quality. The following is a description of the questions based on the quality of the questions.

Calculation , that overall the PAS questions of State Senior High Schools are not reliable where they do not meet the criteria of point f on the Likert scale. Based on the distribution of the quality of the questions, it explains that the presentation as a whole or in general is stated as not good and must be discarded or revised as a whole. The study at this point is to take action follow up on instrument effectiveness assessment in developing critical and creative power of students. Based on the results of data processing and analysis of the questions on the subject of fiqh class X MANPK Yogyakarta, it was found that overall the questions of PAS Madrasah Aliyah Negeri were not reliable where they did not meet the criteria of the point which reads " *In addition to item-based requirements, the test as a whole must be reliable with the provisions that have been explained previously on a Likert scale. "* Based on the distribution of the quality of the questions, it explains that the presentation as a whole or in general is stated to be not good and must be revised as a whole.

Referring to the criteria used by Anderson starting from C1-C6 which will be analyzed as follows:

Questions no . 1 , 2, 4, 11, 13, Using C1 and the question shows a very bad category, then the question is discarded. Whereas at the high school level C1 is not allowed, because the high school level has reached a minimum of C4. This is what makes students in Indonesia less creative and innovative.

Questions no. 3, 5, 6, 7, 17, 20, 21, 25, 26, 27, 28, 30, 32, 34, 35, 36, 40, show questions that are categorized as not good, so these questions are discarded. The provision requires that the questions used must be competent and meet the requirements. When viewed from the category that has been analyzed for the questions, there are 42.5% that are not good. This shows that the quality of teachers in compiling assessment instruments is very low.

Questions no. 8, 9, 10, 14, 15, 16, 18, 19, 22, 23, 24, 31, 33, 37, 38, 39 show questions that are in the moderate category, so the questions are updated. The provision requires that the questions used must be competent and meet the requirements. When viewed from the categories that have been analyzed for their questions, there are 40% that are moderate. It shows that the quality of teachers in compiling assessment instruments has not been able to stimulate students to have critical and creative powers.

Questions no. 12, and 29 Show questions that are categorized as good, then the questions are accepted. The provision requires that the questions used must be competent and meet the requirements. When viewed from the category that has been analyzed for the questions, there are 5% good. Showing the quality of teachers in compiling assessment instruments is still minimal in compiling assessment instruments.

The results of observations conducted on students show that students in the learning process activities do not behave critically and innovatively based on the fact that students only receive material without any deep curiosity. In addition, students are also not yet able to read real problems such as the many students' lack of awareness in disposing of garbage, being wise in using water and the inability to innovate in classroom learning. This data is reinforced by an interview with one of the teachers at the Yogyakarta State Madrasah Aliyah school . Thus, it shows that the effectiveness of the assessment instrument in developing students' critical and creative powers is still very low, as evidenced by only 5% being accepted, while 40% were improved and 55% were rejected, discarded and not accepted. Overall, the assessment instrument is not yet effective. Therefore, an evaluation is needed in making test items.

CONCLUSION

Based on the results of the data analysis above, it can be concluded that the teacher's ability to prepare assessment instruments is not in accordance with the desired competencies at the SMA/MA level, namely KI, KD and Indicators. Analysis of data calculations, there are no questions that are of very good quality, 2 questions of good quality, 16 questions are included in questions that have moderate quality, 16 questions of poor quality, and there are 6 questions that have very poor quality . The effectiveness of the preparation of the instrument is found to be less effective by seeing that many of the question items are not good and must be discarded and seeing that the operational words used in the preparation of the assessment instrument still use C1-C3. So the questions or instruments used are not effective in increasing students' critical and creative abilities. In addition, the results of observations show that students' critical and creative abilities in everyday life are still low. For example, when the imam cancels their prayer, they don't know what to do and they rarely ask about the reality in the field that occurs.

ACKNOWLEDGEMENT

The researcher would like to express his deepest gratitude to all parties who have provided support and contributions so that this research can be completed properly. Furthermore, the researcher would like to thank the Muhammadiyah University of Surakarta for providing support in the form of facilities and resources to carry out this research and also the researcher appreciates the support and guidance from the lecturers and academic staff at the university who have provided valuable insights and constructive input during the research process.

The researcher also thanks the authors and researchers whose work has become the main reference in this study. Without their work and contribution in the field of inclusive education, this study would not have been completed properly.

thank all parties who have contributed and supported in various forms. Hopefully the results of this study can provide a positive contribution to the development of inclusive education in Indonesia

REFERENCES

Abrori, M. S., & Hadi, M. S. (2020). Integral Values in Madrasah: to Foster Community Trust in Education. *Istawa : Jurnal Pendidikan Islam*, *5*(2), 160. https://doi.org/10.24269/ijpi.v5i2.2736

Afida, R. N. (2023). Literature Review : Peran Guru dalam Membangun Ketrampilan 4C Siswa dengan Pembelajaran Berdiferensiasi. *Prosiding Seminar Nasional Pascasarjana Universitas Negeri Semarang*, *1*(1), 643–647. http://pps.unnes.ac.id/pps2/prodi/prosiding-pascasarjana-unnes

Ajmain, M. T., Mahfuz, A. N. A., Rahman, S. N. H., & Mohamad, A. M. (2019). *Challenges of Islamic Education Teachers in Teaching*. *2*(1), 38–47.

Alhumaid, M. M. (2023). *Enhancing Graduate Students ' Attitudes and Self-Efficacy towards Inclusively Teaching Children with Disabilities Using an Inclusivity-Based Physical Activity Program*. *11*(5), 1158–1167. https://doi.org/10.13189/saj.2023.110525

Aref, E. E. M., Hameed, Z. A. A., & Mohamed, H. A. (2023). Cognitive functions in Arabic - speaking children with velopharyngeal insufficiency and their impact on speech intelligibility. *The Egyptian Journal of Otolaryngology*. https://doi.org/10.1186/s43163-023-00500-5

Budiman, A., Samani, M., & Setyawan, W. H. (2021). The Development of Direct-Contextual Learning: A New Model on Higher Education. *International Journal of Higher Education*, *10*(2), 15–26.

D'Angelo, S. C., Mache, J., & Guillén-Gosálbez, G. (2023). Absolute Sustainability Assessment of Flue Gas Valorization to Ammonia and Synthetic Natural Gas. *ACS Sustainable Chemistry and Engineering*. https://doi.org/10.1021/acssuschemeng.3c05246

Daryanes, F., Suzanti, F., Linggasari, M. N., Mahadi, I., Suwondo, S., Arnentis, A., & Sayuti, I. (2023). Pelatihan Penyusunan Soal Berbasis Hots (Higher Order Thinking Skills) Dengan Aplikasi Live Worksheet Untuk Meningkatkan Kompetensi Guru. *JMM (Jurnal Masyarakat Mandiri)*, *7*(1), 293. https://doi.org/10.31764/jmm.v7i1.12061

De Felice, F., Travaglioni, M., & Petrillo, A. (2021). Innovation trajectories for a society 5.0. *Data*, *6*(11). https://doi.org/10.3390/data6110115

Fahmi Husein, Moh Solikul Hadi, E. K. (2024). *Strategi Digital Detox untuk Siswa : Efektivitas Pendidikan Islam melalui Program Bimbingan Konseling pada Anak SMA / SMK di Yogyakarta Penerbangan AAG Adisutjipto , penerapan strategi Digital Detox melalui program*. *13*, 358–369.

Güven, M. (2010). An analysis of the vocational education undergraduate students' levels of assertiveness and problem-solving skills. *Procedia - Social and Behavioral Sciences*, *2*(2), 2064–2070. https://doi.org/10.1016/j.sbspro.2010.03.282

Haak, L. Van Der. (n.d.). *InterAttractive Social Robots : Exploring the acceptance and attitude towards social robot design in an online scenario-based experiment .*

Hadi, M. S. (2018). *Implementasi Model Pengembangan Multiple Intellegence Dalam Meningkatkan Kecerdasan Natural Peserta Didik Melalui Metode Project Based Learning Di Kelas X Manpk Yogyakarta Pada Mata Pelajaran Fikih*. 152. https://digilib.uin-suka.ac.id/id/eprint/33289/

Hadi, M. S., Abrori, M. S., & Zahra, D. N. (2020). Pengembangan Multimedia Interaktif Macromedia Flash Profesional 8 Pada Pembelajaran Al-Qur'an Hadis Kelas X Semester Genap Di Man 1 Yogyakarta. *At-Tajdid : Jurnal Pendidikan dan Pemikiran Islam*, *3*(2), 148. https://doi.org/10.24127/att.v3i2.1123

Hadi, M. S., Amrullah, A. K., & Anwar, K. (2024). *Journal of Islamic Scriptures in Non-Arabic Societies Membentuk Peserta Didik Menjadi Pribadi Ulul al-bab Prespektif*. *1*(1), 61–62.

Hasnida, S. S., Adrian, R., & Siagian, N. A. (2024). Transformasi Pendidikan di Era Digital. In *Jurnal Bintang Pendidikan Indonesia* (Vol. 1, Nomor 1). https://kumparan.com/aan-herdian89/transformasi-pendidikan-di-era-digital-1zG74Ilpzc4/4

Isnaen, F., & Albastiah, F. A. (2021). Pengaruh Kompetensi Sumber Daya Manusia, Pemanfaatan Teknologi Informasi,. *Al-Mal: Jurnal Akuntansi dan Keuangan Islam*, *5*, 139–153.

Lubis, M. A., Yunus, M. M., Embi, M. A., Sulaiman, S., & Mahamod, Z. (2010). Systematic steps in teaching and learning Islamic Education in the classroom. *Procedia - Social and Behavioral Sciences*, *7*(2), 665–670. https://doi.org/10.1016/j.sbspro.2010.10.090

Mackiewicz, J. (2018). A Mixed-Method Approach. In *Writing Center Talk over Time*. https://doi.org/10.4324/9780429469237-3

Mega Aris Saputra, N., Ramli, M., Yusuf Sobri, A., Madihie, A., & Khoirul Fitriyah, F. (2024). Digital Literacy as an Indicator of Adolescent Mental Health and Implementation for the Role of School Counselor: A Systematic Literature Review. *SEEJPH: South Eastern European Jurnal of Public Health*, *23*(3).

Meltzer, D. E. (2002). The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores. *American Journal of Physics*, *70*(12), 1259–1268. https://doi.org/10.1119/1.1514215

Moh. Solikul Hadi, Muhammad Nuril Anam, M. S. A. (2021). Reconstruction Of Martin Heiddeger's Thinking Existentialism Model on Education inThe Industrial Era 4.0. *Journal of Research in Islamic Education*, *03*(02), 47–58.

Moh. Solikul Hadi, D. (2024). Efektifitas Instrumen Penilaian Pendidikan Agama Islam Untuk Meningkatkan Kemampuan Berpikir Kritis Dan Kreatif Peserta Didik Di Smk

Muhammadiyah Mlati Yogyakarta. *Berkala Ilmiah Pendidikan*, *4*(1), 28–34.
Muhammad Jihan Khopia, Asep Tutun usman, N. M. (2024). *Ekstrakurikuler Di Pesantren Modern Sebagai Upaya Dalam Pembentukan Karakter Santri*. 1301–1316.
Munir, M., & Zumrotus Su'ada, I. (2024). Manajemen Pendidikan Islam di Era Digital: Transformasi dan Tantangan Implementasi Teknologi Pendidikan. *Journal of Islamic Education andManagement*, *5*(1), 1–13.
Nikmah, L., & Toha, M. (2019). Implementasi Penilaian Autentik Dalam Pembelajaran Pendidikan Agama Islam. *Journal of Islamic Education*, *2*(2), 102–109.
NurArif Farhan, Wahyu Firman Syah, Abdul Khobir, U. M. (2020). Microlearning Sebagai Strategi Pembelajaran Peningkatan Kemampuan Berpikir Siswa (SPPKB). *Jurnal Basicedu*, *5*(5), 3(2), 524–532. https://journal.uii.ac.id/ajie/article/view/971
Nurlan, F. (2019). Metodologi Penelitian Kuantitatif. In *ParePare: CV. Pilar Nusantara*.
Oey-Gardiner, M., Rahayu, S. I., Abdullah, M. A., Effendi, S., Darma, Y., Dartanto, T., & Aruan, C. D. D. P. A.-1. pdfaniel. D. (2007). *Era Disrupsi: Peluang Dan Tantangan Pendidikan Tinggi Indonesia*. https://www.aipi.or.id/index.php?pg=detilpublikasi&pid=67&type=2#reserach-tool-tab
Oktaviana, D., Barizi, A., & Yaqin, M. Z. N. (2022). The Challenges of The Society 5.0 Era: The Evaluation of Learning in Primary Education. *Al-Adzka: Jurnal Ilmiah Pendidikan Guru Madrasah Ibtidaiyah*, *12*(1), 10. https://doi.org/10.18592/aladzkapgmi.v12i1.5831
Santoso, I., & Madiistriyatno, H. (2021). METODOLOGI PENELITIAN KUANTITATIF. In *Universitas Nusantara PGRI Kediri* (Vol. 01). Indigo Media.
Siregar, M. (2021). Islamic Education Thought of Sheikh Nawawi Al-Bantani: The Disaster in The Qur'an. *Cendekia: Jurnal Kependidikan dan Kemasyarakatan*, *1*(1), 127–143. https://doi.org/10.21154/cendekia.v1i1.2194
Smuts, S., & Smuts, H. (2022). Society 5.0 and the future of work skills for software engineers and developers. *EPiC Series in Computing*, *84*, 169–182. https://doi.org/10.29007/9kzd
Solikul Hadi, M. (n.d.). *PENDIDIKAN AGAMA ISLAM DAN ISU-ISU ANTI KORUPSI (Telaah Materi Pendidikan Agama Islam Di Sekolah) ISLAMIC RELIGIOUS EDUCATION AND ANTI-CORRUPTION ISSUES (Study of Islamic Education Material in Schools)*. 1017–1030. https://www.voaindonesia.com/a/indeks-persepsi-korupsi-ri-
Sugiono. (2014). *Metode Penelitian Pendidikan Kualitatif, kuantitatif, dan R & D*. alfabeta.
Uyuni, B., & Adnan, M. (2020). The Challenge of Islamic Education in 21st Century. *SALAM: Jurnal Sosial dan Budaya Syar-i*, *7*(12), 1079–1098. https://doi.org/10.15408/sjsbs.v7i12.18291
Yasin, M. (2020). Pembelajaran Matematika Untuk Menghadapi Tantangan Abad 21. *ResearchGate*, *March*, 1–35.