EDUKASIA: Jurnal Pendidikan dan Pembelajaran
Vol. 5, 1 (June, 2024), pp. 75-82
ISSN: 2721-1150 EISSN: 2721-1169

Validation Analysis of HOTS-Based Multiple-Choice Questions in Islamic Religious Education at Al-Azhar High School Medan

Salim[1], Mohamad Aji Prasetia[2], Muhammad Bagas F[3], Muhammad Arif Fadhillah Lubis[4]

[1] Universitas Islam Negeri Sumatera Utara, Indonesia; salim@uinsu.ac.id
[2] Universitas Islam Negeri Sumatera Utara, Indonesia; mohamadaji12prasetia@gmail.com
[3] Sekolah Tinggi Ilmu Kesehatan Indah Medan, Indonesia; muhammadbagasf@stikes-indahmedan.ac.id
[4] Politeknik Negeri Medan, Indonesia; muhammadarif@polmed.ac.id

ARTICLE INFO

Keywords:

Validation;
Multiple choice questions;
HOTS;
Islamic education;

Article history:

Received 2023-10-13
Revised 2023-12-19
Accepted 2024-01-06

ABSTRACT

This study aims to determine the quality of multiple-choice questions based on Higher Order Thinking Skill's (HOTS) in Islamic Religious Education (PAI) Subjects for Class XI of Al-Azhar Medan High School in the 2022-2023 Academic Year. Content analysis is the technique used in this research. In this study, data analysis was conducted qualitatively. This means analyzing data from exam question documents on the subject of Islamic Religious Education, dividing the questions into categories C1 to C6, and then presenting the results of the data analysis in the form of percentages for each question category. The subject of this research is multiple-choice questions that have been made by the teacher of Islamic Religious Education class XI at SMA Al-Azhar Medan in the 2022–2023 academic year, totaling 35 questions. The data collection techniques used in this study are observation and analysis of data contained in question documents and answer keys. The results obtained from the analysis of the quality of items according to Bloom's taxonomy in the cognitive domain indicate that of the 35 multiple-choice questions found, 51.43% are in the HOTS categ multiple choice questions ory and 48.57% are included in the LOTS item category. Analysis of the quality of items in the aspect of difficulty level shows 32 questions (91.43%) in the medium category, 3 questions (8.57%) in the easy category, and 0 questions (0%) in the difficult category. Analysis of the quality of the items in the aspect of the differential power of the questions shows 31 questions amounting to (88.57%) good category, 3 questions (8.57%) very good category, and 1 question amounting to (2.86%) sufficient category.

Corresponding Author:
Salim
Universitas Islam Negeri Sumatera Utara, Indonesia; salim@uinsu.ac.id

1. INTRODUCTION

Nowadays, change is a necessity that cannot be avoided. These changes occur in all aspects of life, including education. Responding to these changes that occur can be done by making anticipatory efforts, including improving students' thinking skills through the learning process. Through High Order Thinking Skill (HOTS), students can achieve capabilities or abilities in accordance with the expectations of the times. (Maryanto, 2015) stated that through high-level thinking skills, or HOTS, students will be able to think critically, creatively, thoroughly, be able to solve problems and make decisions, and have good character.

According to the survey results of the Program for International Student Assessment (PISA) and Trends in International Match and Science Survey (TIMSS), Indonesian students have not been able to rank in the top position since their participation in 1999. Based on the survey, it shows that the majority of Indonesian students are still at the Lower Order Thinking Skill (LOTS) level. This indicates that students' science literacy is still low. There is still a lot of memorized material that is hoarded and is in the realm of short-term memory. Thinking skills still only tend to remember (recall), restate, or refer without processing (recite).

Along with the implementation of the independent curriculum, it is expected that there will be a paradigm shift in the implementation of learning. Learning, which was originally teacher-centered, has changed to be student-centered. That way, teachers are expected to be more creative and innovative in presenting subject matter. The application of several learning models, such as project-based learning, problem-based learning, learning with a problem-solving approach, and inquiry learning, is an opportunity for teachers to implement learning activities at the Higher Order Thinking Skill (HOTS) level.

The government has made several efforts to improve students' thinking skills. One of them is something new in the concept of the 2013 curriculum technical guidance, namely the application of Higher Order Thinking Skills (HOTS) in the implementation of Islamic Religious Education (PAI) assessment. HOTS is the ability to think beyond recall, restate, or refer without processing (recite). In principle, HOTS questions aim to measure students' abilities in: (1) transferring one concept to another; (2) processing and applying information; (3) finding links from different information; (4) using information to solve problems; and (5) critically examining ideas and information.

PAI learning in schools is one of the subjects that must be taught because of its very important position. The formation of students' attitudes and characters can be achieved through this PAI learning. To achieve this goal, it is necessary to integrate HOTS into the learning process. In Indonesia, Islamic education is placed in a strategic position. This can be seen in Law No. 20 of 2003 concerning the National Education System Article 3, namely: National Education functions to develop the ability and shape the character and civilization of a dignified nation in order to educate the nation's life and aims to develop the potential of students to become human beings who are faithful and devoted to God Almighty, noble, healthy, knowledgeable, capable, creative, independent, and democratic and responsible citizens.

Assessment using HOTS-based questions in PAI subjects has not been fully carried out by teachers, the questions developed only test students' low-level thinking skills. This can also be seen based on the results of research conducted (Ahmad, 2019) found that the problems that arise from the questions developed for the final exam do not fully have HOTS characteristics, and there are still many questions that only measure low-level thinking skills.

Referring to several studies shows the urgency of using HOTS in learning and assessment, among others: Publication (Asfiyah, 2021) which analyzes HOTS-based assessment in PAI subjects by applying descriptive qualitative methods and Milles and Huberman methods. The results showed that HOTS assessment can be carried out with a series of analyses applied to basic competencies, the preparation of sal grids, the selection of stimuli, the writing of questions, and the preparation of assessment rubrics. A series of applications of the assessment concept are dominated by cognitive aspects of making questions. The results of the HOTS-based assessment can be seen in the escalation of a higher level of

mastery of the material. In addition, the ability to process thoughts systematically. The ability to solve problems and confidence in making decisions are part of the results of HOTS-based assessments that are applied. The concept of HOTS assessment can be applied with the aim of stimulating students' critical-creative thinking skills.

Furthermore, research conducted by (Baharun & Sa'diyah, 2018), focused on HOTS-oriented class-based assessment based on Bloom's taxonomy to streamline PAI learning. The results of the study are oriented towards the ability of students to think creatively, critically, innovatively, argue, solve problems with action options, and have expertise. So that the results of the applied assessment can be used as reference material or as a reference for improvement in the PAI learning process.

Other relevant research conducted by (Feriyanto & Putri, 2020) is oriented towards the development of teaching modules in mathematics lessons that support the implementation of HOTS with the development of the Tessmer model. The development process begins with literature research on literacy-based math modules and HOTS questions. The next step is self-evaluation, which consists of analysis and design and looks at students as research subjects and the materials used in the study. In addition to designing the prototype, the materials used included mathematical induction, linear programs, geometric transformations, and matrices. The results of the study produced a math module based on literacy and HOTS questions that aimed to improve critical thinking skills. This module meets the validity criteria for the valid category and the practicality criteria for the very suitable category.

Based on preliminary observations of PAI teachers at SMA plus Al Azhar Medan, the following problems were found: (1) PAI exam questions prepared by teachers are obtained from PAI textbooks and LKS; (2) PAI exam questions have never been analyzed; (3) PAI exam questions have not been fully prepared based on HOTS (High Order Thinking Skill); and (4) based on mid-semester test data obtained from PAI teachers, the PAI questions given to students are still in the LOTS category, which emphasizes more on the aspects of knowledge, understanding, and application.

2. METHOD

This type of research is qualitative research with descriptive research methods. In the process, this research seeks to describe the content validity and accuracy of HOTS-based multiple-choice questions in Islamic Religious Education Subjects in class XI of Al-Azhar Medan High School in 2022-2023. The sample of this study is a HOTS-based multiple-choice answer sheet distributed to XII grade students of SMA Al-Azhar Medan, while what is analyzed is 35 multiple-choice questions. Data analysis techniques are data reduction, data presentation, and conclusion drawing; data validity checking techniques include credibility, transferability, dependability, and confirmability.

3. FINDINGS AND DISCUSSION

Content Validity

In this validity show from 35 questions, it can be seen that questions 3 and 28 are not worth testing based on material aspects because the questions are not in accordance with material competencies and do not have answers. The research findings show from 35 questions based on construction analysis that all questions are worth testing, but there are several aspects that are not fulfilled such as: (1) the main aspect of the question does not give clues to the answer key contained in the main question number 1 is the same as the answer key, (2) the aspect of the length of the answer choices is relatively the same in questions number 2, 4, 19 and 24 have answer choices that are not the same length as the answer key choices with one another, (3) almost all questions do not meet the construction feasibility aspects of points 7, 9 and 10 because these questions are not equipped with pictures, diagrams except question number 7, do not use the statement "all of the answers above are wrong / correct", and the question items depend on the answers to the previous questions. Language feasibility analysis There are several questions that do not meet the criteria, such as: (1) using language in accordance with Indonesian language rules, not fulfilled in questions 2, 3, 4, 9, 11, 12, 14, 15, 16, 18, 19, 22, 26, and 28; and (2) answer choices that do not repeat the

same word or group of words unless it is a unity of understanding, not fulfilled in questions 1, 4, 6, 14, 14, 23, 24, 30, and 33.

Question Item Validity

In the results of testing the validity of the questions from 35 test items, there was 1 test item that was canceled, namely item number 31. Thus, the number of eligible test items amounted to 34. Validity testing can be done using product moment correlation, as shown in Equation (1).

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2\}}} \quad (1)$$

Description:

r_{xy} = Product moment correlation
N = Number of research subjects
X = Subject score on an item
Y = Total subject-score item score

Furthermore, the results of testing the reliability of the questions showed a reliability coefficient of 0.959. Furthermore, with reference to Surapranata (2004), an instrument is said to be reliable if the coefficient is ≥ 0.70. Thus, the question is reliable because the reliability coefficient is 0.959 ≥ 0.70 (Suharsimi, 2014). The following is the reliability formula using Cronbach's alpha, as shown in Equation (2).

$$r_{hitung} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2}\right) \quad (2)$$

Description:

r_{hitung} = Instrument Reliability
k = Number of items
σ_t^2 = Total variance
$\sum \sigma_i^2$ = sum of the variances of the scores of each item

The results of testing the level of difficulty of the questions were 3 easy category questions, 32 medium category questions, and 0 difficult category questions. Testing the level of difficulty of the questions can be done using Equation (3). While the proportional table can be seen in Table 1,

$$P = \frac{B}{JS} \quad 3)$$

Description:

P = Proportion of correct answers or difficulty level
B = Number of participants answered correctly
Js = Number of students taking the test

Table 1. Proportionality and Interpretation Table

Size P	Interpretation
$P < 0{,}30$	Too difficult
$0{,}30 \leq P < 0{,}70$	Fair (medium)
$P \geq 0{,}70$	Too easy

Differential power analysis is used to evaluate the test's ability to differentiate students in the upper and lower groups. Discriminatory power refers to the ability of a question to differentiate students in the

upper and lower groups based on their learning outcomes. The results of various power analyses conducted on questions or questions given to students. The analysis of the question's differential power can be determined using Equation (4), and the classification of the question's differential power index can be seen in Table 2 (Pradipta & Kurniawan, 2023).

$D_B = P_A - P_B$ (4)

Description:

P_A = level of difficulty in the upper group

P_B = level of difficulty in the lower group

Table 2. Proportionality and Interpretation

Classification of Differentiation Index of Question	Index of discriminating power	Classification
1	0,0 – 0,19	Bad
2	0,20 – 0,39	Fair
3	0,40 - 0,69	Good
4	0,70 – 1,00	Very good

The results of testing the differentiation of questions from 35 test items show that there is 1 test item in the sufficient category, 31 test items in the good category, and 3 test items in the excellent category. In the results of testing question distractors from 35 test items, there are 2 invalid test items, namely item 3 and item 28. Distractors that work on 25 questions, or 71.43%, and distractors that do not work on 10 questions, or 28.57%.

Based on Bloom's taxonomy analysis of 35 multiple-choice questions, there are 17 questions (48.6%) including the LOTS question category, namely question numbers 1, 2, 4, 5, 6, 8, 14, 17, 20, 23, 24, 25, 30, 31, 32, 33, and 35. In the LOTS question category, the question criteria are spread at each level with the following details: C_1 (remembering) = 8 questions (22.8%): questions 1, 4, 23, 24, 25, 30, 33, and 35. C_2 (understanding) = 3 questions (8.7%) questions number 8, 31, and 32; C_3 (applying) = 6 questions (17.1%) questions number 2, 5, 6, 14, 17, and 20. Furthermore, 18 (51.4%) questions belong to the HOTS question category. This shows that half of the total questions are HOTS questions, namely question numbers 3, 7, 9, 10, 11, 12, 13, 15, 16, 18, 19, 21, 22, 26, 27, 28, 29, and 34, but all of them are in the C_4 (analysis) criteria = 18 questions (51.4%), so there is no variation in HOTS-based preparation. A summary of the results of the classification of LOTS and HOTS-based question types can be seen in Table 1.

Table 3. Classification of Question Types Based on LOTS and HOTS

Criteria	Number of Questions	Number of Questions Items	Percentage (%)	Description
C_1 (remembering)	8	1, 4, 23, 24, 25, 30, 33, 35	22,8	LOTS
C_2 (understanding)	3	8, 31, 32	8,7	LOTS
C_3 (applying)	6	2, 5, 6, 14, 17, 20	17,1	LOTS
C_4 (Analysis)	18	3, 7, 9, 10, 11, 12, 13, 15, 16, 18, 19, 21, 22, 26, 27, 28, 29, 34	51,4	HOTS

According to the data above, the questions created by the teacher's level of difficulty showed 91.43% (32 questions) in the medium category and 8.57% (3 questions) in the easy category. When referring to the explanation (Suharsimi, 2014), a good question is one that is neither too easy nor too difficult. Questions that are too easy do not stimulate students to increase their efforts to solve them. Conversely, questions

that are too difficult will cause students to become discouraged and have no enthusiasm to try again because they are out of reach. According to Arikunto's assertion, the majority of the teacher-designed questions fall into the medium category, so the composition of the questions in this study is accurate.

Furthermore, related to the percentage proportion of easy category questions, the percentage proportion of medium category questions, and the percentage proportion of difficult category questions, (Suharsimi, 2014) did not state explicitly; he only said that questions that are too easy or too difficult do not mean that they should not be used; this depends on their use.

The same thing is stated in the guidelines of the Ministry of Education and Culture of the Republic of Indonesia in 2019 in the Module for Preparing Questions on Higher Level Thinking Skills (HOTS) PAI and Budi Pekerti, related to the percentage proportion of easy category questions, the percentage proportion of medium category questions, and the percentage proportion of difficult category questions, which is also not stated explicitly.

Furthermore, the research findings show that the analysis of the quality of items according to Bloom's taxonomy revised by Anderson & Krarthwohl in the cognitive domain can be concluded from 35 multiple-choice questions obtained, of which 51.43% (18 questions) are categorized as HOTS questions and 48.57% (17 questions) are categorized as LOTS questions. In particular, the HOTS category questions are all in the C_4 (analysis) criteria, so there is no variation in HOTS-based preparation at C_5 and C_6.

Regarding the percentage of HOTS questions at 51.43% and LOTS questions at 48.57%, basically it is not too far from the theoretical criteria put forward by Suparman (2021: 2) that 40% of lower order thinking skills (LOTS) and 60% of higher order thinking skills (HOTS).

The situation of the teacher's ability to compile HOTS questions is unique because, in this research case, the percentage of questions in the LOTS category and the HOTS category is not much different in percentage. This is a separate concern in the future to further develop the ability of teachers to develop HOTS category questions, even though teachers have received training related to the preparation of HOTS questions.

PAI questions developed by teachers on C1, C2 and C3 are LOTS category questions while C_4, C_5 and C_6 are HOTS category questions as explained in the Ministry of Education and Culture of the Republic of Indonesia in 2019 in the Module for Formulating High Level Thinking Skills (HOTS) PAI and Budi Pekerti which explains that C_1 (remembering), C_2 (understanding), and C_3 (applying) are included in the Low Order Thinking Skill (LOTS) category, while C_4 (analyzing), C_5 (evaluating) and C_6 (creating) are included in the High Order Thinking Skill (HOTS) category.

The same thing is confirmed by Purnomo (2019: 41) and Suparman (2021: 31) that Bloom's taxonomy, namely the C_1 (remembering), C_2 (understanding), and C_3 (applying) levels are included in the LOTS category, while C_4 (analyzing), C_5 (evaluating) and C_6 (creating/making) are HOTS categories.

In general, the ability of PAI teachers at SMA Plus Al Azhar Medan in preparing HOTS questions has been said to be good from the material aspect because 91.42% (32 questions) are suitable for use, while in the construction aspect it can be said to be good because 88.57% (31 questions) are suitable for use, while the ability of teachers in designing questions in the language aspect still needs to be improved because only 42.85% (20 questions) are suitable for use in language.

The questions made by PAI teachers at SMA Plus Al Azhar Medan are in the form of multiple choices. In this case, multiple-choice questions are easier to check and the work can be left to others (Nursalam, 2012), but working on multiple-choice questions allows test takers to only guess the answers and does not represent the level of knowledge of the test takers. In addition, multiple-choice questions are full of cheating by test takers by sharing answers with each other with certain codes, for example, using hands. Analysis of the quality of the questions in the difficulty level aspect shows that 32 questions are (91.43%) in the medium category, 3 questions are (8.57%) in the easy category, and 0 questions are (0%) in the difficult category. Analysis of the quality of the questions in the aspect of different powers of questions shows that 31 questions are (88.57%) in the good category, 3 questions (8.57%) are in the very good category, and 1 question is (2.86%) in the adequate category. (3) analysis

of the quality of the questions according to Bloom's taxonomy as revised by Anderson and Krarthwohl in the cognitive domain. It can be concluded from the 35 multiple-choice questions that 18 questions (51.43%) are in the HOTS question category and 17 questions (48.57%) are in the LOTS question category.

The recommendation that can be proposed is that the preparation of questions must be based on the question indicators obtained from the competency achievement indicators and basic competencies that have been determined. The questions that are prepared must also have answers and meet the criteria for Content Validation and Question Item Validation.

4. CONCLUSION

Conclusions that can be given in connection with the findings of this research are: (1) The content validity of the questions in the material feasibility aspect shows that 32 questions amounting to (91.43%) are suitable for use, and several other questions that do not meet the material eligibility criteria, 3 questions amounting to (8.57%). The analysis of the quality of items on the construction feasibility aspect of 31 questions amounted to 88.57%, which were suitable for use, and several other questions did not meet the construction feasibility criteria; 4 questions amounted to 11.43%. The analysis of items on the language suitability aspect of 20 questions was (57.14%) suitable for use, and several other questions that did not meet the language suitability criteria were 15 questions (42.86%). (2) The validity of the question items shows that 34 questions are (97.14%) in the valid category and 1 question is (2.86%) in the invalid category. Analysis of the quality of the questions in the reliability aspect shows a reliability coefficient of 0.959 in the reliable category.

The recommendation that can be proposed is that the preparation of questions must be based on the question indicators obtained from the competency achievement indicators and basic competencies that have been determined. The questions that are prepared must also have answers and meet the criteria for content validation and question item validation.

REFERENCES

Ahmad, I. F. (2019). Analisis Higher Order Thinking Skills (Hots) Pada Soal Ujian Akhir Siswa Kelas 6 Kmi Dalam Kelompok Mata Pelajaran Dirasah Islamiyahdi Pondok Modern Tazakka Batang. *Jurnal Pendidikan Agama Islam, 16*(2), 137–164.

Anggraini, N. P., Budiyono, & Pratiwi, H. (2019). Analysis of higher order thinking skills students at junior high school in Surakarta. *Journal of Physics: Conference Series*, *1211*(1). https://doi.org/10.1088/1742-6596/1211/1/012077

Asfiyah, S. (2021). Penilaian Berbasis High Order Thinking Skills dalam Meningkatkan Kemampuan Berpikir Kritis dan Kreatif pada Mata Pelajaran PAI dan Budi Pekerti. *Quality*, *9*(1), 103. https://doi.org/10.21043/quality.v9i1.10136

Baharun, H., & Sa'diyah, K. (2018). Penilaian Berbasis Kelas Berorientasi HOTS Berdasarkan Taksonomi Bloom Pada Pembelajaran PAI. *HIKMAH: Jurnal Pendidikan Islam*, *7*(2), 195.

Feriyanto, F., & Putri, R. O. E. (2020). Developing Mathematics Module Based on Literacy and Higher Order Thinking Skills (HOTS) Questions to Train Critical Thinking Ability of High School Students in Mojokerto. *Journal of Physics: Conference Series*, *1594*(1). https://doi.org/10.1088/1742-6596/1594/1/012014

Jarvis, M. A., & Baloyi, O. B. (2020). Scaffolding in reflective journaling: A means to develop higher order thinking skills in undergraduate learners. *International Journal of Africa Nursing Sciences*, *12*(October 2019), 100195. https://doi.org/10.1016/j.ijans.2020.100195

Kwangmuang, P., Jarutkamolpong, S., Sangboonraung, W., & Daungtod, S. (2021). The development of learning innovation to enhance higher order thinking skills for students in Thailand junior high schools. *Heliyon*, *7*(6), e07309. https://doi.org/10.1016/j.heliyon.2021.e07309

Leighton, J. P. (2011). A cognitive model for the assessment of higher order thinking in students.

Assessment of Higher Order Thinking Skills, 151–181.

Maryanto, W. T. (2015). Model Penilaian untuk Pembelajaran Abad 21 (Sebuah Kajian untuk Mempersiapkan SDM Kritis dan Kreatif). *Prosiding Pendidikan Teknik Boga Busana*, *10*(1).

Pradipta, A. W., & Kurniawan, R. (2023). Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Matakuliah Penelitian Pendidikan. *Jurnal Pendidikan*, *11*(02), 234–241.

Qasrawi, R., & Beniabdelrahman, A. (2020). The Higher And Lower-Order Thinking Skills (HOTS and LOTS) In Unlock English Textbooks (1st And 2nd Editions) Based On Bloom'S Taxonomy: An Analysis Study. *International Online Journal of Education and Teaching (IOJET*, *7*(3), 744–758. https://iojet.org/index.php/IOJET/article/view/866

Robinson, D. R., & Schraw, G. (2011). *Assessment of higher order thinking skills*. Information Age Pub.

Sagala, P. N., & Andriani, A. (2019). Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom's Taxonomy. *Journal of Physics: Conference Series*, *1188*(1). https://doi.org/10.1088/1742-6596/1188/1/012025

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer Games and Instruction*, *55*(2), 503–524.

Suharsimi, A. (2014). *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara.