

## Assessment Based on Higher Order Thinking Skill on SKI Learning

Yelfi Dewi S<sup>1</sup>, Moh In'ami<sup>2</sup>, Alina Audia<sup>3</sup>, Masrukhin<sup>4</sup>

<sup>1</sup> UIN Bukittinggi, Indonesia; yelfi.dewi@iainbukittinggi.ac.id

<sup>2</sup> IAIN Kudus, Indonesia; mohinami@iainkudus.ac.id

<sup>3</sup> IAIN Kudus, Indonesia; alinaaudia@gmail.com

<sup>4</sup> IAIN Kudus, Indonesia; masrukhin@iainkudus.ac.id

---

### ARTICLE INFO

#### *Keywords:*

Learning Assessment;  
Instrument Assessment;  
SKI

---

#### *Article history:*

Received 2022-05-27

Revised 2022-10-11

Accepted 2023-01-31

---

### ABSTRACT

This research was conducted to know the assessment instrument for SKI subjects at a state Madrasah Aliyah in Jepara; piloted HOTS-based assessment instrument, and the results of the HOTS-based assessment on SKI subjects. This type of field research is quantitative. This research is field research, namely direct observation of the object under study. Collecting data through interviews, observation, and documentation. The approach used in conducting research is empirical. The instruments used in the SKI subject are oral tests, multiple choice, and assignments. Meanwhile, the researchers used multiple choice and essay test instruments to obtain HOTS-based assessment results. Then the researcher tested the feasibility of the questions with expert validity and the validity and reliability of the items. The final total of 4 multiple choice questions and two essays obtained the result. The HOTS-based assessment was applied to SKI subjects even though all KD could not be included. The findings of this study also show the results of the HOTS-based evaluation with 23 students in the high group, 60 students in the medium group, and 13 students in the low group.

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



---

#### **Corresponding Author:**

Moh In'ami

IAIN Kudus, Indonesia; mohinami@iainkudus.ac.id

---

## 1. INTRODUCTION

Islamic Religious Education (PAI) is a process of teaching Islamic religion carried out in a formal process environment that involves at least three aspects, namely educators (teachers), students, and learning resources. (Amiruddin, A., Nurdin, N., & Ali, 2021; Munjiat, 2020). The focus of the study of PAI subjects in madrasahs includes a.) Al-Qur'an Hadith focuses on being able to read and write well, learn its meaning, and then practice it in life. b.) Faith is the heart of religion. Akidah is also related to faith which will encourage someone to do righteous deeds, have good character, and obey the law. c.) Jurisprudence is a system of sharia rules relating to human actions (*mukallaf*). d.) History of Islamic Culture (SKI) is an event that occurred in the past. This SKI learning emphasizes the ability to take *ibrah*

(lesson) from history to address and resolve current problems and future trends. (Keputusan Menteri Agama (KMA) Nomor 183 Tentang Kurikulum PAI Dan Bahasa Arab, 2019). Allah said, *"By the 'passage of' time! Surely humanity is in 'grave' loss, except those who have faith, do good, and urge each other to the truth, and urge each other to perseverance,"* (Q.S. Al-'Ashr: 1-3); *"O believers! Be mindful of Allah and let every soul look to what 'deeds' it has sent forth for tomorrow.1 And fear Allah, 'for' certainly Allah is All-Aware of what you do."*(Q.S Al-Hasyr: 18). (RI, 1989)

In carrying out Islamic cultural history learning in madrasahs, (Sabki, A. I. A., & Hardaker, 2013) teachers who have high quality and competence are needed to produce intelligent and highly competitive outputs. Competence is defined as a set of knowledge, skills, and behaviors that must be mastered by teachers or lecturers in carrying out professional duties. (Undang-Undang Nomor 14 Tahun 2005 Tentang Guru Dan Dosen, 2005). Competent teachers will be responsible for their students by realizing the implementation of curriculum development, guiding students to learn, personal, character, and physical development of students, analyzing learning difficulties, assessing the learning progress of students, and compiling and implementing assessment procedures. (Emiliasari, 2019; Powell, B. J., Waltz, T. J., Chinman, M. J., Damschroder, L. J., Smith, J. L., Matthieu, M. M., ... & Kirchner, 2015; Wu, J., Chang, I. S., Bina, O., Lam, K. C., & Xu, 2011)

Assessment is carried out to obtain quantity information (through tests by carrying out measurements) and quality data (questions, discussions, observations, and feedback) which are then used to consider and make decisions about grades. The assessment paradigm is identical to high-stake-testing practices accompanied by scoring, grading, and/or comparison of student performance with external standards as a form of accountability. In Indonesia, the policy in providing assessments is focused on the results of the national exam, mid-semester, daily tests, madrasa exams, and practical exams. (Putri Anggoro Kasih, 2016)

Existing policies in Indonesia regarding assessment arise various questions that cause the emergence of irrelevant news between the testing culture in the assessment practices of the learning process, including 1) separating assessment in the learning process, (Collier, 2010; Harks, B., Klieme, E., Hartig, J., & Leiss, 2014; Linn, M. C., Gerard, L., Matuk, C., & McElhaney, 2016; Taras, 2010) 2) encouraging learning and memorization and superficiality, 3) goals the main assessment is seen as a competition, 4) comparing students with each other rather than personal improvement, 5) not paying attention to learning difficulties that students may experience, 6) being a driver of excessive anxiety, (Taylor, 2018) 7) affecting low self-esteem and self-beliefs as learners, and 8) encourage students to devalue the learning process and the assessment itself. (Purnomo, 2016). Looking at some of the statements above, the assessment is not only seen as accountability in the learning process that has been carried out, with the assessment, is expected to be able to provide feedback for students.

Moreover, with the 2013 curriculum, (Febriyanto, E., & Naufal, 2019; Mulawarman, 2021), the government is upgrading the learning and assessment system to be firmer and more comprehensive. Especially for knowledge assessment or cognitive assessment, because teachers consider cognitive assessment in the 2013 curriculum to be the most complicated and confusing assessment, this happens because in the 2013 curriculum cognitive assessment is based on each subject. In addition, there are also HOTS-based learning and questions or assessments in each subject.

HOTS-based learning is one of the modern learning models that is getting a lot of attention currently. (Pratiwi, N., & Mustadi, 2021). But in practice, there are still many who have misconceptions between HOTS-based learning and HOTS-based assessment. (Widana, 2020). Both are two different things, but both have a close and strong relationship. HOTS-based assessment cannot be carried out properly if the learning process that has been carried out by the teacher does not reflect a learning process that directs students to transfer knowledge, think critically and solve problems faced in factual life. These three abilities can be called higher-order thinking skills (HOTS). In line with what has been described above, HOTS learning aims to encourage students to have critical, creative, innovative thinking skills and be able to solve problems. Currently, creativity and innovation are important assets in the success of the learning process of students. For this reason, teachers must prepare teaching

materials, learning resources, learning media, and test instruments to measure learning outcomes according to the HOTS criteria. (Sumardi, 2020). This must be done so that teachers do not only apply assessments based on their knowledge and understanding. (Nurani, 2019)

In addition, learning is still often compacted so that curriculum targets are met. (Renzulli, J. S., Smith, L. H., & Reis, 1982). While thinking capacity is not increased to an optimal level (Higher Order Thinking Skill). Other finding data show that teachers understand that there is a revision of the 2013 Curriculum including the need to develop HOTS in learning activities including in the development of the assessment instrument, but they have difficulty in formulating the indicators contained in the HOTS into an assessment instrument. (Yayuk, 2019)

According to the teachers themselves, HOTS-based learning and assessment can be applied to SKI subjects although not all KD (basic competencies) can be entered using the HOTS system. It's just that educators must make habituation because these subjects are literacy-based, so there needs to be a habit of reading readings related to learning materials. The results of the assessment itself, depend on the students, how well they understand the material being taught and how well they understand the questions being tested so questions that are considered difficult by educators are sometimes in the medium category. For this reason, the teacher himself must be smart in choosing learning methods and choosing the assessment instrument to be used.

One of the assessments made by the teacher on the knowledge aspect is by conducting a written assessment. (Jolly, B., & Dalton, 2018; Lee, 2017; O'Neill, P., & Adler-Kassner, 2010). The written assessment is done using a written test. Written assessment or written test uses two types of questions, namely objective test and essay one. Objective questions (Doody, O., & Bailey, 2016; Hunt, H., Pollock, A., Campbell, P., Estcourt, L., & Brunton, 2018; Lagan, S., Aquino, P., Emerson, M. R., Fortuna, K., Walker, R., & Torous, 2020) are often referred to as selection items or selected-response items. In this type of question, students provide answers to each item presented by selecting the answer choices that have been provided. Types of objective questions include true-false questions, multiple choice questions, and matchmaking. While the type of description is also called a question of giving answers because students must write down the answer or it is often called a description question. Types of description questions are items that measure students' abilities in organizing, integrating, interpreting the tested learning materials, and expressing their answers in writing. (Gavras, K., Höhne, J. K., Blom, A., & Schoen, 2022; Walker, 2015). Explanation questions are useful for measuring student learning progress, intellectual quality, depth of understanding, and even learning difficulties experienced by students.

Another thing that was also done by researchers in preparing the HOTS questions was by paying attention to the objectivity and practicality of the test. A test is said to be objective if the opinions or considerations of the teacher do not influence the process of determining numbers or the process of scoring. So that a test can be guaranteed objectivity, it can be done by formulating test questions specifically and precisely so that the answers are quite clear, avoiding questions that are difficult to digest so that there is no double interpretation, and compiling a test with short, precise or specific answers while checking the test results using an answer key that has been provided previously, in scoring using a key or guideline in awarding predetermined numbers, in contrast to reliability and validity which can be measured, while objectivity cannot be measured.

In making HOTS-based questions on SKI subjects, you must also pay attention to the steps. Higher Order Thinking Skill itself can be interpreted as, higher order thinking ability (HOTS) is a student's thinking pattern by relying on the ability to analyze, create and evaluate all aspects and problems. Meanwhile, high-level thinking is a thinking skill that combines critical thinking and creative thinking. (Varenina, L., Vecherinina, E., Shchedrina, E., Valiev, I., & Islamov, 2021). Rosnawati explained that higher order thinking skills can occur when someone associates newly received information with information that is already stored in his memory, then connects it and/or rearranges and develops the information so that a goal or a resolution of a difficult situation is achieved break it. (Andrean, 2020). Multiple methods of measurement can be used to assess students' HOTS, including a) choosing

(multiple choice, matching, and items); b) generalizing (short responses, essays); and c) providing justifications. (Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, 2019)

Efforts to increase students' ability to solve problems require skills to find, analyze, create, reflect and argue or what are often called higher order thinking skills. Sharpening these skills can be done through subjective or objective questions that require reasoning to answer which are known as HOTS questions. Higher thinking skills require abilities ranging from remembering, referring, and applying to analyzing, and evaluating. (Awaliyah, 2018)

## 2. METHODS

The type of research used by researchers is field research. (Sugiyono, 2014). Namely, research that studies intensively the background of the current situation, and the interaction of a social, individual, group, institution, and society. Field research is the direct observation of the object under study to obtain relevant and valid data. The important idea of this type of research is that the researcher goes to the field to conduct a direct observation of a certain phenomenon that occurs in a place. (Almalki, 2016; Sugiyono, 2014). The place that the researchers targeted was MAN 1 Jepara.

The type of field research used is qualitative research with a descriptive character. The characteristics of qualitative research are descriptive data. (Kim, H., Sefcik, J. S., & Bradway, 2017). In contrast to the opinion of Denzin and Lincoln which states that a qualitative research approach is a research that uses a natural setting to interpret the phenomena that occur and is carried out by involving various existing methods. (Anggito, A., & Setiawan, 2018)

From some of the opinions above, it can be concluded that qualitative research is collecting data in a natural setting to interpret the phenomena that occur and solving the problems being investigated by describing or describing the state of the subject or object of research (a person, group, educational institution, community, etc.). In the end, the essence of qualitative data analysis activities lies in the description or formulation of research results. This painting and formulation are called descriptive. That is a study that aims to describe a variable, group, or social phenomenon that occurs in an educational institution or the community. In this case, the researcher can get accurate data because the researcher meets or deals directly with the informant. Second, the researcher describes the object being studied systematically by noting all things related to the object under study. (Wilkins, J. J., Savic, R. M., Karlsson, M. O., Langdon, G., McIlleron, H., Pillai, G., ... & Simonsson, 2008)

As for the approach used in conducting this empirical type of research, the researcher uses an educational approach. The educational approach is a process of activities, and actions in the field of education to facilitate the implementation of these educational activities. (Rianie, 2015). In this study, researchers used an educational approach because researchers used educational theories to support the research to be carried out. The theories used are intended to explain the nature of education more fully and specifically. Such as educational theory about assessment, tests, and Higher Order Thinking Skills. (Wasitohadi, 2014).

The subject of the research is the source where the research gets a bright spot on the problem under study, in short, the subject of this research is the madrasah who can provide information related to this research. In this study, the main tool or subject is human, meaning that the researcher uses the ability of the researcher in terms of asking, tracking, observing, understanding, and abstracting as an important tool to support research.

The research location is determined based on research needs that can provide information by the research objectives above. Researchers have also conducted interviews, with several key informants. First, the Head of the madrasah can provide data regarding the presence or absence of training received by educators every time there is a change in the curriculum as well as information about SKI subject teachers; Second, the teacher who is in charge of SKI subjects can provide data in the form of the learning process carried out and the assessment instrument used; Third, the deputy head of the curriculum field provides information about the curriculum used in madrasahs; Fourth, students who provide data in the form of HOTS-based assessment results on SKI subjects.

Data collection techniques used to obtain data are carried out in various settings, various sources, and various ways. (Sugiyono, 2014). The research that has been done by researchers uses data collection techniques in the form of observation, interviews, and documentation.

The interview is a means of gathering information by asking several questions orally to be answered orally as well. (Jamshed, 2014; Qu, S. Q., & Dumay, 2011; Turner III, D. W., & Hagstrom-Schmidt, 2022). The main characteristic of the interview is direct face-to-face contact between the information seeker and the information source. To obtain accurate and objective information, it is necessary to build good relationships with respondents. (Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, 2011). In this case, the researchers conducted research by interview or in-depth interviews to obtain detailed and valid data.

To obtain detailed and valid data, the authors conducted interviews with the head of the Madrasah: *First*, to obtain information about whether or not the training was received by educators every time there was a change in the curriculum, for example, currently a HOTS-based assessment has been established; *Second*, waka curriculum to get information about the curriculum used at MAN 1 Jepara at this time; *Third*, the teacher of Islamic cultural history subjects to obtain information in the form of how the learning process is carried out and to know the assessment instruments used; *Fourth*, students of MAN 1 Jepara to find out students' understanding of the learning delivered by the teacher and the results of the HOTS-based assessment.

After that, the researcher conducted observations, defined as direct observation, and recording of the phenomena that appeared on the object of research. Observations and recordings made of objects during events or called direct observation. (Rolfe, 2020). Observations made by researchers are direct observations where researchers come directly to the field or to the research location to observe and directly record the phenomena and things needed in completing the research final project. This means that researchers become active participants to observe every activity that takes place at MAN 1 Jepara to obtain data on HOTS-based assessment instruments in Islamic Cultural History subjects so that data are obtained in the form of HOTS-based assessment results.

At last, to strengthen the findings in the field, the researchers conducted documentation as physical evidence of a study. Documentation of this data collection is in the form of the researcher's writings, such as autobiographies, pictures, books, personal files, or folders that are included in the data. (Firman, 2018). This study contains documentation in the form of pictures ranging from pictures of geographical location, general condition of MAN 1 Jepara, state of study rooms, facilities and infrastructure, the learning process in class, and pictures of interviews with resource persons as evidence that researchers have conducted a study.

In this study, a plan to test the validity of the data is presented. Researchers only test the credibility of the data by triangulation. Triangulation is an activity to re-examine the data. This re-examination of the data the researcher did before the data was analyzed. Triangulation researchers did increase the degree of confidence and accuracy of the data. Triangulation is done by three strategies, namely source triangulation, method triangulation, and time triangulation.

Through source triangulation, researchers seek other information about the topic being studied from other sources or participants. (Moon, 2019; Triangulation, 2014). Namely by conducting questions and answers with subject teachers, deputy head of curriculum, students, and heads of madrasahs. Because, the more sources, the better the results and the more valid the data obtained. Second, method triangulation is a type of triangulation by combines or uses more than one method in analyzing research data. (Quadir, B., Chen, N. S., & Isaias, 2022; Sullivan, G. M., & Artino Jr, 2013). The method triangulation here does not only use the interview method but uses the observation and documentation method. This means that researchers can explore further information through observation and documentation. In principle, method triangulation requires the use of more than one method for re-examination. Finally, time triangulation means that in conducting this research, it is not carried out one time immediately completed but several times on another day. (Wijaya, 2019)

After obtaining field data later, the researchers conducted an analysis. Qualitative data analysis is systematically compiling the data that has been obtained from the results of interviews, field notes, and documentation by grouping the data and describing it into sub-chapters. Then carry out a synthesis to choose which ones are important to make conclusions so that they are easily understood by themselves and the readers.

### 3. FINDINGS AND DISCUSSION

#### SKI Subjects: Assessment Instruments

The results of the study also revealed that according to the students of MAN 1 Jepara, the instrument used in the SKI subject was telling stories, then explaining the material, then getting used to always remembering what was understood." (*Interview*) After knowing some of the student's answers about the instruments often used by subject teachers, the researcher also asked what kind of assessment instrument the students liked the most. Most of them answered multiple choice questions, written tests, and oral tests or story models. (*Observation*)

HOTS-based assessment is one of the efforts made by educators to familiarize their students with high-level thinking by the recommendations stated in the curriculum set by the government. For now, MAN 1 Jepara uses the 2013 curriculum or the revised 2013 version of the education unit level curriculum. The basis used is KMA No. 184 of 2019 while PAI subjects such as Al-Qur'an Hadith, SKI, Fiqh, and Akidah Akhlak are based on KMA No. 183 of 2019. Every time there is a change in the curriculum, both the government and madrasahs themselves will always make training. Every change in the teacher council will receive training and workshops.

An educator must be ready to adapt to changes. Whether from the curriculum, the learning system as well as the instruments and assessments used by educators. After the researchers conducted a series of interviews and observations, the results showed that all educators at MAN 1 Jepara had to implement a HOTS-based assessment system. Do not escape the SKI subjects. Starting from the assessment and teaching, the HOTS system has been implemented. It's just that there are some obstacles to its implementation where the existing pandemic period is still using a partially offline system, or we are shifting.

Meanwhile, the Deputy Head of Curriculum's opinion is also not much different than HOTS-based assessment must also be accompanied by HOTS-based learning as well. Every teacher council at MAN 1 Jepara must implement HOTS-based learning. But not all basic competencies can use the HOTS system. For services, it is always conveyed to Mr. / Ms. Teachers to make HOTS-based services in all subjects including Islamic Cultural History. Because the teacher has also attended training or training at the MGMP level, all MGMPs are collected and then make into essential materials and scientific learning service techniques and HOTS. Madrasahs use this because the government conducts assessments for teachers, assessments for Madrasah Principals, supervisors, and students. So, there are many competency assessments, namely the AKG, AKK, AKP and AN (National Assessment). (*Interview*)

From the implementation of HOTS learning and assessment at MAN 1 Jepara, there are several difficulties or obstacles experienced by teachers when learning and assessment of Islamic Cultural History are applied using the HOTS system. The first difficulty for the teacher is that if he cannot carry out the offline KBM, the implementation of the HOTS learning format will be difficult for students to also experience difficulties because at home the mindset is different. Using zoom meetings is also different, sometimes they don't understand the students, and the educators are also confused. Because some have a hard long-distance signal, don't have a quota, some really don't understand, and some just wake up. In contrast, if learning is done offline, it will be more effective. Second, the lack of teacher literacy related to learning and the variety of learning methods that will be used so that the HOTS learning service sometimes needs to be reminded at every official meeting. Third, there is no

continuity of HOTS learning from lower levels (SD/MI and MTs/SMP) so children have difficulty understanding the flow of thought patterns for such learning. (*Interview*)

It can be concluded that before the assessment activities are carried out, each educator will prepare a test instrument that will be selected according to the characteristics of the students. Generally, to assess the knowledge of students using multiple choice test instruments and essays. After that, arrange the questions according to the specified instrument.

Based on the expression above, the form of multiple-choice questions consists of a question, or a problem and a list of answers provided. Students are asked to choose one correct answer from the answer choices provided. In principle, multiple choice questions consist of a problem and a list of suggested solutions in the form of direct questions or incomplete statements called stems. The subject matter can be in the form of 1) introductory statements and 2) incomplete statements. In addition to the subject matter, in multiple-choice questions, there is a list of suggested solutions called options. Options can be words, numbers, symbols, or certain phrases.

Multiple-choice questions are suitable to be used to determine the dimensions of knowledge which include the categories of factual knowledge, knowledge, conceptual and procedural knowledge. Based on the cognitive process in the revision of Bloom's Taxonomy, it is often assumed that the multiple-choice questions only measure the category of remembering or low-level thinking skills. Multiple-choice questions are more flexible. Multiple choice items can be used to measure the ability of students in the category of understanding and applying or applying.

While the description questions are items that measure students' ability to compose answers in the form of sentences from teaching materials and the opinions of each student and then write them down in the narrative form. Problem descriptions are useful for measuring student learning progress, intellectual quality, depth of understanding, and even learning difficulties experienced by students. The description questions are divided into two, namely free description questions and short or limited description questions. Basically, description questions can be used to measure various student learning outcomes. In the rules of writing descriptive questions that need to be considered are: *First*, formulate the sentences of the questions in detail and clearly so that students do not misinterpret and make multiple interpretations. *Second*, on each item of the description, the division of the answer points, the length of the description (answer) requirements, and the time to do it are written. *Third*, avoid giving a choice of questions that students must answer. (Endrayanto, H. Y. S., & Harumurti, 2014)

When viewed from the theory related to assessment, it is included in the category of formative assessment, namely an assessment that aims to measure student learning development from time to time. In addition, the instruments used by SKI subject teachers are included in the assessment category on the aspect of knowledge that uses subjective questions. Because in the assessment the teacher uses description questions because they are considered able to prove how well the students understand the material that I have conveyed, and the children also must write down the answers according to their respective opinions. (Hariyanto, 2017)

Based on some of the data obtained, it can be concluded that MAN 1 Jepara has implemented HOTS-based learning and assessment but not all basic competencies can use the HOTS system. Every teacher is obliged and must carry out HOTS-based learning for all subjects including Islamic Cultural History. while the instruments that are often used in SKI subjects are multiple choice questions, written tests, and oral tests or story models.

### **Tested Assessment Results**

From the observations that have been made by the researcher, the researcher wants to prove whether the data that has been obtained is truly valid or not, the researcher conducts an experiment by making HOTS-based test questions on SKI subjects and then a trial will be carried out. The first step is to determine the instrument of the test questions to be made. In this case, the researcher used

multiple choice and essay test as instruments. Then, arrange HOTS questions according to the steps for making HOTS questions.

HOTS questions are one of the types of questions that fall into the difficult category. The solution requires deep thinking and high-level thinking. The HOTS questions that will be tested and used in the research are questions made by researchers by paying attention to the procedures for making them based on the provisions of the Ministry of Education and Culture. First, the researcher will analyze KD first according to the material in this even semester. Where in determining KD researchers are prohibited from only fixating on KKO (operational verb) because KKO in the HOTS assessment is only used as an initial guide. (Suhendro, S., Sugandi, D., & Ruhimat, 2021; Widana, 2020). For the second semester, the material for SKI class X is about the development of Islam during the Khulafaur Rasyidin period. Meanwhile, researchers only used two basic competencies that could be used in the HOTS system. This step was taken after the researchers analyzed which KD, they thought were appropriate and included in the cognitive realm. The KD used are, 3.8 Analyzing the selection process of Khulafaur Rashidin and KD 3.9 Analyzing the substance and strategy of Khulafaur Rashidin's da'wah. (Keputusan Menteri Agama (KMA) Nomor 183 Tentang Kurikulum PAI Dan Bahasa Arab, 2019)

In this case, the researcher chose two experts who were considered worthy and competent in the field according to the title of the research above. The first expert, the researcher, chose the SKI subject teacher at the research site. The reason is that he is experienced in his field and has received training several times and knows how to make and form HOTS questions in the subject of Islamic Cultural History.

The second expert chosen by the researcher was one of the lecturers of the Tarbiyah Faculty of IAIN Kudus in the PAI study program with one of the subjects taught was the deepening of Islamic Cultural History material at MTs-MA. From this background, the researcher can provide suggestions and constructive criticism for the formation of HOTS questions in the SKI subjects that the researchers made. Based on the data description table, the validity aspect of the test instrument assessment. The data obtained from the expert validity of the HOTS-based assessment test instrument on SKI subjects has several weaknesses, namely, the questions developed of course need to be improved again following the construct rules of multiple choice and essay questions even though one of them is the results show that the questions are in accordance with the material and indicators. Some of the rules that still seem inaccurate: a) sentences (questions) that must be readjusted; b) options or answer choices.

The following researchers describe in detail the results of expert validity in each aspect of the assessment. In terms of material, the questions made by the researcher are in accordance with the objectives and indicators, the material being asked is also in accordance with the competence being measured slightly different from the results obtained from the lecturer that it is in the sufficient category. The questions are also in accordance with the grid that the researcher made and are included in the HOTS category and there is only one answer key available for each question with homogeneous and logical answer choices. However, the teacher's results are inversely proportional to the fact that there are still questions that present more than one answer key, namely item number 2.

Meanwhile, in terms of construction, the clarity of the instructions for working on the questions is very good with the subject matter being formulated briefly, clearly, and firmly. Slightly different according to the lecturer that there are some questions that are not formulated in a concise, clear, and firm manner. The subject matter does not give clues to the answer key, the distractors are also good, although there are one or two questions whose distractors can be guessed and the probability that the questions can be solved by students is quite good.

In terms of language, the question has used good and correct Indonesian, used communicative language, and did not use taboo language. In general, the comments given by the SKI subject teacher at MAN 1 Jepara were very positive, namely that there was compatibility between the question material and indicators.

While the general comments and suggestions that the author got from the Tarbiyah Faculty lecturers were, there were several sentences or the arrangement of the questions that had to be readjusted so that the questions were easy to understand, and the questions were clear. From these results, the steps that need to be taken are to fix things that are not as recommended by experts. So that the questions that the researchers compiled were declared eligible for testing.

Based on the data obtained by researchers through data collection through interviews, observations, and documentation at MAN 1 Jepara, the researchers can analyze that the purpose of expert validity on two validators is to get criticism and suggestions on the HOTS-based test instrument that has been made. In the form of a multiple-choice test with 5 questions and an essay test with 2 questions.

To prove the feasibility of the questions, the researcher developed a test instrument in which there were the same steps as the steps to prove the validity of the content, namely in the form of: providing a grid and minimum instrument items. Then, the items of the instrument were given to the two experts as mentioned above to be validated by providing input regarding the items of the instrument as well as filling in the conformity of the items with the indicators listed. (Retnawati, 2016)

After the expert validity results have been obtained later, the researcher revises according to the expert's direction so that it is feasible for field trials. Meanwhile, judging from the related theory, what the researcher has done is in accordance with the theory described earlier about the characteristics of a good test, namely the validity of the test. The type of validity carried out by the researcher is content validity. Content validity is the validity that is determined and decided by several experts on a certain matter. (Fitzpatrick, 1983; Sireci, 1998). Expert in question is a person who has expertise in his field, of course with a field that is in accordance with the instrument for research. (Retnawati, 2016)

Another thing that was also done by researchers in preparing the HOTS questions was by paying attention to the objectivity and practicality of the test. A test is said to be objective if the opinions or considerations of the teacher do not influence the process of determining numbers or the process of scoring. So that a test can be guaranteed objectivity, it can be done by formulating test questions specifically and precisely so that the answers are quite clear, avoiding questions that are difficult to digest so that there is no double interpretation, compiling a test with short, precise or specific answers while checking the test results using an answer key that has been provided previously, in scoring using a key or guideline in awarding predetermined numbers, in contrast to reliability and validity which can be measured, while objectivity cannot be measured.

Furthermore, researchers pay attention to practicality in compiling test questions because tests that have high practicability mean that the test is easy and not complicated in administration. On the other hand, a test that is complex and difficult is said to be a test with low practicability. A good test must be practical, with the following characteristics: a. There are clear instructions so that every student being tested can understand what he or she must do or do on the test; b. Easy to implement; c. Students are free to do the easy questions first; d. The test questions are equipped with answer sheets, answer keys, scoring guidelines and scoring keys to facilitate scoring. (Hariyanto, 2017)

After conducting expert validity, the researcher will conduct a small-scale field trial to prove that the questions that the researchers have made are valid and reliable to use. The data from the validity and reliability test items were obtained after the researchers conducted field trials at MAN 1 Jepara to be precise in class X MIPA 3. This field trial activity was only carried out on a small scale, namely on 35 students because 2 students did not leave due to illness so only on a small scale. 33 students. This trial was conducted on Wednesday, January 26, 2022, at the 3<sup>rd</sup> hour. After the field trial, information was obtained that students needed approximately 20 minutes to work on the HOTS questions that the researchers compiled.

From some indications of a good and practical test above, the researcher has included instructions for working on the questions, in its implementation it also does not require complicated preparation, it is enough to bring writing utensils, giving freedom to work on any number of

questions from 5 multiple choice questions and 2 available essay questions. and it is easy to check the scores obtained by students because the researcher has provided an answer key.

The researcher also observed how the students responded in working on the questions that the researchers shared. Some students revealed that the questions that the researcher gave were in the category of difficult questions. Whether from the questions that are difficult to digest and understand or the answers that deceive and require high thinking to be able to solve them optimally. After the researchers finished testing the questions, then the researchers carried out the validity of multiple choice and essay questions on questions that had been answered by the research object.

The results of the validity of the multiple-choice questions above from the 5 questions tested there is one question in number 3 of the multiple choice test instrument which is invalid. While the results of the validity of the essay questions are both valid and can be continued for reliability testing. Next, perform reliability to prove the consistency and reliability of the questions with the results on multiple choice questions being moderate and the reliability results on essay questions being high and fulfilling the requirements for further field trials. That is, a test will be conducted in three classes, namely class X MIPA 1, X MIPA 2, and X MIPA 4.

After the researchers saw the response from the object and the test questions that had been done, not all students knew about the HOTS questions and they answered that the questions that the researchers made were categories of questions that were difficult and misleading. (*Interview*). Looking at some of the results of the interviews that the researchers have conducted above, the researchers got additional information that HOTS-based learning and assessment has been implemented at MAN 1 Jepara in all subjects including SKI. However, not all basic competencies can be applied by the HOTS system. Meanwhile, some students do not know what HOTS is, they only think that the questions are in a difficult category to digest and solve and the answers provided are very misleading.

Based on the research and calculations carried out by the researchers, the results of the validity of the multiple-choice items are item 1 of (0.795), item 2 (0.541), item 3 (0.026), and are declared invalid for item number 3 because  $r$  count is smaller than  $r$  table. Where the calculated  $r$  is 0.026 and the table  $r$  is 0.355. Item 4 (0.579) and item 5 (0.489). While the results of the validity of the essay questions, the results of the two items were declared valid. With the results of  $r$  count (0.934) in item 1 and (0.385) in the last item. The two  $r$  counts are greater than the results of the  $r$  tables so that both are declared valid.

The formula for finding reliability that the researcher uses is by using  $r_{11}$ . The reliability of the multiple-choice test instrument above is stated to be reliable because based on the analysis of the reliability test instrument obtained on the multiple-choice questions is 0.477 with a moderate interpretation. This was obtained after adjusting for the reliability provisions where 0.40  $r_{11}$  0.60 was declared moderate. The reliability of the essay test instrument was declared reliable because it was based on the analysis of the reliability test instrument obtained on the essay question with a result of 0.73 with a high interpretation. This was obtained after adjusting for the reliability provisions where 0.60  $r_{11}$  0.80 was declared high.

The data from the test item validity were carried out by the researcher after conducting the expert validity test. Namely by distributing questions that have been compiled to the object of research on a small scale. This is done because the researcher feels that just doing expert validity is not enough to prove whether the questions compiled are in accordance with the procedures and are feasible to be tested with larger-scale objects or not yet feasible. After getting the results of the questions being tested, the next step is to test the validity and reliability of the items with the results above. Based on the data obtained by researchers through data collection through interviews, observations, and documentation at MAN 1 Jepara, the researchers can analyze that the analysis used to obtain information about the quality of the questions uses validity analysis by using the Pearson Bivariate Correlation method, then performs a reliability test using the product. Pearson's moment to analyze each item, both multiple choice, and essay.

The results obtained after testing the validity and reliability are the remaining questions that are feasible to be tested, totaling 4 multiple choice questions and 2 essay questions. All analytical data obtained by researchers are based on empirical data from field trials. The quality of the HOTS assessment instrument that the researcher made can be seen based on the characteristics of each item which refers to the level of difficulty, discriminating power and effectiveness of distractors. Meanwhile, seen from the theory with formulas that can be used in calculating validity using product moment correlation (Ida, 2021; Sarkadi, 2017):

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

In contrast to validity, reliability can only be calculated by statistics. Then the instrument/test that has high validity does not necessarily have high reliability. Their relationship is independent and independent. (Sarkadi, 2017). The formula that can be used in determining the reliability of test questions is:

$$r_{11} = \left[ \frac{k}{k-1} \right] \left[ 1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right]$$

From the two formulas above, it can be concluded that what is meant by valid means suitable or in accordance with needs. A valid test means that it measures what it is supposed to measure. The test can really give you an idea of what you want to measure. While a test can be said to be reliable if it can be trusted. A test can be trusted if the results achieved by the test are constant or constant. A test becomes unreliable for several reasons, including the situation at the time the test was carried out and the state of the test itself. (Hariyanto, 2017)

What has been explained above has shown that the questions made by the researcher fall into the theory of a good test feature category. Where is evidenced by the validity and reliability tests conducted by researchers. Do not forget to test the validity of this researcher using a significance level of 5% because the research being carried out is educational research. Calculating r tables by means of r tables for 33 students using the formula (number of students - 2 = 33-2 = 31). Then see table r with a significance level of 5%, the result is 0.355. (Sarkadi, 2017)

### HOTS-Based Assessment Results on SKI Subjects

To obtain the results of the HOTS-based assessment, the researcher has tested the feasibility of the questions first, namely in class X MIPA 3, then the researchers conducted validity and reliability. For the validity itself, the researcher conducted content validity to two experts, one of the lecturers of the Tarbiyah Faculty, especially the PAI study program with the SKI subject teacher at MAN 1 Jepara.

In addition, researchers have also tested the validity and reliability of the items. To obtain data on the results of the HOTS assessment, it is necessary to carry out a series of test work on a larger scale. The researcher did the activity a few days after getting the results. The work on this question is a form of further testing that the researchers conducted in class X MIPA 1, X MIPA 2 and X MIPA 4 with the total number of objects reaching 108 students. Since there were some students who were sick and permission to provide guidance to take part in the competition, the final number of research objects became 96 students. This final field trial data was used by researchers to obtain HOTS-based assessment results on SKI subjects at MAN 1 Jepara.

From the results of the final field trial, the researcher will process the results of student answers so that a score is obtained. Starting from the score and some of the information that researchers got from some of these sources, they can answer some of the problem formulations mentioned above. And achieve the research objectives that the author did. The scoring technique that the researcher uses is as follows. The first step is to add up the multiple choice and essay scores, then the writer looks for the average which is symbolized by the letter (M) and looks for the standard deviation (SD). After getting the results later, the researchers put them into the formula. Since the researcher uses grouping into three (High, Medium, and Low) then the formula used is  $X < M - 1SD$  (Low),  $M -$

$1SD \leq X < M + 1SD$  (Medium) and  $M + 1SD \leq X$  (High). (Saifuddin, 2012). The final result of the formula is grouped to obtain data on the percentage of HOTS-based assessment results in SKI subjects.

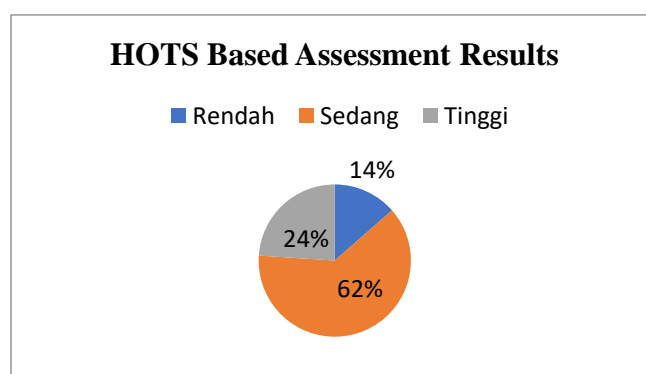
From the results that researchers have done, data is obtained regarding the results of HOTS-based assessments on the subject of Islamic Cultural History at MAN 1 Jepara. The results of the HOTS assessment can be categorized into three levels. The following is the data score based on the HOTS-based assessment. Based on the table below, it can be seen that the results of working on the HOTS questions on the subject of Islamic Cultural History are different.

**Table 1 HOTS Ability Interval of MAN 1 Jepara students**

No.	Standardization Interval	Interval Class	Amount	Group
1	M + 1SD to up	$X \geq 9$	23	High
2	M – 1SD until M + 1SD	$5 \leq X < 9$	60	Medium
3	M – 1SD to down	$X < 5$	13	Low
Amount			N = 96	

Based on the data obtained by the researcher after collecting data by interviewing, observing, and documenting at MAN 1 Jepara, the researcher can analyze that after obtaining scores on the HOTS-based assessment on the SKI subjects the results are varied. Some are low, some are high but most of the results obtained are moderate.

It can be concluded that the average HOTS-based assessment results in SKI subjects obtained the highest percentage results in the medium category, in second place there is a high category, while for the low group the percentage is less. The results of student scores can be seen in the appendix. The following is a picture of the results of the HOTS-based assessment on SKI subjects at MAN 1 Jepara.



**Figure 1 HOTS-Based Assessment Results at MAN 1 Jepara**

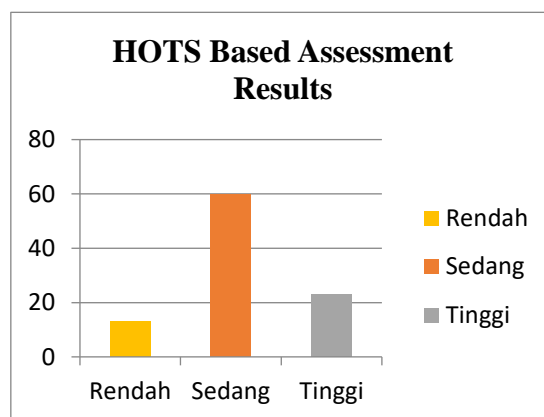
Based on the picture above, the ability to work on HOTS questions on the subject of Islamic Cultural History is different. From the highest group, the percentage (24%), the medium group (62%), and the low group (14%).

The overall results of the HOTS-based assessment on SKI subjects can be seen in the bar chart below.

From the picture, the results of the HOTS-based assessment on SKI subjects at MAN 1 Jepara with the number of students who scored in the high group were 23 students, in the medium group there were 60 students and in the low group there were 13 students. Based on the two pictures above, it can be seen the result of the HOTS ability level of the students of MAN 1 Jepara.

After conducting a search on research that is relevant to the research to be carried out by the researcher, the authors found similarities with the research conducted, including both discussing higher order thinking skills (HOTS) which are specialized in one particular subject, namely

accounting, not SKI and The main topic of the author's discussion is on the instruments and results of the HOTS-based assessment while existing research is on the process and application of HOTS-based learning. (Nursari, E. V., Setiawati, I., & Lismaya, 2021)



**Figure. HOTS Based Rating Bar Chart**

After reviewing several studies that are relevant to the researcher's research entitled "HOTS-Based Assessment of Islamic Cultural History Subjects at MAN 1 Jepara", this study has a position as a complement to previous research.

#### 4. CONCLUSION

Everything about the assessment instruments in MAN 1 Jepara is clearly regulated in Permendikbud No. 23 of 2016. It regulates the form of assessment, assessment instruments, assessment mechanisms and procedures. The assessment instruments used in Islamic Cultural History are oral test instruments, random questions and answers, assignments, and discussions. The implementation of the assessment for all educators at MAN 1 Jepara must implement a HOTS-based assessment system. For the HOTS-based assessment instrument, which was tested using multiple choice and essay test instruments with manufacturing steps in accordance with the provisions of the Ministry of Education and Culture. Before conducting field trials, expert validity tests and item validity tests were carried out. The quality of the items used validity analysis by using the Pearson Bivariate Correlation method, after that, they conducted a reliability test using Pearson's product moment to analyze each item, both multiple choice and essay.

Based on the results of the calculations that have been carried out, the results of the validity of the multiple-choice items are item 1 (0.795), item 2 (0.541), item 3 (0.026) and are declared invalid for item number 3 because  $r$  count is smaller than  $r$  table. Item 4 (0.579) and item 5 (0.489). While the results of the validity of the essay questions are (0.934) in item 1 and (0.385) in the last item. The reliability of the multiple-choice test instrument was declared reliable because based on the analysis of the test instrument the reliability obtained on the multiple-choice questions was 0.477 with a moderate interpretation. While the essay questions with a result of 0.73 with a high interpretation.

After conducting trials in class X MIPA 1, X MIPA 2 and X MIPA 4 with the number of objects reaching 96 students. So, the HOTS-based assessment on SKI subjects at MAN 1 Jepara got a score in the high group there were 23 students, in the medium group there were 60 students and in the low group there were 13 students.

#### REFERENCES

- Almalki, S. (2016). Integrating Quantitative and Qualitative Data in Mixed Methods Research-- Challenges and Benefits. *Journal of Education and Learning*, 5(3), 288–296.
- Amiruddin, A., Nurdin, N., & Ali, M. (2021). Islamic Education Teacher Communication Strategy in

- Increasing Students' Learning Interest. *International Journal of Contemporary Islamic Education*, 3(1), 41–61.
- Andreas, S. (2020). Developing Strategies And Evaluation Of Hots-Based Learning On Thematic Learning In Elementary School. *Sunan Kalijaga International Journal On Islamic Educational Research*, 4(2), 64-76.
- Anggito, A., & Setiawan, J. (2018). *Metodologi Penelitian Kualitatif*. CV Jejak (Jejak Publisher).
- Awaliyah, S. (2018). "Penyusunan Soal HOTS Bagi Guru PPKn Dan IPS Sekolah Menengah Pertama." *Jurnal Praksis Dan Dedikasi Sosial*, 1(1).
- Collier, C. (2010). *Seven Steps to Separating Difference from Disability*. Corwin Press.
- Doody, O., & Bailey, M. E. (2016). Setting A Research Question, Aim and Objective. *Nurse Researcher*, 23(4).
- Emiliasari, R. N. (2019). Lesson Planning in EFL Classroom: A Case Study in Lesson Plan Preparation and Implementation. *Wiralodra English Journal*, 3(2), 367–375.
- Endrayanto, H. Y. S., & Harumurti, Y. W. (2014). *Penilaian Belajar Siswa di Sekolah*. PT. Kanisius.
- Febriyanto, E., & Naufal, R. S. (2019). Attitude Competency Assessment in The 2013 Curriculum Based On Elementary School Prototyping Methods. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, 1(1), 87-96.
- Firman. (2018). *Terampil Menulis Karya Ilmiah*. Aksara Timur.
- Fitzpatrick, A. R. (1983). The Meaning of Content Validity. *Applied Psychological Measurement*, 7(1), 3-13.
- Gavras, K., Höhne, J. K., Blom, A., & Schoen, H. (2022). *Innovating The Collection of Open-Ended Answers: The Linguistic and Content Characteristics of Written and Oral Answers to Political Attitude Questions*. BORA.
- Hariyanto, I. B. dan. (2017). *Asesmen Pembelajaran*. PT Remaja Rosdakarya.
- Harks, B., Klieme, E., Hartig, J., & Leiss, D. (2014). Separating Cognitive and Content Domains in Mathematical Competence. *Educational Assessment*, 19(4), 243–266.
- Hunt, H., Pollock, A., Campbell, P., Estcourt, L., & Brunton, G. (2018). An Introduction to Overviews of Reviews: Planning A Relevant Research Question and Objective for An Overview. *Systematic Reviews*, 7(1), 1–9.
- Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher Order Thinking Skills from Elementary to Master Students in Environmental Learning. *European Journal of Educational Research*, 8(4), 935-942.
- Ida, F. F. (2021). Validitas Dan Reliabilitas Dalam Analisis Butir Soal. *Al-Mu'arrif: Journal Of Arabic Education*, 1(1), 34–44.
- Jamshed, S. (2014). Qualitative Research Method-Interviewing and Observation. *Journal of Basic and Clinical Pharmacy*, 5(4).
- Jolly, B., & Dalton, M. J. (2018). Written Assessment. *Understanding Medical Education: Evidence, Theory, and Practice*, 291-317.
- Keputusan Menteri Agama (KMA) Nomor 183 tentang Kurikulum PAI dan Bahasa Arab, (2019).
- Kim, H., Sefcik, J. S., & Bradway, C. (2017). Characteristics of Qualitative Descriptive Studies: A Systematic Review. *Research in Nursing & Health*, 40(1), 23–42.
- Lagan, S., Aquino, P., Emerson, M. R., Fortuna, K., Walker, R., & Torous, J. (2020). Actionable Health App Evaluation: Translating Expert Frameworks into Objective Metrics. *NPJ Digital Medicine*, 3(1), 1–8.
- Lee, I. (2017). *Classroom Writing Assessment and Feedback in L2 School Contexts*. Springer Singapore.
- Linn, M. C., Gerard, L., Matuk, C., & McElhaney, K. W. (2016). Science Education: From Separation to Integration. *Review of Research in Education*, 40(1), 529–587.
- Moon, M. D. (2019). Triangulation: A Method to Increase Validity, Reliability, and Legitimation in Clinical Research. *Journal of Emergency Nursing*, 45(1), 103-105.
- Mulawarman, W. G. (2021). Overcoming Obstacles in Implementing 2013 Curriculum Policy. *Cypriot*

- Journal of Educational Sciences*, 16(3), 967-980.
- Munjiat, S. M. (2020). Implementation of Islamic Religious Education Learning in Higher Education on The Pandemic Period. *Nazhruna: Jurnal Pendidikan Islam*, 3(2), 285–295.
- Nurani, H. (2019). “Analisis Pelaksanaan Penilaian Kognitif Berbasis Kurikulum 2013 Sabiul Ulum Mayonglor Kabupaten Jepara.” *Indonesian Journal Of Education Research and Review*, 2(2).
- Nursari, E. V., Setiawati, I., & Lismaya, L. (2021). Analisis Perangkat Pembelajaran Berbasis Higher Order Thinking Skill (HOTS) di Masa Pandemi Covid-19. *ALVEOLI: Jurnal Pendidikan Biologi*, 2(2), 78–97.
- O’Neill, P., & Adler-Kassner, L. (2010). *Reframing Writing Assessment to Improve Teaching and Learning*. Utah State University Press.
- Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content Validity—Establishing and Reporting The Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: part 2—Assessing Respondent Understanding. *Value in Health*, 14(8), 978–988.
- Powell, B. J., Waltz, T. J., Chinman, M. J., Damschroder, L. J., Smith, J. L., Matthieu, M. M., ... & Kirchner, J. E. (2015). A Refined Compilation of Implementation Strategies: Results from The Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science*, 10(1), 1–14.
- Pratiwi, N., & Mustadi, A. (2021). Hots-Based Learning in 2013 Curriculum: Is it Suitable?. *JPI (Jurnal Pendidikan Indonesia)*, 10(1), 128-135.
- Purnomo, Y. W. (2016). “Assesment-Based Learning: Sebuah Tinjauan untuk Meningkatkan Motivasi Belajar dan Pemahaman Matematis.” *Sigma*, 6(1).
- Putri Anggoro Kasih, Y. W. P. (2016). “Peningkatan Hasil Belajar Matematika Siswa Sekolah Dasar Melalui Pembelajaran Berbasis Penilaian.” *Journal Of Research and Advance in Mathematics Education*, 1(1), 69–78.
- Qu, S. Q., & Dumay, J. (2011). *The Qualitative Research Interview*. Qualitative Research in Accounting & Management. Sage Publications.
- Quadir, B., Chen, N. S., & Isaias, P. (2022). Analyzing The Educational Goals, Problems and Techniques Used in Educational Big Data Research from 2010 to 2018. *Interactive Learning Environments*, 30(8), 1539–1555.
- Renzulli, J. S., Smith, L. H., & Reis, S. M. (1982). Curriculum Compacting: An Essential Strategy for Working with Gifted Students. *The Elementary School Journal*, 82(3), 185-194.
- Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, dan Psikometrian)*. Parama Publishing.
- RI, D. A. (1989). *Al-Qur'an dan Terjemahnya*. Toha Putra.
- Rianie, N. (2015). Pendekatan Dan Metode Pendidikan Islam (Sebuah Paradigma dalam Konsep Teori Pendidikan Islam dan Barat). *Jurnal: Management of Education*, 1(2).
- Rolfe, S. A. (2020). Direct Observation. In *Doing Early Childhood Research*. Routledge.
- Sabki, A. I. A., & Hardaker, G. (2013). The Madrasah Concept of Islamic Pedagogy. *Educational Review*, 65(3), 342-356.
- Saifuddin, A. (2012). *Penyusunan Skala Psikologi, Edisi 2*. Pustaka Pelajar.
- Sarkadi, K. dan. (2017). *Evaluasi Pembelajaran*. Rizqita Publishing & Printing.
- Sireci, S. G. (1998). The Construct of Content Validity. *Social Indicators Research*, 45(1), 83-117.
- Sugiyono. (2014). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Suhendro, S., Sugandi, D., & Ruhimat, M. (2021). Analysis on Factors Influencing Geography Teachers’ Ability in Constructing High-Order Thinking Skills (HOTS) Assessment Instrument. *Geosfera Indonesia*, 6(2), 205-221.
- Sullivan, G. M., & Artino Jr, A. R. (2013). Analyzing and Interpreting Data from Likert-Type Scales. *Journal of Graduate Medical Education*, 5(4), 541–542.

- Sumardi. (2020). *Teknik Pengukuran dan Penilaian Hasil Belajar*. Deepublish.
- Taras, M. (2010). Student Self-Assessment: Processes and Consequences. *Teaching in Higher Education*, 15(2), 199–209.
- Taylor, J. E. (2018). The Extent and Characteristics of Driving Anxiety. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58(1), 70-79.
- Triangulation, D. S. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum*, 41(5).
- Turner III, D. W., & Hagstrom-Schmidt, N. (2022). *Qualitative Interview Design. Howdy or Hello? Technical and Professional Communication*. OER Pressbooks.
- Undang-Undang Nomor 14 Tahun 2005 Tentang Guru dan Dosen, (2005).
- Varenina, L., Vecherina, E., Shchedrina, E., Valiev, I., & Islamov, A. (2021). Developing Critical Thinking Skills in A Digital Educational Environment. *Thinking Skills and Creativity*, 41, 100906.
- Walker, M. (2015). The Quality of Written Peer Feedback on Undergraduates' Draft Answers to An Assignment, and The Use Made of The Feedback. *Assessment & Evaluation in Higher Education*, 40(2), 232–247.
- Wasitohadi, W. (2014). Hakekat Pendidikan Dalam Perspektif John Dewey Tinjauan Teoritis. *Satya Widya*, 30(1), 49–61.
- Widana, I. W. (2020). The Effect of Digital Literacy on The Ability of Teachers to Develop HOTS-Based Assessment. *Journal of Physics: Conference Series*, 1503(1).
- Wijaya, H. (2019). *Analisis Data Kualitatif: Sebuah Tinjauan Teori & Praktik*. Sekolah Tinggi Theologia Jaffray.
- Wilkins, J. J., Savic, R. M., Karlsson, M. O., Langdon, G., McIleron, H., Pillai, G., ... & Simonsson, U. S. (2008). Population Pharmacokinetics of Rifampin in Pulmonary Tuberculosis Patients, Including A Semimechanistic Model to Describe Variable Absorption. *Antimicrobial Agents and Chemotherapy*, 52(6), 2138–2148.
- Wu, J., Chang, I. S., Bina, O., Lam, K. C., & Xu, H. (2011). Strategic Environmental Assessment Implementation in China—Five-Year Review and Prospects. *Environmental Impact Assessment Review*, 31(1), 77–84.
- Yayuk, E. (2019). “Kemampuan Guru Dlam Implementasi Pembelajaran Dan Penilaian HOTS Pada Siswa Kelas 4 Sekolah Indonesia Bangkok Thailand”. *JINoP (Jurnal Inovasi Pembelajaran)*, 5(2).