Inteligensi: Jurnal Ilmu Pendidikan Vol. 8, No. 1, 2025. Hal. 186 – 197
Tersedia online di https://jurnal.unitri.ac.id/index.php/inteligensi
ISSN 2656-601X*(online)*
ISSN 2656-8675 (cetak)

ASSESSMENT STRATEGY BASED ON VALIDITY AND RELIABILITY IN DETERMINING THE MINIMUM COMPLETENESS CRITERIA FOR ISLAMIC RELIGIOUS EDUCATION

M. Hafidz NurAzizi[1*], Akhmad Robayt Alfauzi[2], Abdul Bashith[3]
[1-3] Magister Pendidikan Agama Islam, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia
Email: *hafidazizi98@gmail.com*[*]

ABSTRACT

This study explores assessment strategies based on validity and reliability in determining the Minimum Competency Criteria (KKM) in Islamic religious education. Validity and reliability are two fundamental concepts educators must understand to ensure that the assessment instruments used are dependable and accurate in measuring student competencies. Validity is defined as the degree to which an instrument accurately measures the intended object, while reliability pertains to the consistency of assessment results. The document also emphasizes the importance of conducting trial tests on assessment instruments before widespread implementation to identify and address existing shortcomings. Additionally, this research highlights the use of the Delphi technique as a systematic method to collect and evaluate expert opinions in determining KKM. Through in-depth analysis, the study employs a library research method, reviewing books, journals, and various other literature sources. The findings conclude that the effective application of the Delphi technique, combined with a focus on validity and reliability, can enhance the quality of assessments in Islamic religious education.

Keywords*: Validity; reliability; Minimum Competency Criteria, Islamic religious education; Delphi technique.*

ABSTRAK

Penelitian ini mengeksplorasi strategi penilaian berdasarkan validitas dan reliabilitas dalam menentukan Kriteria Ketuntasan Minimal (KKM) dalam pendidikan agama Islam. Validitas dan reliabilitas adalah dua konsep dasar yang harus dipahami oleh pendidik untuk memastikan bahwa instrumen penilaian yang digunakan dapat diandalkan dan akurat dalam mengukur kompetensi siswa. Validitas didefinisikan sebagai sejauh mana instrumen secara akurat mengukur objek yang dimaksud, sedangkan keandalan berkaitan dengan konsistensi hasil penilaian. Dokumen ini juga menekankan pentingnya melakukan uji coba pada instrumen penilaian sebelum implementasi luas untuk mengidentifikasi dan mengatasi kekurangan yang ada. Selain itu, penelitian ini menyoroti penggunaan teknik Delphi sebagai metode sistematis untuk mengumpulkan dan mengevaluasi pendapat ahli dalam menentukan KKM. Melalui analisis mendalam, penelitian ini menggunakan metode penelitian perpustakaan, meninjau buku, jurnal, dan berbagai sumber literatur lainnya. Temuan menyimpulkan bahwa penerapan teknik Delphi yang efektif, dikombinasikan dengan fokus pada validitas dan keandalan, dapat meningkatkan kualitas penilaian dalam pendidikan agama Islam.

Kata kunci: Validitas, Reliabilitas, Kriteria Ketetuan Minimal, Pendidikan Agama Islam, Teknik Delphi.

Cara mengutip: NurAzizi, MH., Alfauzi, AR., & Bashith, A. (2025). *Assessment Strategy Based on Validity and Reliability in Determining the Minimum Completeness Criteria for Islamic Religious Education*. *Inteligensi: Jurnal Ilmu Pendidikan, 8(1), 186 – 197* http://doi.org/10.33366/ilg.v8i1.7248

Retrieved from https://jurnal.unitri.ac.id/index.php/inteligensi/article/view/7248

Inteligensi: Jurnal Ilmu Pendidikan Vol. 8, No. 1, 2025. Hal 186 – 197

INTRODUCTION

Islamic Religious Education (IRE) plays a pivotal role in shaping the character and morals of the younger generation. It is not merely an academic subject but a comprehensive approach to instilling values, ethics, and religious principles that guide students in their daily lives. In this context, assessment becomes a critical component in determining the effectiveness of Islamic education. A robust assessment system should not only evaluate students' academic understanding but also measure their ability to apply religious teachings in real-life situations. However, the current assessment practices in IRE often face significant challenges, particularly in terms of validity and reliability, which directly impact the determination of the Minimum Completeness Criteria (MCC) or Kriteria Ketuntasan Minimal (KKM).

The importance of validity and reliability in educational assessments cannot be overstated. Validity refers to the degree to which an assessment accurately measures what it is intended to measure, ensuring that the data collected truly reflects the students' abilities and understanding (Sugiyono, 2018). On the other hand, reliability pertains to the consistency of the assessment results over time, ensuring that the outcomes are stable and reproducible under similar conditions. These two aspects are fundamental in ensuring that the assessment process is fair, accurate, and accountable. Without them, the determination of MCC risks being arbitrary, leading to potential injustices and a lack of motivation among students.

Data from the Ministry of Education and Culture in 2021 highlights the urgency of addressing these issues. A significant decline in the final exam results of students at the madrasah level was reported, raising concerns about the effectiveness of the current assessment methods (Makarim, 2021). This decline underscores the need for a comprehensive evaluation and improvement of the assessment strategies employed in IRE. The determination of MCC, which serves as a benchmark for student achievement, must be based on a thorough analysis of assessment validity and reliability to ensure that it reflects students' true capabilities and learning outcomes.

This study aims to explore strategies for improving the quality of assessment in Islamic Religious Education by focusing on validity, reliability, and the determination of appropriate MCC. Specifically, the research seeks to (1) analyze the current challenges in the assessment practices of IRE, particularly concerning validity and reliability; (2) develop a framework for implementing assessment strategies that are both valid and reliable; and (3) provide practical recommendations for determining MCC that are fair, transparent, and aligned with the educational objectives of IRE.

By addressing these objectives, the study contributes to the ongoing efforts to enhance the quality of Islamic education and ensure that it fulfills its role in shaping morally and ethically sound individuals. The novelty of this research lies in its integrated approach to addressing the issues of validity and reliability in the assessment of Islamic Religious Education. While previous studies have explored various

aspects of assessment in education, few have specifically focused on the unique challenges faced in IRE, particularly in the context of determining MCC. This study bridges the gap by providing a comprehensive analysis of how validity and reliability can be operationalized in the assessment process to ensure that MCC is determined in a manner that is both equitable and reflective of students' true abilities.

Moreover, this research emphasizes the dual role of IRE assessments in evaluating both academic performance and the application of religious values in daily life. This dual focus is critical in ensuring that the assessment process aligns with the holistic objectives of Islamic education. By integrating theoretical insights with practical recommendations, this study offers a valuable contribution to the field of educational assessment and provides a roadmap for improving the quality of IRE assessments.

The findings of this study have significant implications for policymakers, educators, and researchers in the field of Islamic education. For policymakers, the study provides evidence-based recommendations for developing assessment policies that prioritize validity and reliability. For educators, it offers practical strategies for implementing assessments that are fair, transparent, and aligned with the objectives of IRE. Finally, for researchers, the study opens new avenues for exploring the intersection of assessment practices and educational outcomes in Islamic education.

In conclusion, this research addresses a critical gap in the field of Islamic Religious Education by focusing on the validity and reliability of assessments and their role in determining MCC. By doing so, it contributes to the broader efforts to enhance the quality of Islamic education and ensure that it fulfills its mission of shaping morally and ethically sound individuals.

RESEARCH METHOD

This study employs a literature review method to analyze the concepts of validity and reliability in educational assessment, particularly in the context of determining the Minimum Completeness Criteria (MCC) for Islamic Religious Education (IRE). The literature review method is appropriate for this research as it allows for an in-depth exploration of existing theories, frameworks, and empirical studies related to the assessment instruments used in IRE. By synthesizing and critically analyzing relevant literature, this study aims to provide a comprehensive understanding of how validity and reliability can be applied to improve the quality of assessment and MCC determination.

The primary focus of this research is to examine the theoretical and practical aspects of validity and reliability in educational assessment. Validity, as defined by Widodo et al. (2023), refers to the extent to which an instrument accurately measures what it is intended to measure. This study explores three types of validity: content validity, construct validity, and criterion validity. Content validity ensures that the material being assessed aligns with the subject's objectives and context, such as the use of authentic sources like the Qur'an and Hadith in IRE. Construct validity evaluates whether the

Inteligensi: Jurnal Ilmu Pendidikan Vol. 8, No. 1, 2025. Hal 186 – 197

instrument effectively measures conceptual variables, such as attitudes and motivation, while criterion validity assesses the instrument's accuracy based on internal and external benchmarks.

Reliability, on the other hand, pertains to the consistency of assessment results over time. This study examines two key aspects of reliability: response reliability and the consistency reliability of combined items. Response reliability evaluates whether repeated measurements yield consistent results, while consistency reliability assesses the uniformity of results across different items within the same instrument. These concepts are crucial in ensuring that the assessment instruments used in IRE are both accurate and dependable.

The data for this study is collected through a systematic review of academic journals, books, government regulations, and other credible sources. The inclusion criteria for the literature are as follows: (1) studies that focus on validity and reliability in educational assessment, (2) research related to MCC determination in the Indonesian education system, and (3) publications that provide empirical evidence or theoretical insights into the assessment practices in IRE. The selected literature is analyzed using a thematic approach, which involves identifying, organizing, and interpreting key themes related to validity, reliability, and MCC.

The analysis process begins with data reduction, where the collected literature is sorted and categorized based on its relevance to the research objectives. Key elements, such as definitions, methodologies, and findings, are extracted and organized into thematic categories. For example, studies discussing content validity are grouped together to provide a comprehensive understanding of how this concept is applied in IRE assessments. Similarly, literature on reliability is analyzed to identify common challenges and best practices in ensuring consistent assessment results.

The next step is data presentation, where the findings from the literature are summarized and presented in a structured format. This includes the use of narratives, tables, and graphs to illustrate the relationships between validity, reliability, and MCC determination. For instance, a table may be used to compare different methods of testing validity, such as expert judgment for content validity and statistical analysis for construct validity.

Finally, the study concludes with drawing conclusions and verification, where the synthesized findings are interpreted to address the research objectives. This involves evaluating the implications of validity and reliability for MCC determination and providing recommendations for improving assessment practices in IRE. For example, the study may suggest the use of the Delphi technique, as described by Nasrullah (2013), to involve expert panels in the development and validation of assessment instruments.

The literature review method is particularly suitable for this research because it allows for a comprehensive analysis of existing knowledge while identifying gaps and opportunities for future research. By integrating theoretical insights with practical recommendations,

this study contributes to the ongoing efforts to enhance the quality of assessment in IRE and ensure that MCC determination is both fair and effective. Additionally, the findings of this study are expected to provide valuable guidance for educators, policymakers, and researchers in developing assessment instruments that are valid, reliable, and aligned with the objectives of Islamic education.

RESULT AND DISCUSSION

1. Validity and Reliability

Reliability and validity are two important concepts for educational assessment that need to be well understood by educators, especially in the context of Islamic religious education. Validity refers to a measure that indicates the extent to which an instrument can be considered valid or valid (Widodo et al. 2023). In 1937, Garret defined validity as the degree of accuracy of a test in measuring a particular object. Furthermore, in 1971, Cronbach stated that validity must be based on the various available evidence in order for the test to be considered more accurate. Although the definition of validity is still a subject of debate, in 1987, Mehres proposed that the most appropriate definition of validity is a precise measure of a measurement in the context of making tests (Anshari et al. 2024). Validity instruments can be tested in several ways, including content validity, construct validity, and criterion validity.

a. Content Validity

The validity of the content means that the material or content being tested must be in accordance with the subject's abilities, knowledge, lessons, and context (Widodo et al. 2023). For example, in the intructional of Islamic religious education, such as learning materials that are in accordance with authentic and recognized sources, such as the Qur'an, hadith and books that have been recognized by scholars.

b. Construct Validity

Construct validity, also known as *construct validity* , is a validity that requires a measuring tool that has several indicators to measure constructs (Ida and Musyarofah 2021). Construct validity is usually applied to tools used to evaluate conceptual variables related to performance, such as attitudes, interests, self-concept, loci of control, leadership style, motivation for achievement, and so on. And it can also be maximum *performance* such as aptitude tests, intellectual and emotional intelligence.

c. Validity of Criteria

Criterion validity, also known as empirical validity, refers to the determination of validity based on internal and external criteria. To ensure its validity, the results of the trial were carried out on participants who were comparable to the subjects investigated or assessed. The internal criteria are related to the instrument itself which is used as a reference, while the external criteria are related to the results of other instruments or tests that are not related to the instrument used as a reference (Ramadhan, Siroj, and Afgani 2024). However, in certain conditions, external criteria can also consist of other measures that are considered standard or reliable.

Inteligensi: Jurnal Ilmu Pendidikan Vol. 8, No. 1, 2025. Hal 186 – 197

Reliability comes from the word *reliability*, which refers to how accurate the results of an examination can be considered credible if they show consistent results in various tests on the same group of subjects, as long as the elements measured do not change (Ida and Musyarofah 2021). So reliability can be interpreted as the level of consistency of individual deviation scores, where the z-score shows relatively stable consistency when administrative repetition is carried out using the same test or equivalent tests. In the context of learning Islamic religious education, reliability can be measured through the analysis of the reliability of responses and the reliability of the consistency of the combined items.

a. Response Reliability

Response reliability refers to the extent to which respondents' responses to a particular test or instrument are reliable and consistent. In this case, if an instrument is used to measure the size of an object and then a remeasurement is performed on the same object, then it is important to know if the results of the second measurement remain consistent with the first. If the results of the second measurement show inconsistencies, then it is clear that the test results of the measurement do not accurately show the object observed (Ramadhan et al. 2024).

b. Consistency Reliability of Combined Grains

The reliability of the combined items is closely related to the consistency between the items in a test. This can be explained by asking, is the measurement result of one item no different from the measurement result of another item on the same part of the object being measured? If the measurement results on the same part of the object show inconsistencies between different items, then the measurements made by the test cannot be considered a reliable unit (Ramadhan et al. 2024).

From the explanation of validity and reliability that has been explained, the two concepts refer to the suitability between the assessment instrument and the goals to be achieved. These two concepts are interrelated and are very important in ensuring that the assessments carried out in Islamic religious education are trustworthy and provide an accurate picture of the student's competence. Without good validity and reliability, the assessment results will not provide useful information for further educational development.

2. Urgency of Validity and Reliability

Validity in the assessment has a very crucial role to ensure that the assessment really measures the expected competencies. In the context of Islamic religious education, validity is becoming increasingly important because the material taught is often related to moral and spiritual values. For example, if an assessment instrument only measures the memorization of verses of the Qur'an without considering the understanding of its meaning and implementation in daily life, then the assessment can be considered invalid.

Reliability also has a significant impact on the quality of assessment results. Stable and trustworthy results will give confidence to students, parents, and schools that the assessments carried out are fair and objective. In the context of Islamic religious education, reliability is very important to

ensure that the assessments carried out are not affected by bias or other external factors.

Therefore, educators need to integrate assessment into the entire learning system. This is due to the fact that assessment is an integral element of the learning process itself (Affandi et al. 2024). In addition, it is also important to test assessment instruments before they are widely implemented. By conducting this trial, educators can identify and correct deficiencies in the assessment instrument, so that the results obtained are more accurate and reliable. In the context of Islamic religious education, high reliability will help in assessing the understanding and application of religious teachings more objectively.

3. Minimum Completeness Criteria (KKM)

One of the important components in the assessment of education in Indonesia, including Indonesian Religious Education (PAI), is the Minimum Completeness Criteria (KKM). According to Government Regulation of the Republic of Indonesia Number 57 of 2021 concerning National Standards, KKM is the minimum standard for the unity of attitudes, skills, and knowledge that shows the achievement of students' abilities at the end of the education level (RI 2021). This shows that KKM functions as an indicator of students' success in understanding the material being taught. In other words, KKM is not just a number, but is the result of an in-depth analysis of students' abilities and curriculum guidance. The determination of the right KKM will help teachers in evaluating and improving the learning process.

The function of KKM in Islamic religious education is very crucial, especially in the context of learning assessment and evaluation. First, KKM functions as a measuring tool that determines the extent to which students have achieved the expected competencies. With the KKM, teachers can clearly identify students who need additional or remedial assistance. This is in line with the principle of student-oriented education, where each individual has a different learning speed and style. Second, the KKM also plays a role in improving the quality of learning. By setting realistic and challenging KKM, teachers can create more effective learning strategies. Third, the KKM functions as a motivator for students. When students know about KKM, they tend to be more motivated to learn and strive to achieve these scores. That way, KKM does not only function as a limit on grades, but also serves as a driver to improve the quality of education. The determination of the right and relevant KKM in Islamic religious education will have a positive effect on student learning outcomes and the quality of education as a whole.

With many curricula and a variety of teaching methods, effective data collection techniques are needed to get an accurate picture of student development. The following are data collection techniques to assess student achievement:

a. Observation

Matthews and Ross define observation as the process of gathering information through the use of human senses. According to this description, human

Inteligensi: Jurnal Ilmu Pendidikan Vol. 8, No. 1, 2025. Hal 186 – 197

senses such as sight, hearing, smell, taste and so on are the main instruments for observation (Sidiq and Choiri 2019). This technique is carried out with the aim of directly knowing the student learning process (Nuralan 2019).

b. Interview

In this process, questions are needed to students or teachers to find out what the learning indicators are used and what are the constraints found in the learning process (Nuralan 2019).

c. Documetry

Documentation is one of the methods of data collection that does not involve participants directly. This technique is carried out to collect data on evidence of the learning process as a completeness of data on the implementation of the learning assessment implementation process (Nuralan 2019).

d. Questionnaire

Questionnaire is a data collection method in which participants are given a series of written questions to be answered with the aim of meeting the user's information needs (Purnomo and Palupi n.d.).

There are data analysis methods to determine completeness and improvement, namely:

1. Data reduction

Data reduction means sorting, selecting key elements, and focusing on the most significant elements. The reduced data will provide a clearer view and make it easier for researchers to collect additional data. Electronic devices such as computers can help in this process by assigning codes to specific elements (Sidiq and Choiri 2019).

2. Data presentation

After the data is successfully reduced, the next step is to present the data. To determine the completeness of learning, data can be presented in the form of narratives, graphs, or tables to make it easier to understand. The presentation of this data helps researchers understand the situation that occurs, so that they can take the next step based on the understanding that has been obtained (Sidiq and Choiri 2019).

3. Withdrawal Conclusion

After presenting the data, it is necessary to draw conclusions or it can also be called verification (Sidiq and Choiri, 2019). In this process, core data is taken from several previously collected data that has been analyzed to assess whether the learning objectives are achieved using the Minimum Completeness Criteria (KKM) (Sarastiwy 2018).

4. Development of Remidial Programs for Students Who Have Not Reached KKM

Steps in the Development of a remidial program for students who have not reached the Minimum Completeness Criteria:

1. Tes diagnosis

A diagnostic test is a test given to students before a remidial exam is conducted, this test is done to see the student's learning difficulties on fractional material. In the process of the diagnosis test, the researcher carried out several stages such as checking the attendance of students and then conditioning the readiness of students to learn after which perception activities were carried out. The test results that students have taken are then

Inteligensi: Jurnal Ilmu Pendidikan Vol. 8, No. 1, 2025. Hal 186 – 197

corrected by the researcher (Sarastiwy 2018).

2. Remidial Learning

This remidial learning is given to students who have not reached the Minamal Completeness Criteria (KKM) in the diagnosis test. After delivering material about fractions, the researcher held a question and answer session for students who did not understand the material. Furthermore, students who still do not understand the material are given a remedial test (Sarastiwy 2018).

3. Application of assessment evaluation techniques

Basically, evaluation is not only limited to assessing academic achievement, but also includes the potential of each student and identifying which ones need to be improved. In this article, it will be explained about the assessment instrument carried out by Muhammad Nasrullah using the Delphi assessment instrument.

In determining the Minimum Completeness Criteria (KKM), a new technique is needed. Among the methods that can be used is the delphi technique. According to Wiliam Dunn (2008) in the book Public Policy Analysis: *an Introduction that "delphi technique is an intuitive forecasting procedure for obtaining, exchanging, and developing opinion about future events"* Delphi technique is a method used to predict future events by asking questions, searching, and collecting, as well as developing opinions from experts individually (Adib 2019). In another sense, the Delphi technique can be described as a method that systematically seeks out, collects, evaluates, and analyzes the opinions of experts. In this technique, it is hoped that it can obtain KKM scores that are in line with the teacher's expectations and not from the results of grade engineering (Nasirullah 2013).

Prediction verification in the Delphi technique involves the participation of experts, and predictions about future phenomena are based on empirical data. An agreement was reached after this verification. In this process, the development of the instrument aims to gain support from experts in the field related to the instrument. Using the Delphi technique, you can do the following: conduct a needs analysis to find problems; determine the priority, type, components, and process of making the instrument; set the objectives of the development of the instrument; and determine problem-solving methods, especially for the development of research instruments. (Adib 2019).

The first step in determining the KKM score using the Delphi technique is to determine the criteria for the required teachers. First, determine the subjects whose KKM scores will be determined. Second, make sure that the teacher's educational background is in accordance with the subjects he teaches. Third, pay attention to the number of teachers who teach the subject. If there is only one teacher who teaches the subject, then it is necessary to cooperate with other educational units or schools that have almost the same categories and conditions. After determining the teacher criteria, the next step is to select teachers who are in accordance with the criteria that have been set. In determining the KKM score, teachers can make a questionnaire using multiple-choice questions or descriptions

Inteligensi: Jurnal Ilmu Pendidikan Vol. 8, No. 1, 2025. Hal 186 – 197

which are then distributed to objects which are then analyzed by a panel of experts (Nasirullah 2013)

4. Challenges and Solutions of Delphi Engineering in determining KKM

Although the delphi technique offers many benefits, there are some challenges that may be encountered in its application. There are several challenges in the application of the delphi technique, including:

1. Engagement and commitment from a panel of experts

In some cases, experts may not have enough time to actively participate in this process, which can result in inadequate responses or even abandonment of the questionnaires disseminated. To address these challenges, it is important for organizers to select a panel of experts who not only have relevant knowledge, but are also willing to commit to the process.

2. Potential bias in expert panel selection

If the expert panel does not reflect the required diversity of perspectives, then the results obtained may not be representative. Therefore, it is important to ensure that the expert panel is made up of individuals with diverse backgrounds, including educational experience, and areas of expertise which will help produce a more comprehensive view and reduce the risk of bias.

3. Relatively long time

The Delphi process can take quite a long time, especially if there are many expert panels involved. To overcome this, organizers can design questionnaires that are more concise and focused, so that the data collection process can be carried out more efficiently. This can use technologies such as online surveys that can accelerate data collection and analysis (Fink-Hafner et al. 2019).

Finally, it is important to ensure that the results of the Delphi technique are communicated clearly and transparently to all stakeholders. By proactively addressing these challenges, the delphi technique can be a very effective tool for determining relevant and beneficial KKM for the educational process.

CONCLUSION

This study shows that validity and reliability are two fundamental concepts that must be understood by educators to ensure that the assessment instruments used are reliable and accurate in measuring student competence. Validity is defined as a measure of the accuracy of the instrument in measuring the intended object, while reliability is related to the consistency of assessment results. One of the methods proposed in this study is the Delphi technique, which is a systematic approach to collecting and evaluating the opinions of experts. While the Delphi technique offers many benefits, the study also identifies some of the challenges that may be encountered in its application.

These challenges include the involvement and commitment of the expert panel, the potential for bias in the selection of the expert panel, and the relatively long time required for the process. To address these challenges, it is important for organizers to select a panel of experts who have relevant knowledge and are willing to commit, as well as ensure diversity of perspectives within the expert panel.

Overall, the effective application of Delphi techniques, along with an emphasis on validity and reliability, can improve the quality of assessment in Islamic religious education. The results of this study are expected to make a significant contribution to the development of better and more equitable education policies, as well as improve the learning process for students.

REFERENCES

Adib, Helen Sabera. 2019. "Teknik Pengembangan Instrumen Penelitian Ilmiah Di Perguruan Tinggi Keagamaan Islam." *Sains Dan Teknoogi* 139–57.

Affandi, Lalu Hamdian, H. Husniati, N. Nurhasanah, and Khairun Nisa. 2024. "Strategi Guru Sekolah Dasar Dalam Menentukan Soal Penilaian Hasil Belajar Siswa." *Jurnal Ilmiah Pendidikan Citra Bakti* 11(4):1113–24. doi: 10.38048/jipcb.v11i4.4128.

Anshari, Muhammad Isa, Rodiah Nasution, Muhammad Irsyad, Alifia Zuhriatul Alifa, and Indah Aminatus Zuhriyah. 2024. "Analisis Validitas Dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI." *Edukatif : Jurnal Ilmu Pendidikan* 6(1):964–75. doi: 10.31004/edukatif.v6i1.5931.

Fink-Hafner, Danica, Tamara Dagen, May Dousak, Meta Novak, and Mitja Hafner-Fink. 2019. "Delphi Method: Strengths and Weaknesses." *Metodoloski Zvezki* 16(2):1–19. doi: 10.51936/FCFM6982.

Ida, Farida Far, and Anna Musyarofah. 2021. "Validitas Dan Reliabilitas Dalam Analisis Butir Soal." *Al-Mu'Arrib: Journal of Arabic Education* 1(1):34–44. doi: 10.32923/al-muarrib.v1i1.2100.

Makarim, Nadiem Anwar. 2021. *Laporan Kinerja 2021*.

Nasirullah, Mohammad. 2013. "Penetapan Nilai Kriteria Ketuntasan Minimal Dengan Teknik Delphi Di SMA Negeri Kabupaten Pamekasan." *Jurnal Kebijakan Dan Pengembangan Pendidikan* 1:35–41.

Nuralan, Siti. 2019. "Implementasi Penilaian Pembelajaran Kurikulum 2013 Dalam Meningkatkan Partisipasi Aktif Siswa SDN 1 Buga Sitti." *Tolis Ilmiah; Jurnal Penelitian* 1(2):124–29.

Purnomo, Puji, and Maria Sekar Palupi. n.d. "Buku Teknik Penyusunan Instrumen Penelitian."

Ramadhan, Muhammad Fakhri, Rusydi A. Siroj, and Muhammad Win Afgani. 2024. "Validitas and Reliabilitas." *Journal on Education* 6(2):10967–75. doi: 10.31004/joe.v6i2.4885.

RI, BPK. 2021. *Peraturan Pemerintah (PP) Nomor 57 Tahun 2021 Tentang Standar Nasional Pendidikan*.

Sarastiwy, Nadella. 2018. "Pelaksanaan Pembelajaran Remedial Untuk Mengatasi Kesulitan Belajar Pada Materi Pecahan Siswa Kelas V SD." 1–23.

Sidiq, Umar, and Moh. Miftachul Choiri. 2019. *Metode Penelitian Kualitatif Di Bidang Pendidikan*. Vol. 53.

Sugiyono. 2018. *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D*. Bandung: ALFABETA.

Widodo, Slamet, Festy Ladyani, La Ode Asrianto, Rusdi, Khairunnisa, Sri Maria Puji Lestari, Dian Rachma Wijayanti, Ade Devriany, Abas Hidayat, Dalfian, Sri Nurcahyati, Tessa Sjahriani, Armi, Nurul Widya, and Rogayah. 2023. *Metodologi Penelitian*.