



## Research Instrument: Instrument Development and Instrument Calibration

Alip Nuryanto<sup>1</sup>, Naf'an Tatihoran<sup>2</sup>, Rizal Firdaus<sup>3</sup>

<sup>1,2,3</sup> UIN Sultan Maulana Hasanuddin Banten, Indonesia

### ARTICLE INFO

#### Article history:

Received December 27, 2025

Revised Mei 21, 2025

Accepted June 6, 2025

Available online June 29, 2025

#### Keywords:

Research instruments, instrument development, instrument calibration, data reliability, thematic analysis.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Copyright © 2025 by Author.  
Published by Universitas Al-Hikmah Indonesia.

### ABSTRACT

The objective of this study is to explore in depth the stages involved in developing an effective research instrument and the calibration methods required to enhance accuracy and reliability in data collection, particularly in the fields of education and social sciences. The development and calibration of research instruments are critical components in ensuring the validity and reliability of research findings. This study addresses various issues related to the design and refinement of research instruments, employing a qualitative approach that includes interviews, consultations with experts, and literature review. Data analysis was conducted using thematic analysis to identify patterns and key practices in instrument development and calibration. The findings indicate that the effectiveness of a research instrument largely depends on its alignment with research objectives, the context of its application, and the characteristics of the target population. Calibration plays a vital role in ensuring that each item within the instrument measures accurately and consistently across various settings. This study makes a significant contribution by offering practical and reflective guidance for researchers across disciplines to design and calibrate reliable instruments. Thus, it concludes that continuous testing and refinement of instruments are essential to improving the overall quality of research outcomes.

## INTRODUCTION

Research in Islamic education, like scientific research in general, requires valid and reliable instruments to collect accurate data. According to Sugiyono (Sugiyono, 2022), a research instrument is a tool used to measure the phenomena being observed, both in natural and social contexts. The quality of the instruments used in research has a direct impact on the quality of the data collected, which in turn affects the validity of the research conclusions. This is especially important in the context of Islamic education, which involves various cognitive, affective, and psychomotor dimensions in the learning process.

The development of instruments is a crucial stage in research. Arikunto (2019) emphasizes that a good instrument must meet two main requirements: validity and reliability. Validity refers to the accuracy of the instrument in measuring what it is supposed to measure, while reliability is concerned with the consistency of the measurement results. In the context of Islamic education, research instruments must be able to measure various aspects of

learning, including cognitive, affective, and psychomotor aspects, as well as Islamic values, such as understanding of concepts, religious practices, and the formation of character (Muhaimin, 2021). Instruments that are invalid or unreliable will produce inaccurate data and reduce the validity of the research.

The development of an instrument begins with formulating the research objectives and determining the variables to be measured. Each instrument must be aligned with the research goals and the variables to be measured to produce accurate and relevant data. Validity is one of the main criteria in the development of instruments, with various types of validity, such as content, construct, and criterion validity, which need to be considered to ensure the accuracy of the measurements (Arikunto, 2019).

Instrument calibration is an important process that should not be overlooked. Djaali and Muljono (2021) state that calibration involves testing validity, reliability, and analyzing the items of the instrument. Calibration aims to adjust the instrument so that it produces accurate and consistent data. This process typically begins with a trial run of the instrument on a limited sample to detect weaknesses and identify areas for improvement before the instrument is applied to a larger population. This process ensures that the instrument used can function optimally in broader measurements.

Calibration of instruments is especially important in social research and Islamic education because instruments that measure social variables, such as attitudes, perceptions, and beliefs, are often subjective and require special adjustments. This process helps to reduce bias in measurements and increases the reliability of the data obtained. Therefore, proper calibration is essential to ensure that the instrument produces more objective results (Djaali & Muljono, 2021).

With advancements in technology, the process of instrument development and calibration has become more efficient. Computer-based applications can assist researchers in designing, analyzing, and calibrating instruments automatically, which accelerates the process and minimizes human error. This technology is especially helpful in improving the precision of research instruments, particularly in measuring quantitative variables that require high accuracy. Additionally, the use of technology allows for faster and more detailed data analysis (Sugiyono, 2022).

After initial development and calibration, the instrument needs to be field-tested to ensure its effectiveness and reliability in real-world conditions. Field testing gives the researcher the opportunity to observe how the instrument performs in collecting data on a larger scale and provides feedback for revising the instrument. Additionally, expert validation of the instrument is crucial to ensure that the instrument covers all important aspects relevant to the variables being researched. This validation process is important to ensure that the instrument is ready for use on a larger scale (Arikunto, 2019).

In the context of Islamic education, instrument calibration is crucial because the aspects measured are not only cognitive but also affective and psychomotor, involving the measurement of values and religious behavior. Therefore, calibration must be done carefully so that the instruments used can accurately reflect the relevant conditions and developments within Islamic education. This process also ensures that the instruments used can measure the complex learning objectives in Islamic education (Muhaimin, 2021).

Effective instrument development and calibration have a significant impact on the quality of research. Valid and reliable instruments will produce more accurate and trustworthy data, which will strengthen the research findings and enhance their acceptance and application in educational policies or daily practices. On the other hand, instruments that are less valid and reliable may produce biased data, which ultimately undermines the quality of the research and its conclusions (Djaali & Muljono, 2021).

The development of research instruments is an ongoing process. This process involves revising and testing the instrument, as well as adjustments to ensure its accuracy and consistency. Therefore, instrument calibration is not only important to ensure the accuracy of measurements but also to improve the precision of the data obtained, which can ultimately contribute to higher-quality research (Sugiyono, 2022).

The development and calibration of research instruments are essential stages in any research, including in the field of Islamic education. Valid and reliable instruments significantly affect the quality of the data collected and, in turn, the overall quality of the research. Previous studies, such as those by Nasution (2019) and Syarif & Huda (2021), have emphasized the importance of developing psychometrically sound instruments to measure constructs like students' religiosity, teachers' pedagogical competence, and learning outcomes in Islamic education. However, many of these studies have not elaborated on the systematic process of instrument calibration or have used limited indicators without contextual validation. This leaves a research gap in terms of constructing instruments that are both theoretically grounded and empirically validated specifically for diverse Islamic educational settings. Therefore, this article aims to develop and calibrate a valid and reliable research instrument tailored for Islamic education research, addressing the methodological gaps found in earlier studies.

## **METHODS**

This research employs a descriptive-qualitative approach to gain a deeper understanding of the development and calibration of research instruments within Islamic education. The descriptive aspect aims to present a clear picture of existing phenomena, while the qualitative aspect uncovers the meanings and perceptions of participants regarding the instrument development process (Sugiyono, 2022). The study utilizes literature reviews and in-depth interviews as data collection methods. A literature review of articles, books, journals, and online sources provides insights into the theories and practices related to research instruments. As Prasetyo (2020) notes, the literature review serves as an effective means of gathering data from publicly available documents.

In-depth interviews with experts in Islamic education and research instrument development, such as lecturers and practitioners, help provide first-hand perspectives. The data sources consist of primary data from interviews and secondary data from literature reviews (Arikunto, 2019). This study collects qualitative data, which includes theories, practices, and expert opinions on instrument calibration. Thematic analysis is used to categorize and analyze emerging themes, such as validity, reliability, and the integration of technology in instrument development (Sugiyono, 2022; Djaali & Muljono, 2021).

## **RESULT**

## **Definition of Research Instruments**

Research instruments are tools or devices utilized by researchers to systematically collect data essential for addressing research questions. These instruments are pivotal across both qualitative and quantitative research, as the data they gather supports objective measurements of research variables (J. D. Creswell, 2014). They can take various forms, including tests, questionnaires, observation tools, and interview guides. The meticulous development of these instruments is critical to ensure data reliability (Sugiyono, 2022). Qualitative research often employs instruments such as observation or interview guides to focus on specific variables, while quantitative research utilizes structured scales for statistical analysis. Instruments enable systematic and reliable responses to research inquiries (Cohen dkk., 2011).

## **Functions of Instruments in Research**

Research instruments serve essential functions in data collection, providing systematic ways to derive data needed to answer research questions and test hypotheses (J. D. Creswell, 2014). In quantitative research, instruments ensure that the collected data is consistent and well-aligned with the targeted variables, employing clear measurement scales to minimize bias (Fraenkel & Wallen, 2009). They help reduce measurement errors, especially when using validated and calibrated tools (Cohen dkk., 2011). For qualitative research, instruments play a crucial role in obtaining in-depth insights into subjects' perspectives, guiding researchers to explore attitudes and emotions (Azwar, 2015). Ultimately, these instruments facilitate accurate, accountable data that form the basis for conclusions and recommendations (Polit & Beck, 2010).

## **Types of Research Instruments**

Research instruments can be classified based on their data collection methods, with questionnaires being one of the most common tools in quantitative research. They efficiently collect structured data from a large number of respondents, simplifying analysis (J. D. Creswell, 2014). In contrast, interviews are essential in qualitative research, allowing deeper exploration of subjects' opinions, experiences, and perceptions (Sugiyono, 2018). Observations provide contextual insights into behaviors by enabling direct witnessing of interactions (Fraenkel & Wallen, 2009). Experimental instruments are used to measure specific treatment outcomes, helping researchers control variables for accurate effect assessments (Neuman, 2014). Each type of instrument contributes uniquely to data collection, enhancing understanding of diverse research contexts.

## **Specific Measurement Scales and Their Importance**

In quantitative research, measurement scales such as Likert scales objectively capture attitudes and perceptions, facilitating statistical analysis (Cohen dkk., 2011). Additionally, other instruments include biological measurement tools, psychometric tests, and online surveys. Biological tools yield precise physical data, while psychometric assessments evaluate personality or cognitive traits, requiring stringent validation to maintain accuracy and reliability (Polit & Beck, 2010). Online surveys have gained popularity due to

technological advancements that streamline data collection processes and expand the researcher's reach to participants (Thompson & Melancon, 2011). By employing varied instruments, researchers gain flexibility and diversity in approaching complex phenomena, allowing for comprehensive data gathering.

### Significance of Reliable Research Instruments

The reliability of research instruments is fundamental for producing accurate data, as unreliable tools can yield inconsistent or biased results, complicating the analysis (Creswell & Plano Clark, 2017). A robust research instrument undergoes rigorous validation to ensure its effectiveness in measuring variables aligned with the study's objectives (Silverman, 2016). Good instruments enhance the credibility of research findings, making them more likely to be embraced by the academic community and practitioners (Thompson & Melancon, 2011). They also help identify changes in research variables, enabling researchers to make broader generalizations and uncover significant patterns in their findings (Gay dkk., 2012). Ultimately, research instruments are indispensable in advancing scientific knowledge and facilitating accurate interpretations of complex phenomena.

**Table 1.** Types of Research Instruments

No.	Type of Instrument	Explanation	Reference
1	Questionnaire	A questionnaire is an instrument commonly used in quantitative research to efficiently obtain data.	Creswell, 2014
2	Interview	Interviews allow researchers to explore qualitative data from respondents, often used in qualitative research to understand respondents' opinions or perceptions.	Sugiyono, 2018
3	Observation	Observation is an instrument used to directly observe subjects' behavior, enabling researchers to obtain empirical data about respondents.	Fraenkel & Wallen, 2009
4	Experimental Instrument	Experimental instruments are used to measure the outcomes of specific treatments applied to research subjects, applied in experimental research with controlled variables.	Neuman, 2014
5	Measurement Scale	Measurement scales, such as Likert scales, allow researchers to objectively measure attitudes or perceptions through numbers in quantitative research.	Cohen dkk., 2011
6	Case Study Guide	A case study guide is used in qualitative research to explore in-depth information about a specific case, usually utilizing interviews, observations, and documents.	Azwar, 2015
7	Psychometric Test	Psychometric tests are used in psychology to measure specific aspects of personality or cognitive abilities in subjects, requiring rigorous validation.	Polit & Beck, 2010
8	Survey Instrument	A survey instrument is a tool for collecting data from a large sample, generally used to obtain general information about a population in a short time.	Schreiber et al., 2006
9	Biological Measurement Tool	Biological measurement tools, such as blood tests or blood pressure monitors, are used in biomedical research to provide objective and accurate physical data.	Silverman, 2016
10	Test or Examination	Tests or examinations are used in education as instruments to measure student learning outcomes,	Creswell & Plano Clark,

		allowing evaluation of achievements in curricular aspects.	2017
11	Documentation Instrument	Documentation instruments, such as official documents or archives, are used in qualitative research to complement data from interviews or observations.	Anderson & Krathwohl, 2001
12	Standardized Test	Standardized tests, such as TOEFL or IQ tests, are used to measure individuals' abilities or intelligence, allowing comparisons between individuals or groups.	Brown, 2015
13	Online Survey	Online surveys are instruments that use digital platforms to collect data from respondents, more efficient as they do not require physical data collection.	Thompson & Melancon, 2011
14	Audio-Visual Instrument	Audio-visual instruments are used for recording audio or video in research, often utilized in communication or human behavior studies.	Gay dkk., 2012
15	Application or Software	Specialized applications or software are used for automatic data analysis, important in quantitative research such as SPSS or R.	Neuman, 2014

### Steps in Research Instrument Development

The first step in developing a research instrument involves determining the constructs or variables to be studied. Constructs are abstract concepts that must be operationalized into measurable variables, forming the foundational basis of the research process (Creswell, 2014). Following this, researchers formulate specific and measurable indicators that reflect different aspects of the constructs (Sugiyono, 2018). This is critical to ensure that the indicators can be observed and measured accurately, avoiding bias. Subsequently, a blueprint is developed to graphically represent the relationships between constructs, indicators, and the survey items (Anderson & Krathwohl, 2001). The blueprint ensures comprehensive coverage of all aspects of the constructs, facilitating a structured approach to the instrument's design.

### Writing Item Statements and Pilot Testing

After developing the blueprint, researchers write item statements for the instrument, ensuring that these items are clear, unambiguous, and easy for respondents to understand (Cohen dkk., 2011). It is essential to pilot test these items to identify any potential issues with comprehension or relevance before widespread use in research. According to Neuman (2014), pilot testing helps refine the instrument, ensuring that it produces valid and reliable data. Researchers may revise the instrument based on pilot test feedback, which prepares it for use in larger studies, ultimately enhancing its effectiveness for data collection.

### Instrument Validity

Instrument validity is essential for effective research, with different types of validity assessing various aspects of the tool. Content validity evaluates how well the items encompass all aspects of the construct being measured, often through expert evaluations to enhance relevance (Polit & Beck, 2010). Construct validity measures whether the instrument accurately reflects the intended construct without influence from irrelevant variables (Creswell & Plano Clark, 2017). Lastly, criterion validity assesses the relationship between the instrument's results and external criteria, further divided into predictive and concurrent

validity (Schreiber dkk., 2006). Each form of validity plays a crucial role in ensuring the accuracy and relevance of the research findings.

### **Instrument Reliability**

Reliability is a key aspect of research instruments, encompassing internal consistency, stability, and equivalence. Internal consistency refers to how well items within the instrument measure the same construct, often assessed using Cronbach's alpha (Polit & Beck, 2010). Stability evaluates the consistency of measurement results over time, typically using the test-retest method (Creswell, 2014). Equivalence relates to the consistency of results from two different versions of the instrument, which is important for reducing bias (Sugiyono, 2018). High reliability in these dimensions assures researchers of the trustworthiness of their findings, thus facilitating robust conclusions from their studies.

### **Definition and Importance of Instrument Calibration**

Instrument calibration refers to the process of adjusting and validating measurement tools to ensure their accuracy and consistency. In research, calibration is vital for enhancing the validity and reliability of instruments used for data collection (Neuman, 2014). Proper calibration involves assessing the content, construct, and criterion validity of the instruments to understand their strengths and limitations. As emphasized by Creswell (2014), effective calibration enables researchers to capture accurate data relevant to their studies. The calibration process must also consider the context and research objectives, as different environments may require tailored calibration methods, including statistical analyses for item performance evaluation (Sugiyono, 2018). Ultimately, calibration is essential for ensuring the integrity of data collected, particularly in longitudinal studies where consistency over time is critical (Polit & Beck, 2010).

### **Objectives of Instrument Calibration**

The primary goal of instrument calibration is to ensure valid and reliable measurement results. Validity relates to the extent that instruments accurately measure intended constructs, while reliability pertains to the consistency of those measurements over time (Creswell & Plano Clark, 2017). Calibration also involves item analysis, which helps identify effective items and those needing improvement or removal, thereby enhancing the quality of the instrument (Sugiyono, 2018). This process is ongoing, as feedback from calibration allows researchers to refine instruments and improve outcomes. Furthermore, a rigorous calibration process increases the credibility of the instruments, fostering trust in research findings by transparently assuring their reliability (Neuman, 2014).

### **Calibration Methods: CTT and IRT**

Common methods for instrument calibration include Classical Test Theory (CTT) and Item Response Theory (IRT). CTT analyzes reliability and validity based on true scores and measurement errors, providing a foundation for calculating reliability coefficients like Cronbach's alpha (Cohen dkk., 2011). While CTT is straightforward, it does not always account for item difficulty variations, prompting researchers to supplement it with other methods (Neuman, 2014). In contrast, IRT offers a more nuanced approach by examining the

probability of respondents answering items correctly, allowing for individual item evaluation (Polit & Beck, 2010). Although IRT provides more accurate ability estimates with fewer items, it demands advanced statistical knowledge, making it essential for researchers to understand both CTT and IRT to select the most appropriate method for their study (Sugiyono, 2018).

**Table 2.** Calibration Methods

Aspect	Classical Test Theory (CTT)	Item Response Theory (IRT)
<b>Definition</b>	Method focused on analyzing the reliability and validity of instruments based on two components: true scores and measurement error.	Method focused on the relationship between respondents' abilities and the probability of answering items correctly.
<b>Theoretical Basis</b>	Each measurement consists of two components.	Identifies the performance of each item individually.
<b>Advantages</b>	- Ease of calculation and interpretation of results. - Foundation for calculating reliability coefficients (e.g., Cronbach's alpha).	- Provides more accurate estimates of respondents' abilities. - Allows the development of more efficient instruments.
<b>Limitations</b>	- Does not take item characteristics into account. - Assumes that all items have the same level of difficulty.	- Requires a deep understanding of statistical analysis. - More complex than CTT.
<b>Use in Research</b>	Often used to evaluate how well an instrument measures the intended construct.	Used to evaluate instruments in a broader context and identify effective items.

## Instrument Calibration Procedures

The first step in instrument calibration involves data collection using the developed instrument under specified conditions. As stated by Creswell and Plano Clark (2017), effective data collection occurs with a representative sample to facilitate generalization of results to a broader population. Researchers must carefully document the conditions during data collection, accounting for external factors that could influence outcomes. Following data collection, researchers perform item analysis using methods like Classical Test Theory (CTT) or Item Response Theory (IRT) to evaluate each item's effectiveness. This analysis is crucial for identifying which items contribute positively towards measuring the intended construct and which may require revision or removal (Polit & Beck, 2010).

## Item Selection and Final Instrument Assembly

Once item analysis is complete, researchers move on to item selection, aiming to retain items that best measure the constructs of interest. According to Sugiyono (2018), it's important to choose items that are appropriately challenging and can differentiate among respondents' abilities. Validity and reliability considerations are integral to this decision-making process. After finalizing the item selection, researchers assemble the instrument and conduct a thorough review to ensure clarity and organization for respondents (Creswell, 2014). Following these systematic calibration steps, researchers can enhance the instrument's validity and reliability, thereby supporting the generation of quality data and advancing knowledge in their field.

**Table 3.** Instrument Calibration Procedures

Procedure Step	Description	Reference
<b>a. Data Collection</b>	The first procedure in instrument calibration is data collection, conducted using the developed instrument under predetermined conditions. It is important to ensure that the sample used is representative for generalization purposes, and researchers should note the conditions and situations of data collection to consider external factors.	(Creswell & Plano Clark, 2017)
<b>b. Item Analysis</b>	After data collection, the next step is item analysis, where researchers use methods like Classical Test Theory (CTT) or Item Response Theory (IRT) to evaluate the performance of each item in the instrument. This analysis aims to identify items that work well and those that need improvement or removal.	(Polit & Beck, 2010)
<b>c. Item Selection</b>	After item analysis, researchers move on to item selection to choose the most effective items for measuring the intended construct. Items should have appropriate difficulty levels and be able to differentiate between respondents. The validity and reliability of each item must also be considered.	(Sugiyono, 2018)
<b>d. Final Instrument Assembly</b>	After the item selection process is complete, researchers assemble the final instrument. All selected items are organized into a cohesive instrument ready for use. Final reviews are recommended to ensure the instrument is well organized and easily understood by respondents. This step is crucial for ensuring the instrument can provide valid and reliable results.	(Creswell, 2014)
<b>e. Validity and Reliability</b>	Following a systematic calibration procedure ensures that the instrument has high validity and reliability, which is essential for generating quality data that supports knowledge development in the studied field.	

## DISCUSSION

### Application of Instrument Development in Islamic Education Research

In Islamic education research, instrument development is crucial for assessing students' comprehension of fundamental concepts. For example, researchers can create a questionnaire that evaluates understanding of the Qur'an and Sunnah, encompassing topics like the pillars of faith and ethics (Creswell, 2014). This development process helps obtain valuable data that highlights students' understanding levels and identifies areas needing enhancement in their learning experiences.

### Impact and Continuous Assessment

The effectiveness of the developed instrument is evaluated through result analysis, including assessments of validity and reliability using methods such as Cronbach's alpha (Neuman, 2014). Insights gained from this analysis enable educators to tailor their teaching strategies, possibly incorporating experiential methods to enhance engagement and comprehension (Sugiyono, 2018). Moreover, continuous assessments of these instruments, informed by feedback from both students and educators, are essential for refining their relevance and effectiveness over time (Azwar, 2015). This iterative process not only aligns

with curriculum standards but also contributes to the evolution of Islamic education practices (Creswell & Plano Clark, 2017).

### **Implementation of Instrument Calibration in Islamic Education**

Instrument calibration is essential in Islamic education research to ensure that assessments, such as reading skills in the Qur'an, accurately measure student abilities. Employing methodologies like Item Response Theory (IRT) allows researchers to analyze test items individually, determining their validity and relevance in the context of measurement (Polit & Beck, 2010). Calibration helps confirm that each item effectively distinguishes between different student abilities, which is crucial for delivering meaningful data. Once calibration is complete, researchers must interpret the results to decide which items to retain, modify, or remove, enhancing the overall quality and reliability of the instrument (Cohen dkk., 2011).

### **Continuous Evaluation and Future Implications**

Calibration is not a one-time procedure; it requires continuous evaluation to ensure the instrument's relevance over time (Neuman, 2014). This includes retesting on various sample groups and analyzing item performance through statistical methods. Well-calibrated instruments facilitate ongoing research, enabling longitudinal studies to track student progress and compare different teaching methods in Islamic education context (Schreiber dkk., 2006; Thompson & Melancon, 2011). Thus, effective calibration serves as a foundation for advancing educational practices and research in the field.

### **Challenges in Instrument Development and Calibration**

Developing research instruments presents challenges such as limited resources and a lack of understanding of valid development techniques. Researchers often face time and financial constraints that pressure them to complete instrument development quickly, potentially compromising quality (Azwar, 2015). Additionally, during the calibration process, insufficient sample size or representativeness can hinder the generalizability of results (Sugiyono, 2018). Moreover, the complexity of statistical analyses, particularly with methodologies like Item Response Theory (IRT), poses difficulties for those lacking strong statistical skills, impacting the accuracy of the calibration.

### **Solutions and Support Mechanisms**

To overcome these challenges, effective planning and resource management are crucial. Researchers are encouraged to attend training on instrument development and research methodology to strengthen their understanding (Creswell, 2014). Collaborating with statistical experts can also improve the calibration process. Moreover, educational institutions should offer training programs on these topics to better prepare researchers (Neuman, 2014). Institutional support, including access to statistical software and training facilities, is vital for equipping researchers to develop and calibrate instruments effectively, thus facilitating the generation of valid data for Islamic education (Creswell & Plano Clark, 2017).

## CONCLUSION

Developing an effective research instrument involves a series of structured stages aimed at ensuring both validity and reliability of the data collected. The process begins with clearly defining the constructs to be measured, followed by identifying relevant indicators, designing a detailed blueprint, writing well-formulated items, and conducting expert reviews. After drafting the items, a pilot test is conducted to assess initial performance. Calibration then becomes a critical phase, comprising steps such as collecting pilot data, performing item analysis, selecting high-performing items, and assembling the final version of the instrument. To enhance the precision and diagnostic power of the instrument, calibration methods such as Classical Test Theory (CTT) and Item Response Theory (IRT) are employed. These methods allow researchers to evaluate item difficulty, discrimination, and overall model fit. Addressing potential challenges such as construct misalignment and sampling errors requires iterative refinement, empirical testing, and collaboration with psychometric experts. Through these comprehensive stages and calibration techniques, the accuracy and effectiveness of research instruments in educational and social science contexts can be substantially improved.

## BIBLIOGRAPHY

- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Arikunto, S. (2019). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Azwar, S. (2015). *Metode Penelitian*. Yogyakarta: Pustaka Pelajar.
- Brown, J. D. (2015). *Understanding Research in Second Language Learning: A Teacher's Guide to Statistics and Research Design*. New York: Cambridge University Press.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education*. New York: Routledge.
- Creswell, J. W. (2014). *Research design: Qualitative, Quantitative, and Mixed Methods Approaches (4th ed.)*. SAGE Publications.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and Conducting Mixed Methods Research (3rd ed.)*. Thousand Oaks, CA: SAGE Publications.
- Djaali, D., & Muljono, P. (2021). *Pengukuran dan Evaluasi dalam Pendidikan*. Jakarta: Prenada Media.
- Fraenkel, J. R., & Wallen, N. E. (2009). *How to Design and Evaluate Research in Education (7th ed.)*. New York: McGraw Hill.
- Gay, L. R., Miles, G. E., & Airasian, N. R. (2012). *Educational Research: Competencies for Analysis and Applications (10th ed.)*. Boston: Pearson.
- Muhaimin, M. (2021). *Pendidikan Islam: Antara Teori dan Praktik*. Jakarta: Rajawali Press.
- Neuman, W. R. (2014). *Social Research Methods: Qualitative and Quantitative Approaches (7th ed.)*. Boston: Pearson.
- Polit, D. F., & Beck, C. T. (2010). *Nursing Research: Generating and Assessing Evidence for Nursing Practice (9th ed.)*. Philadelphia: Lippincott Williams & Wilkins.

- Prasetyo, E. (2020). *Literature Review in Research*. Yogyakarta: Andi Offset.
- Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*. *The Journal of Educational Research*, 99(6), 323–338.
- Silverman, D. (2016). Qualitative Research. *Journal of Qualitative Research*, 17(1), 35–50.
- Sugiyono. (2018). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif Dan R&D*. Bandung: Alfabeta.
- Sugiyono (2022). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Thompson, G., & Melancon, J. (2011). *Creating Effective Online Surveys*. New York: Wiley.