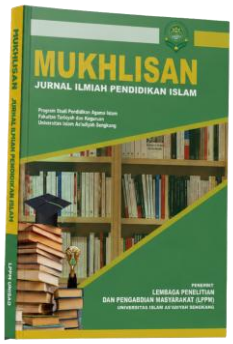
Vol. 3 | No. 1 | 2026



MUKHLISAN
Jurnal Ilmiah Pendidikan Islam
e-ISSN: 3047-3519
Available online at https://journal.unisad.ac.id/index.php/mukhlisan
Email: mukhlisan@unisad.ac.id

Design and Implementation of Valid and Reliable Learning Assessment Instruments

Desain dan Implementasi Instrumen Penilaian Pembelajaran yang Valid dan Reliabel

DOI: https://doi.org/10.1038/s41586-020-2649-2

Author Name	Author Affiliation
Ariswanto	Universitas Islam As'adiyah Sengkang

Corresponding Author: ariswanto879@gmail.com

Abtstract

Learning assessment is a crucial part of the educational process because it measures the achievement of learning objectives and monitors student progress. To ensure assessment results truly reflect student abilities, valid and reliable instruments are required. This article aims to explain the concept, design, and implementation of learning assessment instruments that meet validity and reliability criteria. Instrument design involves defining objectives, developing a framework, writing items or rubrics, conducting trials, and analyzing and revising. Instrument implementation is systematic through assessment, processing results, and providing follow-up in the form of remedial and enrichment programs. Valid instruments ensure that the measurement tool aligns with the competencies being measured, while reliability ensures the consistency of the measurement results. With proper planning and implementation, assessment instruments can improve the quality of learning and provide an accurate picture of student abilities.

Keywords: Assessment Instruments, Validity, Reliability, Assessment Design, Learning Evaluation

Article History
Received: 15 November 2025 | Revised: 10 February 2026 | Accepted: 12 March 2026

Abstrak

Penilaian pembelajaran merupakan bagian penting dalam proses pendidikan karena berfungsi untuk mengukur ketercapaian tujuan pembelajaran serta mengetahui perkembangan peserta didik. Agar hasil penilaian benar-benar mencerminkan kemampuan siswa, diperlukan instrumen yang valid dan reliabel. Artikel ini bertujuan untuk menjelaskan konsep, desain, dan implementasi instrumen penilaian pembelajaran yang memenuhi kriteria validitas dan reliabilitas. Desain instrumen dilakukan melalui tahapan penentuan tujuan, penyusunan kisi-kisi, penulisan butir soal atau rubrik, uji coba, serta analisis dan revisi. Implementasi instrumen dilakukan secara sistematis melalui pelaksanaan penilaian, pengolahan hasil, serta tindak lanjut berupa remedial dan pengayaan. Instrumen yang valid memastikan bahwa alat ukur sesuai dengan kompetensi yang diukur, sedangkan reliabilitas menjamin konsistensi hasil pengukuran. Dengan perencanaan dan pelaksanaan yang tepat, instrumen penilaian dapat meningkatkan kualitas pembelajaran dan memberikan gambaran yang akurat tentang kemampuan peserta didik.

Kata-kata Kunci: *Instrumen Penilaian, Validitas, Reliabilitas, Desain Penilaian, Evaluasi Pembelajaran*

Introduction

Assessment in learning is not simply the act of assigning grades to students. It is a conscious and planned process to determine how well students have understood the material being taught. Therefore, the focus is not solely on the final result, but also on the learning process they have undergone. Through assessment, teachers can understand students' abilities, development, and difficulties. Assessment results also serve as the basis for determining learning improvement measures to make the teaching and learning process more effective and meaningful.

In practice, assessments that focus solely on final scores often do not reflect students' actual abilities, as these scores are highly dependent on the quality of the instruments used. If the measuring instrument is not well designed, for example, if the questions do not align with competencies or learning objectives, the results obtained can be misleading and do not reflect true abilities. Therefore, assessment instruments must be designed systematically, based on clear indicators, and referring to the competencies being measured so that the assessment results are truly accurate, fair, and can be used as a basis for improving the learning process.

A good assessment instrument must meet two main requirements: validity and reliability. Validity means the instrument accurately measures the competency it is intended to measure, while reliability means consistent results when used under similar conditions. Without these two requirements, the assessment can be unfair and less objective.

In the modern learning era, teachers are required to design various instruments, both tests and non-tests, that can measure cognitive, affective, and psychomotor aspects in a balanced manner. This requires teachers to be skilled in designing instruments, developing assessment frameworks, creating assessment rubrics, and analyzing assessment results.

Based on this, this article provides an in-depth discussion of how to design and implement valid and reliable learning assessment instruments. With a sound

understanding of design and implementation, it is hoped that assessment can be an effective tool for improving the quality of learning and student learning outcomes.

The introduction section should present the background of the study, the urgency and significance of the research, relevant phenomena, and the objectives of the article. Authors are expected to discuss previous studies relevant to the topic as part of the theoretical review and scholarly argumentation. The theoretical framework should be integrated into the introduction rather than presented in a separate subsection. The introduction must also identify the research gap, novelty, or contribution of the current study compared to prior research. This section should be written systematically, logically, and critically by using recent and relevant references, particularly from reputable national and international scientific journals.

Research Method

The method used in this article is a library research approach. This approach was chosen because discussing the design and implementation of valid and reliable assessment instruments requires a strong theoretical foundation and analysis of various concepts developed by experts in the field of educational evaluation. Through the library research method, the author collected, read, understood, and analyzed various written sources relevant to the topic of discussion. (Sugiyono, 2019).

The data sources used in this article come from books on learning evaluation, scientific educational journals, regulations or guidelines for assessment in the curriculum, and other references related to the validity and reliability of instruments. These sources were selected based on their relevance to the topic, their level of trustworthiness, and their relevance to the current development of educational assessment systems.

Data collection techniques were conducted through documentation studies, namely by identifying and reviewing important theories regarding assessment concepts, principles of instrument development, types of validity and reliability,

and steps for implementing assessment in the classroom. The data obtained were then recorded, classified, and systematically organized according to the article's discussion framework.

Furthermore, descriptive-qualitative analysis is a method of analyzing data that emphasizes explanation and meaning, rather than numbers or statistical calculations. Data obtained from various sources such as books, journals, or documents is first read and understood, then explained coherently according to the topic of discussion. Therefore, the focus is on clearly describing and explaining the data content for easy comprehension.

In this analysis, the author doesn't simply copy existing theories, but rather attempts to understand the essence of each concept and then rephrase it in his own, simpler language. This process helps prevent the discussion from becoming too rigid or overly theoretical. This way, readers can grasp the main meaning of the concepts discussed without struggling to understand scientific terminology.

During the writing process, the author also synthesized, combining several expert opinions to develop a more comprehensive understanding of how to design valid and reliable assessment instruments. Furthermore, these concepts were interpreted to ensure they could be applied to real-life learning situations.

Using a literature review and descriptive-qualitative analysis, this article provides examples of the stages of designing and implementing a quality assessment instrument. This method also allows for in-depth discussion without being limited to a single case, allowing the results to serve as a reference for teachers and prospective educators in developing better assessment systems.

The research method section explains the procedures and methodological approaches employed in conducting the study. This section should contain information regarding the type and approach of the research, research setting, participants or objects of the study, data collection techniques, research instruments, and methods of data analysis. The explanation of the methodology should be concise yet sufficiently clear to enable readers to understand the research process and evaluate its scientific validity. For conceptual papers or

literature review articles, this section may be adapted into a description of the analytical framework or review method applied in the study.

Results and Discussion

Designing valid and reliable assessment instruments requires systematic and planned steps. Assessment instruments developed without careful planning tend to fail to comprehensively reflect students' abilities. Therefore, the first step is to clearly formulate learning objectives and then translate them into measurable indicators. These indicators serve as the basis for developing the test outline and test items. (Eko, P, W. 2018).

In the design process, creating a grid proved to be a crucial part, as it served as a conceptual map for the assessment. The grid helped teachers ensure that each question aligned with the core competencies, material, and level of thinking they were intended to measure. By using a grid, the risk of material deviation or imbalance in difficulty levels was minimized.

In terms of validity, the results of the discussion indicate that a good instrument must align the question content with the learning objectives. If the learning objectives emphasize analytical or problem-solving skills, then the questions should encourage students to think deeply, not just memorize facts. Validity is also evident in the use of clear, unambiguous language that is appropriate to the students' developmental level. Instruments that have undergone review and testing are more likely to be accurate in measuring competency.

Meanwhile, reliability relates to the consistency of assessment results. An instrument is considered reliable if it produces consistent results when used under relatively similar conditions. Achieving reliability requires a sufficient number of questions, a balanced variety of difficulty levels, and clear and detailed scoring guidelines. For non-test assessments such as observations or practicals, the use of a detailed assessment rubric is crucial for maintaining consistency and reducing teacher subjectivity.

The implementation of assessment instruments is also a determining factor in the quality of results. Assessments must be conducted in an orderly and professional manner. Teachers need to provide clear instructions, ensure students understand the assigned tasks, and create a conducive atmosphere. Furthermore, the time allotted for completing the tasks must be adjusted to the difficulty level of the questions so that students can demonstrate their abilities optimally.

After the assessment is conducted, the processing and analysis of the results is equally important. Assessment results need to be analyzed to determine the level of competency achievement, difficulties experienced by students, and the effectiveness of the learning methods implemented. Based on this analysis, teachers can determine follow-up actions, including remedial measures for students who have not yet achieved mastery and enrichment programs for those who have exceeded the target.

This discussion demonstrates that assessment is more than just assigning numbers, but rather a continuous process aimed at improving the quality of learning. Valid instruments ensure that what is being measured truly aligns with competencies, while reliability ensures confidence in the results. When designed and implemented systematically, assessment can be an effective tool for improving the overall quality of education.

A. Understanding Learning Assessment Instruments

Instrument Learning assessment can be understood as any tool teachers prepare to obtain an overview of student learning outcomes. Through these tools, teachers can determine students' level of understanding of the material, the skills they have mastered, and the attitudes they demonstrate during and after the learning process. Assessment tools are not simply tools for assigning grades, but also a medium for collecting targeted data so teachers can assess whether learning objectives have been achieved (Zainal, Arifin 2017).

In the educational process, assessments are not only intended to provide final grades but also to measure students' overall learning progress. Therefore, assessment instruments must be designed appropriately to objectively and accurately reflect students' abilities. If the instruments used are inappropriate, the assessment results can be misleading and not reflect true abilities.

Simply put, an assessment instrument can be defined as a measuring tool in learning. Just as a scale is used to measure weight, an assessment instrument is used to measure learning outcomes. However, in learning, what is measured is not only knowledge, but also attitudes and skills. Therefore, the forms of assessment instruments vary and are tailored to the aspects being measured.

Assessment instruments in learning are generally divided into two types: tests and non-tests. Tests are typically used to measure students' cognitive abilities or knowledge. Examples include multiple-choice questions, essays, true-false questions, and short answer questions. Non-test instruments, on the other hand, are used to assess attitudes and skills, such as observations, interviews, questionnaires, practical assessments, projects, portfolios, and reflective journals.

A good assessment instrument must have several important characteristics, including:

1. It must be aligned with the learning objectives. This means that what is being measured must align with the competencies to be achieved.
2. It must be clear and easy for students to understand so as not to cause misinterpretation.
3. Must be able to provide reliable results, so that teachers can use these results as a basis for decision making in learning.

Furthermore, assessment instruments also serve as evaluation tools for teachers. From the assessment results, teachers can determine whether the learning methods or strategies used are effective or need improvement. If many students have not achieved mastery, it means there is a part of the learning process that needs to be reviewed. Therefore, it can be concluded that learning

assessment instruments are important tools in education, used to systematically measure and evaluate student learning outcomes.

B. Concept of Validity And Reliability

1. Validity Concept

Validity is the degree to which an instrument accurately measures what it is supposed to measure. This means that the assessment tool truly aligns with the objectives and competencies it is intended to assess. If teachers want to measure students' critical thinking skills, the questions they create must require students to analyze, compare, or solve problems, not simply memorize material.

Simply put, validity can be likened to a doctor using the right instrument to examine a patient. If you want to measure blood pressure, you should use a blood pressure monitor, not a thermometer. Similarly, in learning, the instrument must be appropriate for the competency being measured.

In the world of educational evaluation, the concept of validity is widely discussed by experts such asAzwar, Saifuddinwhich explains that validity is related to the extent to which a measuring instrument is able to carry out its measuring function accurately. (Saifuddin, A. 2019)

Validity has several important aspects:

a. Content Validity

Content validity is the degree to which the content of an assessment instrument matches the learning objectives or competencies it aims to measure. This means that each question or assessment item must accurately represent the material and indicators taught. If the teacher has clearly formulated learning objectives, the instrument must reflect those objectives. In other words, content validity ensures that what is being tested is indeed derived from what has been learned. (Sudijono, A. 2018)

Content validity is closely related to lesson planning. Before creating questions, teachers must develop a framework that includes core competencies, indicators, material, and the format of the questions. This framework serves as a

guide to ensure the questions do not stray from the topic or deviate too much from the initial objectives. Without a framework, teachers are very likely to create questions that are actually good but do not align with the competencies they are trying to achieve.

Content validity also requires balanced coverage of the material. Questions should not focus solely on one small section of the material while leaving other important sections unmeasured. For example, if a chapter contains three main topics, the assessment should cover all three equally. If only one topic appears predominantly in the questions, the instrument lacks content validity because it doesn't represent the entire material.

Content validity emphasizes that assessment instruments must be relevant, representative, and aligned with learning objectives. If content validity is met, assessment results will be fairer and truly reflect students' mastery of the material. Therefore, teachers need to be meticulous and systematic in developing instruments to ensure assessments do not deviate from the planned competencies.

b. Construct Validity

Construct validity is the degree to which an instrument accurately measures a specific concept or ability according to its underlying theory. A construct itself is an abstract concept that cannot be directly observed, such as attitude, motivation, responsibility, honesty, or critical thinking skills. Because of its abstract nature, a construct must first be described in clear indicators so that it can be measured using an assessment instrument. (Saifuddin, A. 2019).

Construct validity ensures that each item or statement truly reflects the concept it is intended to measure, not other irrelevant aspects. For example, if a teacher wants to measure critical thinking skills, the questions should require students to analyze, compare, reason, or solve problems. If a question simply asks students to memorize definitions, it doesn't align with the critical thinking construct, even if the material is the same.

Construct validity also relates to the fit between the indicators and the theory used. Before creating an instrument, teachers must first understand the definition and characteristics of the concept they wish to measure. For example, to measure responsibility, indicators might include completing assignments on time, maintaining trust, and admitting mistakes. If the indicators used do not reflect the theoretical characteristics of responsibility, the instrument lacks construct validity.

Construct validity emphasizes that an instrument must be fully aligned with the concepts underlying its measurement. An instrument with good construct validity will be able to describe student abilities or attitudes more accurately and in-depth. Therefore, a clear understanding of theory and the formulation of indicators are key to maintaining construct validity in learning assessment.

c. Empirical Validity

Empirical validity is the level of accuracy of an instrument, demonstrated through trial results and real-world data. This means that a question or measuring instrument is not only assessed theoretically but must also be tested to determine whether it truly measures student abilities accurately. Therefore, empirical validity is based on analytical evidence after the instrument has been used, not simply on estimates or planning. (Sumarna, S. 2017).

Empirical validity is seen in a question's ability to distinguish between students who truly understand the material and those who don't. If a question is answered correctly by students with high scores and incorrectly by students with low scores, then the question has good discriminatory power. This indicates that the question is effective in measuring students' level of mastery of the material.

Conversely, if a question is too easy so that almost all students answer correctly, or too difficult so that almost all students answer incorrectly, then the question is empirically inadequate. Such questions fail to demonstrate differences in ability between students. Therefore, through analysis of trial results, teachers can identify which questions need to be improved, replaced, or retained.

Empirical validity emphasizes the importance of validating the instrument through real-world data before widespread use. By conducting trials and analysis, teachers can ensure that the instrument is truly effective in measuring the expected competencies. This makes assessment results more accurate, objective, and reliable as a basis for decision-making in learning.

d. Reliability Concept

Reliability is the level of consistency of measurement results. An instrument is considered reliable if it produces relatively consistent results when used under similar conditions. In other words, assessment results are trustworthy because they do not fluctuate unreasonably. (Saifuddin, A. 2019).

Reliability can be likened to a scale used to measure body weight. If a person weighs themselves several times within a short period of time and the results are always the same or nearly the same, then the scale is reliable. However, if the results fluctuate drastically for no apparent reason, the scale is unreliable.

In descriptive questions or practical assessments, reliability is highly dependent on the assessment rubric. If the rubric is unclear, different teachers may assign different scores to the same answer. This can reduce the level of trust in the assessment results.

It's important to understand that reliability is a prerequisite for trustworthy assessment results. However, reliability alone is not enough; the instrument must also be valid. A measuring instrument can be consistent but not accurate. Therefore, reliability and validity must go hand in hand. Overall, the concept of reliability emphasizes that good assessments are those that produce stable, consistent, and accountable results, enabling teachers to make more fair and accurate learning decisions.

C. Valid and Reliable Assessment Instrument Design

Designing valid and reliable assessment instruments is essentially a process of carefully designing learning measurement tools so that the assessment results

truly reflect students' true abilities and can be trusted. In education, assessment is not simply a matter of assigning numbers; it serves as the basis for making decisions about whether learning objectives are being achieved, whether teaching methods need improvement, and how students' overall development is progressing. Therefore, the instruments used should not be haphazardly created but rather systematically designed.

When a teacher designs an assessment instrument, the first thing they should consider is the learning objectives. A good instrument always stems from clear objectives. If the learning objective emphasizes analytical skills, then the measurement tool must also require students to think analytically. If practical skills are being measured, then the assessment should be hands-on, not just a written test. This is where validity lies. An instrument is considered valid if it truly measures the competencies targeted in the learning process. Validity is not only about the content of the questions being appropriate to the material, but also about the appropriateness of the assessment format to the type of ability being measured. (Djemari, M. 2016).

In practice, valid instrument design is inseparable from careful planning. Teachers need to ensure that each question item or assessment indicator directly relates to the established core competencies. No questions should deviate from the objectives. Furthermore, the language used must be clear, free from ambiguity, and appropriate to the student's developmental level. Ambiguous language can lead to students misunderstanding the questions, resulting in results that no longer reflect their true abilities. This situation will reduce the instrument's validity.

However, validity alone is not enough. An instrument must also be reliable. Reliability relates to the consistency of results. This means that if an instrument is used under relatively similar conditions, the results will also be relatively consistent. If a student demonstrates consistent ability, their grades should not fluctuate drastically without clear reason. A reliable instrument provides a sense of fairness because the results can be trusted. To achieve reliability, teachers

need to pay attention to clear instructions, consistent scoring methods, and the use of structured rubrics, especially when assessing essays or practicals. Without detailed scoring guidelines, assessments can be subjective and the results inconsistent.

Valid and reliable instrument design also requires pilot testing before official use. Through pilot testing, teachers can determine whether questions are too easy or too difficult, whether they are able to distinguish between students who understand the material and those who don't, and whether any questions are confusing. Based on these results, the instrument can be refined. This revision process is important because it demonstrates that assessment is not static but is continually refined for greater accuracy.

In a learning context such as Islamic Religious Education, instrument design must encompass all aspects of competency: knowledge, attitudes, and skills. Measuring understanding of religious teachings alone is not enough; it is also necessary to assess attitudes and practices in daily life. Therefore, good instrument design often combines several forms of assessment, such as written tests for cognitive aspects, observations for attitudes, and hands-on practice for skills. This comprehensive approach allows assessment results to better reflect students' actual conditions.

Overall, designing a valid and reliable assessment instrument is the process of designing a measurement tool that is targeted and consistent through careful planning, careful writing, clear language, the development of objective scoring guidelines, and testing and revision before use. A good instrument not only produces numbers but also provides an honest picture of student abilities. With valid and reliable instruments, teachers can make more informed learning decisions, students feel fairly assessed, and the quality of education can be continuously improved.

D. Implementation of Assessment Instruments

Implementation of assessment instruments is the actual process where previously designed measuring tools are actually used in learning activities to determine the extent to which students have achieved the targeted competencies. At this stage, assessment is no longer confined to the planning stage but enters the implementation, processing, and utilization stages of the results. Good implementation determines whether the instruments, which have been designed in a valid and reliable manner, are truly capable of producing accurate and reliable information.

In practice, assessment implementation should not be carried out spontaneously or without preparation. Teachers must ensure that the instruments used align with the formulated learning objectives. For example, if the learning objective is for students to be able to analyze a concept, then the assessment implementation must also provide space for students to think critically, not simply answer rote questions. This alignment between objectives and implementation is what ensures the assessment remains meaningful.

Before the exam, teachers need to conduct both technical and non-technical preparations. Technical preparations include preparing questions, answer sheets, scoring rubrics, attendance lists, and time management. Non-technical preparations include preparing the classroom atmosphere, explaining the rules, and creating a supportive environment for students to complete the assessment honestly and calmly. If the atmosphere is not conducive, for example, if it is too noisy or students do not understand the instructions, the assessment results can be affected by external factors unrelated to the students' abilities.

During assessments, teachers must administer the instruments according to the established procedures. For written tests, teachers ensure sufficient time, no cheating, and that all students understand the instructions. For practical assessments, teachers must observe each indicator specified in the rubric and assign scores based on clear criteria. For attitude assessments, teachers need to conduct ongoing observations, not just once, to ensure more objective results.

Consistency in implementation is crucial to maintaining the reliability of the results.

Once the assessment is complete, implementation doesn't stop at assigning grades. The next stage is processing and analyzing the results. Student grades need to be analyzed to identify achievement patterns. Teachers can identify areas the majority of students have mastered and areas where they remain weak. These results allow teachers to reflect on the learning process. If many students have not achieved a particular competency, perhaps the teaching method needs to be improved or the material re-explained. Thus, assessment becomes a tool for improving the quality of learning, not simply an administrative formality. (Nana, S. 2017).

Assessment implementation must also uphold the principles of fairness and transparency. Students have the right to know how they are being assessed and the basis for the scores. Therefore, using rubrics and scoring guidelines is crucial to preventing subjective assessments. When students understand the assessment criteria, they feel valued and are more motivated to improve.

Furthermore, assessment implementation must encompass all aspects of student abilities. Assessment should focus not only on cognitive aspects or knowledge, but also on attitudes and skills. In certain subjects, such as Islamic Religious Education, for example, in addition to assessing understanding of religious concepts through written tests, teachers also need to assess students' attitudes and religious practices through direct observation and practice. This comprehensive approach allows assessment results to better reflect students' actual conditions.

Overall, the implementation of assessment instruments is the process of using measurement tools in a planned, systematic, and objective manner to obtain accurate information about student abilities. This process encompasses preparation, implementation, processing of results, and utilization of the results for learning improvement. When done well, assessments not only produce scores

but also serve as a means for reflection, decision-making, and continuous improvement of educational quality.

The results and discussion section presents the findings of the research obtained through data collection and analysis processes. Findings may be presented in the form of narrative descriptions, tables, figures, or charts relevant to the objectives of the study. The discussion should critically analyze and interpret the findings by relating them to existing theories, concepts, and previous studies. Authors are expected not only to describe the data but also to provide scholarly interpretation and demonstrate the contribution of the findings to the development of knowledge within the field. Every table, figure, and chart included in the article must be properly numbered and titled.

Conclusion

Learning assessment instruments are tools teachers use to measure student learning achievement, including knowledge, attitudes, and skills. For assessment results to be credible, instruments must meet two main requirements: validity and reliability. Validity means the instrument accurately measures the competency it is intended to assess, while reliability means the assessment results are consistent and stable when used under the same conditions.

To produce valid and reliable instruments, systematic design based on learning objectives, clear and indicator-based question formulation, and objective and planned implementation are required. Good implementation ensures that assessments are fair and that the results can be used to evaluate and improve the learning process. Thus, assessments are not simply about assigning grades, but rather serve as a tool for improving the quality of education.

References

Arifin, Z. (2017). *Evaluasi Pembelajaran: Prinsip, Teknik, dan Prosedur.* Bandung: Remaja Rosdakarya.

Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Azwar, S. (2019). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.

Djemari, M. (2016). Pe*ngukuran, Penilaian, dan Evaluasi Pendidikan.* Yogyakarta: Parama Publishing.

Hasan, H. (2017). *Evaluasi Kurikulum*. Bandung: Remaja Rosdakarya.

Sudijono, A. (2018). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.

Sudjana, N. (2017). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.

Sugiyono. (2019). *Metode Penelitian Pendidikan*. Bandung: Alfabeta.

Sukardi. (2019). *Evaluasi Pendidikan: Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara.

Surapranata, S. (2017). An*alisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes.* Bandung: Remaja Rosdakarya.

Widoyoko, E. P. (2018). *Evaluasi Program Pembelajaran*. Yogyakarta: Pustaka Pelajar.