Development and Role of Measurement Tools in Educational Evaluation

[1]Namiyah Fitriani, [2] Sintya Rahmadewi, [3] Nurhadi, [4] Shuhada
[1,2,3]*Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia. [4] University College MAIWP International, Malaysia*
[1]*Namiyahfitriani@gmail.com, [2]Sintyarahmadewi1904@gmail.com, [3]nurhadi@pba.uin-malang.ac.id, [4] syuhada.zain@ucmi.edu.my.*

Abstract: *Evaluation in education and learning is essential for assessing student competencies, measuring teaching effectiveness, and ensuring continuous educational improvement. This study explores the development of measurement tools and their role in enhancing learning quality. Using a qualitative approach with a library research method, data collection is conducted through an in-depth review of academic literature, including books, journal articles, research reports, and educational policy documents. The study follows four stages: (1) identifying and selecting credible literature sources, (2) categorizing information based on key evaluation concepts, (3) conducting theoretical comparisons to identify similarities and differences, and (4) synthesizing findings to develop a comprehensive understanding of evaluation in education. The findings highlight the significance of valid and reliable measurement tools in accurately assessing learning outcomes. Effective evaluation not only measures student achievement but also functions as a feedback mechanism for educators, enabling them to refine teaching methods and curriculum design. Additionally, the study emphasizes that a well-structured evaluation system contributes to institutional development by promoting accountability, improving instructional strategies, and supporting evidence-based decision-making in education. To ensure the reliability of findings, the research applies a multi-source approach, comparing diverse academic references to validate information. The study concludes that an effective and structured evaluation process is vital for optimizing educational outcomes, fostering student success, and enhancing overall teaching quality. Recommendations for future research include expanding the scope of evaluation methods and integrating technology-based assessment tools for more accurate and comprehensive learning evaluations.*

Keywords*: Educational Assessment, Measurement Tools, Learning Evaluation, Teaching Effectiveness*

INTRODUCTION

The quality of a country's human resources (HR) is strongly influenced by the quality of its education. Education plays an important role in creating competent human resources, both in soft skills and hard skills,

according to employment needs and development demands. As a long-term investment, education has a noble and strategic value, especially for the younger generation who will determine the progress of a nation.(Ansyoriah et al., 2024)

Improving the quality of education can be done by improving the quality of learning and assessment systems(Sholeh, 2024). The two are interrelated: good learning produces good learning outcomes, while effective assessment helps teachers develop appropriate teaching strategies and motivate students.

Evaluation is important for assessing learning progress and providing a basis for future improvement, and is regulated in the National Education System Law and the National Education Standards(Maesyaroh, 2024). For optimal assessment, assessment standards are needed as guidelines for teachers and cooperation from various parties. Evaluation also plays a role in measuring the efficiency and effectiveness of achieving learning objectives.(Soulisa et al., 2022)

Assessment of learning outcomes aims to measure changes in student behavior in line with established instructional objectives(Sholeh & Muzakki, 2024). Currently, there are three domains of behavior that become the reference in developing assessment instruments, namely cognitive (thinking), affective (attitudes and emotions), and psychomotor (physical skills)(Johnson et al., 2024). This classification is known in education as Bloom's Taxonomy, which helps in designing and evaluating the learning process comprehensively.(Inanna et al., 2021)

Assessment on the other hand is a grading process that produces qualitative values(Sholeh et al., 2023). Tests are tools used to assess student learning success, while measurement is the process of comparing something to a predetermined standard, the result of which is a number(Ma et al., 2024). Evaluation is the determination of the value or implications of a behavior, be it individual or institutional behavior. Thus, evaluation has a broader meaning than measurement and assessment, where measurement, assessment,

and evaluation are in a hierarchy.(Ibrahim, 2012)

Educational progress relies heavily on effective evaluation(Ahmed et al., 2024). Evaluation plays an important role in measuring the achievement of learning objectives and identifying areas for improvement. Without structured evaluation, it is difficult to assess the success of an education program or ensure the learning process is up to standard. Therefore, the author is interested in exploring the topic of Evaluation in Education and Learning Processes (Assessment, Measurement and Testing: Steps in Developing Measurement Tools, and their Role in Education).

This study focuses on three main aspects: assessment, measurement and testing. The author will also discuss the steps to develop valid and reliable measurement tools, as well as the role of evaluation in improving the quality of learning. Through this research, the author hopes to contribute to educators in designing effective evaluations and to policy makers in developing quality education programs. Thus, evaluation is not only a measuring tool,

but also a driver of overall educational progress.

METHOD

This study employs a library research method as the primary approach to exploring and analyzing information related to evaluation in education and learning(McKenney & Reeves, 2012). Library research is conducted by reviewing various academic sources to gain a comprehensive understanding of a phenomenon. The literature used in this study includes academic books discussing evaluation theories in education, peer-reviewed journal articles, research reports from educational institutions, and relevant educational policy documents.

The data collection process in this research follows several stages. First, the researcher identifies relevant literature sources related to the research topic. Next, literature is searched and selected through academic databases such as Google Scholar, ResearchGate, Springer, and reputable educational journals. Once the appropriate literature is found, a structured reading and

note-taking process is carried out to examine the content, record key findings, and categorize information based on research themes.

For data analysis, this study adopts the thematic analysis method(Yin, 2009). The initial step involves categorizing information based on key concepts such as evaluation, measurement, and feedback in education. Subsequently, theoretical comparisons from various sources are conducted to identify similarities and differences in perspectives. Following this, a synthesis of findings is carried out to summarize the analysis results and develop a more comprehensive understanding of the role of evaluation in improving learning quality.

To ensure the validity and reliability of the research findings, several essential steps are taken(Fink, 2019). First, only credible literature sources, such as reputable journals and academic books from trusted publishers, are used. Second, a multiple sources approach is applied to compare and verify information, enhancing the accuracy of the findings. Lastly, the consistency and alignment of the theories used with the

synthesized results are rigorously examined to ensure the reliability of the study.

RESULT AND DISCUSSION

The Concept of Evaluation in the Education and Learning Process

Evaluation comes from the English word "evaluation". This word was adopted into the Indonesian vocabulary with a slight pronunciation adjustment to "evaluation." In the English dictionary, "evaluation" means evaluation, assessment, or appraisal. Etymologically, the word evaluation comes from English and refers to judgment or assessment. In Arabic, evaluation is known as "imtihan," which means test, as well as the term "khataman," which is used to assess the final results of the educational process.(Rohmad, 2017)

Based on this understanding, every evaluation activity has three implications, namely: (1) evaluation is a continuous process that begins before learning, not just at the end. (2) evaluation must be goal-directed to improve teaching. (3) evaluation requires appropriate measurement tools to gather information

needed for decision-making. Thus, evaluation helps educators assess learning progress and plan for future improvements.

In relation to learning activities, Norman E. Gronlund (1976) defines evaluation as follows: "Evaluation... a systematic process of determining the extent to which instructional objectives are achieved by pupils." Thus, evaluation is a systematic process used to assess the extent to which learners have achieved various learning objectives. (Febriana, 2021)

Law No. 20/2003 Article 39 Paragraph 2 on the National Education System states that educators are professionals who are responsible for planning and implementing learning, assessing learning outcomes, providing guidance, and conducting research and community service, especially in higher education. One of the important competencies that educators must have is the ability to evaluate the learning process and assess learning outcomes. This ability is the basis of educator professionalism. In addition, learning evaluation is also related to teacher skill assessment instruments, which include

evaluation indicators as assessment criteria.(A. Kurniawan, Nanang, et al., 2022)

Every activity carried out by humans must have a purpose. In the context of education, measurement and assessment carried out by teachers are also aimed at achieving certain goals.(Asrul et al., 2022) In accordance with the definition of evaluation that has been mentioned, the main purpose of measuring and assessing students is to plan, collect, and provide information needed in making alternative decisions. However, some experts do not distinguish between the purpose and function of measurement and assessment. (Wahib, 2021)

Assessment is used by teachers and educational supervisors to measure the effectiveness of teaching experiences, learning activities and methods applied. Thus, assessment has a very important role and function in the learning process.

Assessment

1. Definition of Assessment

Assessment is a systematic and structured process that aims to collect information about student learning

outcomes and the achievement of predetermined competencies). (Rahmah & Nasryah, 2019)

Assessment is a systematic and planned process to collect, analyze and interpret information about students' learning outcomes(Abror et al., 2024). This process involves the application of various methods and the use of diverse assessment tools, such as tests, observation, portfolios and self-assessment. The main purpose of assessment is to determine the extent to which learners have achieved the expected competencies, which include a series of abilities, knowledge and attitudes that must be mastered.

2. Types of Assessment

According to Aries in accordance with the objectives there are five types of assessment as follows:(Munaroh, 2024)

a. Diagnostic Assessment

Written and oral tests are used for diagnostic assessment (pretests or prates) which aims to detect student learning difficulties and find solutions. This assessment also helps teachers identify students' interests, strengths, weaknesses, and need for assistance in learning.

b. Formative Assessment

Formative assessment is conducted when students are struggling to complete tasks and focuses on the teaching and learning process. It helps educators adjust learning and students improve learning techniques. According to Jocsmani, this assessment monitors student progress within a segment, such as a unit or chapter, through daily tests, quizzes and assignments.

c. Summative Assessment

Summative assessments are used to determine final grades and measure the extent to which students understand the material before moving on to the next topic. This assessment technique is tailored to the needs of students and teacher judgment, and can be done individually or in groups. The form of the test can be oral, written, or

performance to measure process skills.

d. Selective Assessment

Selective assessment is an assessment carried out to select students who will represent the school in certain competitions, as well as for broader purposes such as new student admissions or labor recruitment. This selection assesses skills, talents, interests, attitudes, and behavior. According to Jocsmani (1988:27), the purpose of this assessment is to select students due to limitations, especially in educational facilities.

e. Placement Assessment

Placement assessment is an assessment that aims to determine the prerequisite skills needed before starting a learning program. It focuses on students' readiness for the new program and the suitability of the program to their abilities. The assessment is conducted when there is a need to place students in an educational program that suits their abilities.

3. Assessment Process

Teachers design assessment tools that cover three aspects of ability: affective, cognitive, and psychomotor. Teachers carry out the assessment through five main stages :

a. reviewing the syllabus ,

b. creating an assessment grid

c. conducting the assessment

d. analyzing the results ,

e. compiling the report.(Mustika et al., 2021)

Measurement

1. Definition of Measurement

Measurement, according to Cangelosi (1995), is the process of collecting data through empirical observation to obtain information related to a predetermined goal. In this case, teachers assess student achievement by observing their behavior, evaluating performance, listening to speech, and using senses such as seeing, hearing, and touching. Zainul and Nasution (2001) add that measurement has two main characteristics: first, the use of certain

numbers or scales; second, following predetermined rules or formulas. (Wajihah, 2024)

According to Djali and Mulyono (2000), there are measurement objects in the field of education which consist of 8 objects, namely :

a. Learning outcomes or achievements. Learning outcomes or student achievement is an achievement obtained in the implementation of learning activities. Learning outcomes use measuring instruments in the form of tests .

b. Attitude. measured by Student attitudes here are measured using attitude scale instruments such as those developed by Likert, etc. In measuring this attitude scale, the teacher makes an instrument in the form of a student attitude questionnaire using certain indicators related to student attitudes .

c. Motivation. In this case, motivation measurement is also carried out using a scale-shaped instrument developed and adjusted to motivational theories.

The teacher makes an instrument in the form of a student learning motivation questionnaire using existing indicators and in accordance with student learning motivation.

d. Intelligence. Intelligence here is measured using intelligence tests, for example, such as the Stanford Binet test, Simon Binet test, etc .

e. Talent. Different student talents can be measured using aptitude tests. Student aptitude tests include artistic, mechanical, sports, numerical, etc. aptitude tests.

f. Emotional Intelligence. Emotional intelligence, like attitude and motivation, is also measured using instruments developed and adjusted to emotional theories.

g. Interests. Interest is also measured using instruments developed and adjusted to the theories of student interest in learning.

h. Personality Measurement of students' personalities is carried out using measuring instruments in the form of personality tests, there are several

personality tests, one of which is the q sort personality test.(A. Kurniawan, Febriant, et al., 2022)

It can be concluded that measurement in learning is an assessment process that gives value to the process and learning outcomes, based on certain criteria, rules, or formulations that are clear and in accordance with predetermined goals, to produce decisions about the process and learning outcomes.(Faiz et al., 2022)

2. Measurement Tools and Methods

Measurement Tools and Methods refer to the tools and techniques used to collect data and information in a research or evaluation. Measurement tools include physical or digital instruments, such as questionnaires, tests, measuring instruments (e.g., rulers, scales), and data analysis software. These tools are designed to ensure that the data collected are accurate and reliable .

Measurement Methods are the approaches or strategies used to collect data, such as surveys, observations, experiments, or interviews. The method determines how the data will be obtained and analyzed, as well as ensuring that the measurement process is in line with the objectives of the research.The right combination of measurement tools and methods is essential to obtain valid and relevant results in a research or evaluation.

a. Quantitative Measurement Tools

1) Graded Scale

The scale describes the value in the form of numbers related to the results of a consideration. As stated by Oppenheim, "Rating gives a numerical value to a type of judgment," so the scale is always presented in numerical format.

2) Questionnaires

Questionnaires are non-test evaluation instruments that aim to measure affective aspects both inside and outside the classroom. In general, a questionnaire consists of a list of questions that must be answered by the individual to be measured

(respondent). Measurements made through questionnaires tend to focus on the opinions, feelings of learners, as well as their views on learning Arabic, and others. (Riinawati, 2021)

b. Qualitative measurement tools

1) Interview

Interview is a method used to obtain answers from respondents through a one-sided question-and-answer process. In social research, interviews are used to collect data from informants, while Arabic teachers usually ask oral questions directly in class to find out students' understanding of certain situations.

2) Observation

Observation is an assessment method carried out by educators using the senses directly to assess the implementation of actions and their accuracy. This activity uses pre-designed instruments to identify learners' attitudes. In Arabic lessons, the aspects observed include students' accuracy, diligence, and speed in solving problems. (Tina, 2023)

When viewed from the technical implementation, observation can be achieved in three ways, namely:

a) Direct observation, which is observation carried out directly on the object under investigation .

b) Indirect observation, which is observation made through intermediaries, both techniques and certain tools .

c) Participatory observation, which is an observation carried out by taking part or involving oneself in the situation of the object under study.(Kuantitatif, 2016)

Test

1. Definition of test

A test is a tool for gathering information, consisting of a series of items or questions designed to be

administered to students under certain conditions. When tests are administered, measurement also takes place, although measurement does not always involve tests. As part of measurement, tests must have consistent and accountable standards. In addition to being used to obtain information about a person's condition, tests can also serve to set goals, so that the activity can be measured.(Susilawati, 2018)

2. Forms of Tests

In terms of usefulness for measuring students, there are three kinds of tests:

a   Diagnostic tests

Diagnostic tests are tests used to identify student weaknesses, so that appropriate treatment can be given based on these weaknesses. In general, this test is known as an entrance assessment or in English it is called an entering behavior test. This test aims to measure the level of mastery of basic knowledge required to receive advanced knowledge. Therefore, it is also known as a pre-

requisite test. In addition, this test functions as a placement test.

b   Formative Test

From the meaning of the word "form" which is the basis for the term formative, formative evaluation aims to determine the extent to which students have developed after following a particular program. Formative evaluations, or formative tests, are given at the end of each program. This test serves as a post-test or final test of the program.

c   Summative Test

Summative tests or summative evaluations are conducted after the completion of the administration of a group of programs or a larger program. In the school context, formative tests can be equated with daily tests, while summative tests are similar to general tests which are usually carried out at the end of the semester.(Arikunto, 2010)

3. Test Validity and Reliability

a.   Validity

Validity is a measure of the accuracy and precision of a measuring instrument (test) in carrying out its function. Test validity shows how well the test measures what it is supposed to measure. In other words, validity assesses whether the test actually achieves its intended purpose. The higher the validity of a test, the more accurate it is in reflecting the characteristics or conditions of the object being measured.

b. Reliability

Reliability comes from the word reliability, which means the level of confidence in the results of a measurement. A test is considered reliable if it produces consistent results when given to the same group at different times or occasions. The reliability of the assessment tool refers to the accuracy and persistence of the tool in assessing the target being measured, meaning that the tool will provide relatively the same results every time it is used.(Ramadhan et al., 2024)

Steps to develop a Measurement Tool

In general, a tool refers to something that can assist a person in completing a task or achieving a goal in a more effective and efficient manner. The term "tool" is often also referred to as "instrument". Thus, an instrument is a device used by researchers to collect data through the measurement process. (Afifah, 2021)

General steps in developing a measurement tool:

1. Establish Measurement Objectives
   Determine what you want to measure (e.g., knowledge, attitudes, skills). Clearly define the purpose of the measurement, such as assessing learning outcomes, diagnosing learning difficulties, or evaluating a program.

2. Identify Competencies or Indicators to be Measured
   Determine the specific competencies or indicators that are relevant to the purpose of the measure. Use curriculum standards or taxonomies (such as Bloom's Taxonomy) as a reference.

3. Determine the Type of Measurement Tool

   Choose an appropriate form of measurement tool, such as a written test (multiple choice, essay), observation, questionnaire, or performance assessment. (Pariama et al., 2025)

4. Develop a Grid (Blue Print)

   Create a grid that includes indicators, materials, cognitive levels (for example, C1-C6 in Bloom's Taxonomy), and question forms. Make sure the grid reflects a balanced and representative proportion.

5. Developing Problem Items or Instruments

   Create question items or instruments based on the grid. Make sure the questions/instruments are valid, reliable, and in accordance with the indicators being measured. Use clear language and avoid ambiguity.

6. Try Out

   Test the measurement tool on a representative sample. Analyze the test results to evaluate the validity, reliability, difficulty level, and differentiation of the questions.

7. Revision and Refinement

   Improve the measurement tool based on the results of the trial analysis. Eliminate or revise items that do not meet the criteria.

8. Implementation

   Use the refined measurement tool for actual measurement purposes. Ensure that the measurement is carried out objectively and consistently.

9. Evaluation and Reflection

   Evaluate the effectiveness of the measurement tool after use. Reflect for further improvement and development.(H. Kurniawan, 2022)

Measurement Objectives

The purposes of measurement in the context of education and evaluation can vary, but in general, they can be divided into several main categories:

1. Assessing Learning Outcomes

   One of the main purposes of measurement is to assess the extent to which learners have achieved set learning objectives. This includes measuring the

knowledge, skills and attitudes that have been acquired during the learning process.

2. Diagnosing Learning Difficulties

Measurement is also used to identify difficulties or obstacles that learners face in the learning process. By knowing the areas that need improvement, teachers can provide appropriate interventions to help students.

3. Evaluation of Teaching Programs

Measurement can be used to evaluate the effectiveness of a teaching program or curriculum. By analyzing the measurement results, educators can determine whether the program is successful in achieving the desired goals or needs to be adjusted.(Achadah, 2019)

4. Providing Feedback

Measurement provides feedback to learners on their progress. This feedback is important for student motivation and self-development, as well as helping them understand areas for improvement.

5. Determining Eligibility or Placement

In some cases, measurement is used to determine learners' eligibility for a particular program or for placement in a learning group that matches their abilities.

6. Supporting Decision Making

Measurement results can be used as a basis for decision-making in educational contexts, such as promotion, graduation, or curriculum development.

7. Setting Standards and Criteria

Measurement helps in setting clear standards and criteria for assessment, so that all learners can be assessed in a fair and consistent way.

8. Improving the Quality of Education

By measuring systematically, educational institutions can identify strengths and weaknesses in the learning process, which in turn can be used to improve the overall quality of education.(Efendi & Sholeh, 2023)

By understanding the purpose of measurement, educators can design more effective and relevant measurement tools and ensure that the evaluation process provides useful information for learner development and educational program improvement.

The Urgency/Role of Measurement and Assessment

Measurement and assessment are very important activities in the teaching process. Both are one of the four main tasks of a teacher. The four main tasks include planning, implementing, assessing the success of teaching, and providing guidance. In teaching practice, these four activities are interrelated and form an inseparable whole.

In carrying out their teaching duties, a teacher tries to create learning situations that allow students to learn, motivate them, present teaching materials, and use methods and media that have been prepared(Habibulloh et al., 2024). In addition, teachers also process and interpret student learning outcomes, and make decisions to improve the effectiveness of future teaching. To achieve optimal educational goals, teachers provide guidance to students by trying to understand the learning difficulties they face and their backgrounds, and providing assistance in accordance with their abilities and authority to overcome these problems .

Measurement and assessment contribute significantly to all components of teaching and learning activities(Habibulloh et al., 2024). The information obtained from analyzing the results of measurement and assessment is very important for making policies that need to be taken by a teacher to improve the quality of teaching and learning activities in the classroom.

Measurement and assessment play an important role in assessing the performance of the components of teaching and learning activities to achieve the ultimate goal. Information from the analysis of measurement and assessment results is needed to assist teachers in formulating policies to improve the quality of learning in the classroom.

If student learning outcomes are unsatisfactory, the steps that teachers can take are as follows:

1. If many students do not achieve satisfactory results on most of the material, the teaching program needs to be revised and repeated.

2. If only a certain part of the material is unsatisfactory, the teacher should include that part in the next teaching plan.

3. If most students are unsatisfactory, the teaching program should be repeated.

4. If only a few students are unsatisfactory, the teacher should provide remedial programs for them.

The results of measurement and assessment also affect the steps teachers take in developing teaching and learning activities. If teachers do not realize and utilize these results, ineffective teaching and learning activities will continue, and the results of the teaching program will not be optimal(Habibulloh et al., 2025). Teachers' failure to recognize students' learning difficulties can lead to ongoing problems that negatively affect their education. In addition, data from measurement and assessment are also important for providing guidance. This information helps teachers identify successful and unsuccessful students, so that successful students can be given enrichment activities, while unsuccessful ones can be given remedial activities.

The implementation of teaching and learning activities, the results of measurement and assessment of student learning outcomes will influence the steps that teachers must take to develop teaching and learning activities. If a teacher does not realize and utilize the results of measurement and assessment in the implementation of teaching and learning activities, then ineffective activities will continue, and the results of the teaching program will not be optimal. In addition, if teachers are not aware of some students' failure to complete learning tasks in a teaching program, these students will face ongoing difficulties, which may negatively affect their education in primary school and their overall educational journey.

Another important role of measurement and assessment is in providing guidance. The data obtained from this activity provides information about students who successfully complete learning tasks well and those who do not. Successful students can be provided with enrichment activities and materials, while unsuccessful ones can be provided with remedial activities and materials .

The role of measurement and assessment is very broad, as stated by Linn and Gronlund (1995), who mentioned that assessment goes hand in hand with the teaching process from beginning to end. Dressel (1991) also emphasizes that education is a complex process involving the selection of concepts, values, and skills, as well as the planning of experiences to facilitate the mastery of these ideas. Therefore, assessment is inevitable in education.

The above highlights the importance of assessment in teaching programs and the breadth of what needs to be assessed. The effectiveness of everything planned in teaching and learning activities should be monitored and measured through assessment. Given the significance of assessment in school teaching, teachers should have the knowledge and skills to conduct assessment. This is one of the competencies that teachers must have, as expressed in the results of the teacher education curriculum workshop organized by P3G. Competence in assessment at least includes the ability to develop assessment instruments, especially

tests; administer tests or other instruments; and process and interpret learning outcome data.

CONCLUSION

Evaluation in the education and learning process is a fundamental element that plays a crucial role in assessing students' understanding, skills, and the effectiveness of teaching methods. Through assessment, measurement, and testing, educators can obtain valuable insights that help improve the quality of instruction and ensure meaningful learning experiences for students. The development of valid and reliable measurement tools is essential to ensure that evaluation accurately reflects students' competencies and provides constructive feedback for educators. To enhance the effectiveness of evaluation in education, it is recommended that future research focuses on innovative approaches to designing measurement tools that are adaptable to diverse learning contexts. Educators should integrate technological advancements, such as digital assessment platforms and data-driven analytics, to improve the accuracy and

efficiency of evaluations. Furthermore, collaboration between educators, policymakers, and researchers is necessary to establish standardized yet flexible evaluation frameworks that cater to different educational needs. Continuous improvement in evaluation practices will not only contribute to better educational outcomes but also foster an inclusive and dynamic learning environment. By developing and implementing effective evaluation strategies, educators can support students in reaching their full potential and create a more equitable and high-quality education system.

REFERENCE

Abror, S., Mutrofin, M., & Hardinanto, E. (2024). Reimagining Teacher Professional Development to Link Theory and Practice. *JTL: Journal of Teaching and Learning*, *1*(1), 22–36.

Achadah, A. (2019). Evaluasi dalam pendidikan sebagai alat ukur hasil belajar. *An-Nuha Jurnal Pendidikan Islam*, *6*(1), 91–107.

Afifah, W. (2021). *Pengembangan Alat Evaluasi Pembelajaran Matematika Berbasis Two Tier Multiple Choice Berlevel HOTS Menggunakan Ispring Suite 9* [PhD Thesis, Universitas Muhammadiyah Gresik]. http://eprints.umg.ac.id/5170/

Ahmed, N. H., Andersion, J., & Martínez, A. (2024). Innovative Blended Learning Approaches to Enhance Student Engagement in University. *JTL: Journal of Teaching and Learning*, *1*(1), 1–21.

Ansyoriah, S., Boeriswati, E., & Asep Supriyana, S. S. (2024). *Evaluasi Pembelajaran Bahasa*. Penerbit Adab. https://books.google.com/books

Arikunto, S. (2010). Dasar-dasar evaluasi pendidikan. jakarta: Bumi Aksara. *Pedagogy*.

Asrul, A., Saragih, A. H., & Mukhtar, M. (2022). *Evaluasi pembelajaran*. http://repository.uinsu.ac.id.

Efendi, N., & Sholeh, M. I. (2023). Manajemen pendidikan dalam meningkatkan mutu pembelajaran. *Academicus: Journal of Teaching and Learning*, *2*(2), 68–85.

Faiz, A., Putra, N. P., & Nugraha, F. (2022). Memahami makna tes, pengukuran (measurement), penilaian (assessment), dan evaluasi (evaluation) dalam pendidikan. *Jurnal Education and Development*, *10*(3), 492–495.

Febriana, R. (2021). *Evaluasi pembelajaran*. Bumi Aksara. https://books.google.com/books

Fink, A. (2019). *Conducting research literature reviews: From the internet to paper.* Sage publications.

Habibulloh, M., Sholeh, M. I., Farisy, F. A., & Nasihuddin, M. (2025). Effectiveness of Leadership in Managing Religious-Based Educational Institutions. *IJEMR:*

International Journal of Education Management and Religion, *2*(1).

Habibulloh, M., Sholeh, M. I., & Idawati, K. (2024). Exploring Technological Innovations and Approaches in Modern Education. *SAHRI: Journal of Studies in Academic, Humanities, Research, and Innovation*, *1*(1), 49–66.

Ibnu Sholeh, Muh., Tanzeh, A., Fuadi, I., & Kojin. (2023). Kepemimpinan Profetik (Study Proses Peningkatan Lembaga Pendidikan Islam di Indonesia). *JMPI: Jurnal Manajemen, Pendidikan dan Pemikiran Islam*, *1*(1), 27–44. https://doi.org/10.71305/jmpi.v1i1.9

Ibrahim, M. M. (2012). *Pengembangan Pengukuran Non-Tes Bidang Pendidikan*. Makassar: UIN Press.

Inanna, I., RAHMATULLAH, R., & Hasan, M. (2021). *Evaluasi pembelajaran: Teori dan praktek*. Tahta Media Group. https://eprints.unm.ac.id/21188/

Johnson, E., Mendoza, C., & Sobirin, M. S. (2024). Strategies of School Principals in Improving Educational Quality: An Analysis of Best Practices in American Schools. *JMPI: Jurnal Manajemen, Pendidikan, Dan Pemikiran Islam*, *2*(2).

Kuantitatif, P. P. (2016). Metode Penelitian Kunatitatif Kualitatif dan R&D. *Alfabeta, Bandung*.

Kurniawan, A., Febriant, A. N., Hardianti, T., Ichsan, I., Desy, D., Risan, R., Maya Sari, D. M., Sitopu, J. W., Dewi, R. S., & Sianipar, D. (2022). *Evaluasi pembelajaran*. PT. GLOBAL EKSEKUTIF TEKNOLOGI.

http://repository.uki.ac.id/8714/3/EvaluasiPembelajaran.pdf

Kurniawan, A., Nanang, N., Arifannisa, A., Noflidaputri, R., Supriyadi, A., Rahman, A. A., Arrobi, J., Jamaludin, J., Arissandi, F., & Sianipar, D. (2022). *Metode pembelajaran di era digital 4.0*. PT Global Eksekutif Teknologi. http://repository.uki.ac.id/9835/

Kurniawan, H. (2022). *Pengantar praktis penyusunan instrumen penelitian*. Deepublish. https://books.google.com/books

Ma, X. Z., Ertmer, P. A., & Pelgrumen, C. P. (2024). The Impact of Technology Integration on Student Learning Outcomes. *JTL: Journal of Teaching and Learning*, *1*(1), 73–90.

Maesyaroh, A. (2024). Strategi Kepala Madrasah Aliyah Swasta Al Falah Wuluhan Jember Menuju Madrasah Unggul. *JMPI: Jurnal Manajemen, Pendidikan, dan Pemikiran Islam, 2*(1).

McKenney, S. E., & Reeves, T. C. (2012). *Conducting educational design research*. Routledge.

Munaroh, N. L. (2024). Asesmen dalam Pendidikan: Memahami Konsep, Fungsi dan Penerapannya. *Dewantara: Jurnal Pendidikan Sosial Humaniora*, *3*(3), 281–297.

Mustika, D., Ambiyar, A., & Aziz, I. (2021). Proses penilaian hasil belajar kurikulum 2013 di Sekolah Dasar. *Jurnal Basicedu*, *5*(6), 6158–6167.

Pariama, B., Elfarisyah, D., Harahap, K., Prabudi, M. I., & Fiqri, M. A. (2025). Bentuk dan Teknik Evaluasi Hasil Belajar. *MUDABBIR Journal*

Research and Education Studies, *5*(1), 01–13.

Rahmah, A. A., & Nasryah, C. E. (2019). Evaluasi pembelajaran. *Evaluasi Pembelajaran*. https://repository.bbg.ac.id/handle/83 7

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, *6*(2), 10967–10975.

Riinawati, R. (2021). *Pengantar Evaluasi Pendidikan*. Thema Publishing. https://perpustakaan.iaiskjmalang.ac. id/wp-content/uploads/2023/11/18.-Pengantar-Evaluasi-Pendidikan-ISI.pdf

Rohmad, R. (2017). *Pengembangan Instrumen Evaluasi dan Penelitian*.

Sholeh, M. I. (2024). Pengaruh Kinerja Guru dan Pengembangan Kurikulum Terhadap Prestasi Belajar Siswa di SDI Al-Badar Tulungagung. *Jurnal Karya Ilmiah Pendidik dan Praktisi SD&MI (JKIPP)*, *3*(1).

Sholeh, M. I., & Muzakki, H. (2024). The Effectiveness of Principal Leadership in Managing a Tahfidz- based Curriculum. *Indonesian Research Journal in Education*, *8*(2).

Soulisa, I., Supratman, M., Rosfiani, O., Renaldi, R., Utomo, W. T., Hermawan, C. M., Ariati, C., Riyanti, A., Tauran, S. F., & Astiswijaya, N. (2022). *Evaluasi Pembelajaran*. Penerbit Widina. https://books.google.com/books

Susilawati, D. (2018). *Tes dan pengukuran*. UPI Sumedang Press. https://books.google.com/books

Tina, W. (2023). *Kompetensi Pedagogik Guru Dalam Evaluasi Sumatif Pada Pembelajaran Sejarah Semester Genap Di Sma Negeri 1 Buay Madang Kecamatan Buay Madang Kabupaten Oku Timur Tahun Ajaran 2022/2023*. http://digilib.unila.ac.id/id/eprint/765 22

Wahib, A. (2021). Manajemen evaluasi program supervisi pendidikan dalam meningkatkan mutu pendidikan. *Auladuna: Jurnal Prodi Pendidikan Guru Madrasah Ibtidaiyah*, *3*(1), 91–104.

Wajihah, N. (2024). Evaluasi Program Pendidikan Agama Islam di Madrasah: Perspektif Stakeholder dan Implikasinya untuk Perbaikan. *Res: Review Of Education Studies*, *1*(1), 64–83.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). SAGE Publications.