

The Quality Of Learning Outcome Tests In The Learning Process

Ismail Marzuki*

[*ismailmarzuki@umt.ac.id](mailto:ismailmarzuki@umt.ac.id)

*Postgraduate Program, Master of Islamic Religious Education Muhammadiyah University of Tangerang

ABSTRACT

The quality of a test of learning outcomes is largely determined by the quality of the items. Learning outcomes test that contains high-quality, though deep items a small amount will be far more useful than a test of learning outcomes contains dozens of low-quality items that will degrade the test and function misleading measurement results. Processing learning outcomes tests in order to improve the learning process can be done by making a problem analysis (item analysis). Analysis Quantitative is done by testing the instruments that have been Qualitatively analyzed to a number of students who have characteristics the same as students who will be tested with the instrument. Characteristics Quantitative internals are intended to include validity, reliability, power distinguishing, difficulty level and effectiveness of deceptive functions (distractor). As for the benefits or usefulness of teacher-made question analysis according to Suke Silverius in his book entitled Evaluation of Learning Outcomes and Feedback, is as follows: (1) Determine whether the items functioning exactly as intended by the teacher, (2) Feedback for students regarding their performance and is the basis for class discussion, (3) Feedback for teachers about student learning difficulties, (4) Areas curriculum that requires improvement, (5) Improvement of items, and (6) Improve question writing skills. Learning outcomes are the level of mastery achieved by students in follow the learning program in accordance with the educational objectives that have been set.

Keywords: *evaluation, test quality and learning outcomes*

Copyright © 2025 Tadarus Tarbawy, Jurnal Kajian Islam dan Pendidikan is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

A. INTRODUCTION

No matter how frequently evaluation systems change, their presence in education and learning cannot be ignored. The interconnectedness between educational objectives, learning activities/processes, and evaluation is so close that they are inseparable. Evaluation is necessary to assess the success of learning activities in achieving educational goals.

There are three main reasons why educational activities always require evaluation. *First*, when viewed from a process approach, there is a relationship between educational objectives, learning processes and evaluation. *Second*, the activity of evaluating learning outcomes is one of the characteristics of a professional educator. *Third*, when viewed

from an institutional approach, educational activities are management activities, which include activities *planning, programming, organizing, actuating, controlling, And evaluating*.

In the field of learning, evaluation aims to (1) determine the specific learning content competencies possessed by students, (2) improve the teaching and learning process. Meanwhile, in the field of learning outcomes, evaluation aims to (1) determine differences in student abilities, (2) measure their success both individually and in groups. Therefore, adequate instruments are needed. This is because the success of revealing the results of student learning as they are (the objectivity of the assessment results) is very dependent on the quality of the assessment tool in addition to the method of implementation.

Theoretically, students in a class constitute a heterogeneous population or group, meaning each student has different abilities from the others. Therefore, when a test is conducted, the results will be reflected in a normal curve. Most students fall in the middle zone, a small portion fall in the left tail, and a small portion fall in the right tail of the curve. However, if the conditions after the test results are analyzed are not as expected in the normal curve, or in other words, if all students (*testee*) get a bad score or vice versa if all students get a good score, then there must be "something" with the test.

The quality of a learning outcome test is largely determined by the quality of its items. A learning outcome test containing high-quality items, even if only a small number, will be far more useful than one containing dozens of low-quality items. Low-quality items not only diminish the test's functionality but also produce misleading measurement results.

One way to determine the quality of a learning outcome test is to conduct a question analysis (*item analysis*). Item analysis can primarily be performed for objective tests. This does not mean that essay tests cannot be analyzed, but there are no standard guidelines for analyzing essay test items. Therefore, objective learning outcome tests are easier to analyze than essay learning outcome tests, both in terms of validity, reliability, difficulty level, and discriminatory power. Assessment of student learning activities and outcomes is intended to collect data as consideration in assisting further development and/or determining student success. In addition to this assessment, student assessment is an integral part of educational activities, intended to obtain information about student learning activities and progress.

B. METHOD

The method used in this study was library research, which is an approach conducted by reviewing and analysing various relevant literature sources, including books, scientific articles, research reports, and education policy documents. This study did not involve direct field data collection but focused on conceptual and theoretical exploration of the issue of teacher professionalism and its contribution to

improving the quality of education, particularly in shaping the character of students.

Through this approach, researchers can identify, classify, and synthesise findings from previous studies discussing teacher competency development, the role of teachers as role models, and strategies for improving the quality of educators in the context of character education. This approach enables researchers to build a solid conceptual framework, reveal research gaps, and formulate practical implications for the development of education policies and practices oriented towards shaping student character.

C. RESULTS AND DISCUSSION

1. Test Instrument Quality

One of the most effective ways to improve the learning process is by evaluating the learning outcome tests obtained from the learning process itself. In other words, we process the test results in such a way that we can identify which components of the learning process are weak.

Processing learning outcome tests in order to improve the learning process can be done by creating a question analysis (*item analysis*).

a. Understanding Test Item Analysis

Question analysis is an activity related to the process of collecting, summarizing, and using information about students' answers to test questions. Sudjana (1991: 135) states that test item analysis is the study of test questions in order to obtain a set of questions that have adequate quality. According to Saifuddin (2000: 130), test item analysis is the testing of all test items based on empirical items (data obtained from the results of the actual test administration), in order to obtain evidence regarding the quality of the test items.

Based on the definitions above, it can be stated that test item analysis is the process of assessing learning outcome test items based on students' answers to the test, so that the quality of a test as a tool for measuring student learning outcomes can be determined.

b. Elements of Test Item Analysis

An instrument should be analyzed before use. There are two types of analysis available: qualitative and quantitative. Qualitative analysis is conducted by colleagues within the same field of expertise. The goal is to assess the material, construction, and language used to ensure they meet guidelines and are comprehensible to students.

Quantitative analysis is conducted by testing the instrument that has been analyzed qualitatively on a number of students who have the same characteristics as the students who will be tested with the instrument. (Majid, 2004:223)

Quantitative question analysis emphasizes the analysis of the internal characteristics of the test through empirically obtained data. (Sumarna, 2005:10) The internal characteristics quantitatively intended include validity, reliability, discriminating power, level of difficulty and effectiveness of the distractor function.

1) Test Validity

Test validity needs to be determined to determine the test's quality in relation to measuring what it is supposed to measure. The word "valid" is defined as "precise, correct, authentic, and valid." Therefore, the word "validity" can be interpreted as accuracy, truth, legitimacy, or validity. When the word "valid" is associated with the test's function as a measuring tool, then a test is said to be valid if the test accurately, correctly, authentically, or legitimately measures what it is supposed to measure. (Sudijono, 1991: 93)

The definition of validity according to Sumarna (2005:50) is "A concept related to the extent to which a test has measured what should be measured" according to Mudjijo (1995:40), a test is called valid if the test can measure what it wants and should measure. Furthermore, according to Nana Sudjana (1991:12), validity is the accuracy of the assessment tool for the concept being assessed so that it truly assesses what should be assessed.

From the several definitions above, it can be stated that a test can be said to be valid if the test can measure what it is intended to and should measure.

Meanwhile, what is meant by the validity of a test item is the accuracy of measurement possessed by an item (which is an inseparable part of the test as a whole), in measuring what should be measured by the item. (Sudijono, 1991: 182)

2) Types of Validity

Validity is the most important requirement in an evaluation tool. (Purwanto, 1994: 137) To determine whether a learning outcome test has validity or measuring accuracy, it can be seen from two aspects, namely from the aspect of the test itself as a totality (test validity), and from the aspect of its items, as an inseparable part of the test (test item validity).

(a) Test validity

The validity of a test can be determined from the results of thinking and from the results of experience. The first thing that will be obtained is logical validity (*logical validity*) and the second thing is to obtain empirical validity (*empirical validity*). These two things are the basis for grouping test validity as follows: 1) Logical validity. Logical validity means logical/reasoning, so logical validity for an evaluation instrument refers to the condition for an instrument that meets the requirements for validity based on the results of reasoning and has been well designed, in accordance with applicable theories and provisions. There are two types of logical validity that can be achieved by an instrument, namely content validity and construct validity, 2) Empirical Validity. Empirical validity means having an experience meaning, so an instrument is said to have empirical validity if it has been tested from experience. Thus, empirical validity cannot be obtained only by

compiling an instrument based on provisions such as logical validity, but must be proven by the results of analysis carried out on data from observations in the field, proving that the learning outcome test has been able to accurately measure the learning outcomes that should be measured. (Arikunto, 2002:6)

There are two ways to find out whether the learning outcome test has empirical validity or not, namely in terms of its predictive accuracy. (*predictive validity*) and its comparative accuracy or "present" (*concurrent validity*).

(b) Item validity

The validity of a test item is the accuracy of measuring an item (which is an inseparable part of the test as a totality), in measuring what should be measured through the item. (Sudijono, 1991:182)

Actually, every item in the learning outcome test is an inseparable part of the learning outcome test as a validity in measuring or revealing the learning outcomes that have been achieved by each individual student after they have followed the learning process for a certain period of time.

3) Item Validity Testing Techniques

An item can be said to have high validity or can be declared valid, if the scores on the item in question have a match or direction parallel to the total score or in statistical language: "There is a significant positive correlation between the item score and the total score". The total score here is positioned as a dependent variable (*dependent variable*), while the item score is positioned as the independent variable (*independent variable*). Thus, to arrive at a conclusion about whether the items whose validity we want to know are valid or not, we can use correlation techniques as an analysis technique. A question can be declared valid if the score of the item in question is proven to

have a significant positive correlation with the total score. As is known, in objective tests there are only two possible answers: correct and incorrect. Each item answered correctly is generally given a score of 1 (one), while each incorrect answer is given a score of 0 (zero). This type of data in the world of statistics is known as pure discrete data or dichotomous data. Meanwhile, the total score owned by each item is continuous data.

An item has high validity if the scores on the item align with the total score. This alignment can be interpreted as correlation, so the correlation formula is used to determine item validity. (Arikunto, 2002:162)

According to the existing theory, if variable I is in the form of pure discrete data or dichotomous data (item scores), while variable II is in the form of continuous data (total item scores), then the appropriate correlation technique to use in finding the correlation between variable I and variable II is the Point Biserial Correlation Technique, where the correlation index is given the symbol (r_{pbi}) (Sudijono, 1991: 245)

2) Test Reliability

A reliable test provides a consistent measure of a student's ability to assess performance toward a specific goal. Reliability indicates consistent values. An instrument with a high level of reliability can be relied upon as a basis for drawing conclusions and decisions.

Reliability is a translation of the word *reliability* which has the origin of the word *rely* and *ability*. Reliability is often interpreted as dependability. This means that a test is reliable if the test is used to measure repeatedly and produces the same results. (Thoah, 1996:118)

Reliability is the accuracy or precision of an evaluation tool. A test or evaluation tool is said to be reliable if it is trustworthy, consistent, stable, and productive. (Purwanto, 2002: 139). So, what is important here is the accuracy,

the extent to which the test or tool can be trusted for its accuracy.

A learning outcome test is considered good if it has reliability. When this term is associated with the test's function as a measuring tool for student learning success, the test can be said to be reliable. *reliable* if the measurement results carried out using the test repeatedly on the same subject always show the same results or are fixed and stable. (Sudijono, 1991: 95) the same or stay here does not always have to be the same, but follows changes steadily.

From the several definitions above, the measurement results can be trusted only if the results obtained from several measurements on the same group of subjects are relatively the same, as long as the aspects being measured in the subjects being measured have not changed.

Factors that influence Reliability. Some factors that influence reliability are as follows: (a) The extent of reliable sampling, (b) Differences in the talents and abilities of the students being tested, and (c) The atmosphere and conditions of the testing.

Reliability testing techniques. To determine whether an objective learning outcome test prepared by a tester has achieved consistent measurement, three techniques can be used to test the level of reliability of test items, namely:

(a) Method *Test-retest* (reshape method)

This method is used to test using an assessment tool on the same subject, administered twice at different times, and then correlated. The resulting correlation coefficient indicates the instrument's level of consistency, which also serves as the correlation coefficient value. The results of this technique are reliable if the instrument measures relatively constant variables.

The steps that can be taken in this reliability test are: (1) Compiling a test whose reliability will be measured, (2) Testing the compiled

test (stage I), (3) Calculating the score of the test results from stage I, (4) Retesting the compiled test (stage II), (5) Calculating the score of the retest results (stage II), and (6) Calculating the reliability of the test by correlating the test scores I with test score II with formula *Product Moment Person Correlation*.

(b) Approach *equivalent form reliability* (parallel form)

This approach is used with two forms of the same test administered to one group at the same time. The forms of the test are identical in the sense that they are designed to measure the same abilities.

This method can be used to find out the test stability coefficient with the assumption that the system being measured by the test will not change by only using two forms of test. The steps taken are as follows: (1) Compiling two equivalent tests, (2) Compiling both tests (at the same time or in parallel), (3) Giving a score of the test results that have been tested, arranged by separating test A from test B, and (4) Finding the stability coefficient of both tests (A and B) by looking for the correlation using the formula *Product Moment Correlation*.

(c) Approach *Split-Half*

This method is used to determine reliability by measuring a group of subjects, using only one type of measuring instrument, and only once. In other words, this method is carried out with one group of subjects, one type of measuring instrument, and one measurement, or one group of testees, one type of test, and one testing session. The general steps taken to find test reliability are: (1) Compiling a test with at least an even number of numbers, so that when divided the total is the same, (2) Testing the test on one sample, (3) Calculating the score of each student in two score groups, which

can be grouped into odd and even scores, or can also be grouped into upper and lower half scores, (4) Finding the reliability of half the test by correlating the two scores using the formula *Product Moment* or look for deviations in the odd-even halves, and (5) Look for the reliability of a full test using the formula *Spearman Brown*, formula *Flanagan* and formula *Roll*.

(d) Internal Consistency (*internal consistency*)

To measure the consistency coefficient, an approach can be used that does not divide the test into two. This is caused by two possibilities: (1) the number of items is odd, so it cannot be divided into two, (2) the composition between odd and even items is not homogeneous, so that when divided, it tends not to have a positive correlation.

Internal consistency is based on the homogeneity or correlation between answer scores on each test item. If the average correlation between test items is high, then the reliability is also high. If the average correlation approaches zero, then the internal consistency is also zero and the reliability is low. There are several techniques and equations used to find reliability with this internal consistency, namely L: (1) coefficient *alpha*, (2) Kuder-Richardson₂₀, (3) Kuder-Richardson₂₁, and (4) Hoyt technique.

2. Learning outcomes

The ability of students to perform various expected activities, which they must learn through instructional activities, is called learning. Learning outcomes are the level of mastery achieved by students in participating in a learning program in accordance with established educational objectives. *Taxonomy of Educational Objectives* Bloom classifies learning outcomes into three domains, namely: Cognitive domain, Affective domain, and

Psychomotor domain. (Benjamin, 1971: 271-277).

Briggs states that learning outcomes are all the skills and results achieved through the learning process. Teaching in schools is expressed in numbers or grades measured by learning outcome tests. (Lislie J, 1993: 149)

Likewise, Gronlund stated that learning outcomes are a systematic procedure for determining how much a student has learned. (Norman Gronlund, 1982:1) Nitko further said that learning outcomes are a systematic procedure for obtaining information used in decision-making about students, curriculum and programs and educational policies by observing and describing one or more characteristics using a numerical scale or classification scheme.

Learning outcomes are the abilities that students possess after they have received their learning experiences. The abilities that students possess are the result of learning activities and can be observed through student performance (*learner's performance*). (Dick, 1988:36) The intended student performance refers to the abilities achieved and applied by students in responding to each object they encounter. Dick and Reiser explain that learning outcomes are the abilities students possess as a result of learning activities. (Dick, 1988:11)

Kingsley divides learning outcomes into three types, namely: (1) Skills and habits, (2) Knowledge and understanding, and (3) Attitudes and characteristics. These learning outcomes are obtained by students after participating in learning activities. To determine the level of achievement of student learning outcomes, teachers use learning outcome tests and are usually expressed in the form of scores.

Won *deep* Burhan stated that competence and capability are concrete evidence of learning outcomes and can be divided into five categories: (1) Intellectual Skills (*Intellectual Skill*), (2) Cognitive Strategy (*Cognitive Strategy*), (3) Verbal Information (*Verbal Information*), (4) Motor Skills (*Motor Skill*), and (5) Attitude (*Attitude*). (Nurgianto, 1995:22-24).

While students are studying, they will be faced with problems to solve and overcome

(*Problem Solving*). A problem can be defined as a question that must be solved. Problem solving is something that is integrated within the learner and the learning outcomes.

According to Polya (*deepSetiabudi*) there are four steps that need to be taken in solving a problem, namely: (1) Understanding the problem, (2) Developing a strategy, (3) Carrying out the chosen strategy and (4) Re-evaluating the work that has been done. (Budi, 2003:2)

In order to differentiate between the Learning Outcome Test (THB) and the Learning Achievement Test (TPB), Nasoetion and Suryanto explain as follows: The Learning Outcome Test (THB) is a measuring tool that is able to determine a person's abilities after participating in learning. The material stated is not only about the material obtained from the teacher but also about things outside of what is given, practiced and discussed with the teacher, while the Learning Achievement Test (TPB) is a measuring tool that is able to determine how much of the lesson which has been followed can be mastered/absorbed by students.

It should be emphasized that the assessment of student learning outcomes does not only concern cognitive aspects, but also the application or *performance*, the affective aspect which concerns the attitude of internalizing values that need to be instilled and fostered through the material that has been provided.

From several theories regarding the understanding of learning outcomes above, it can be stated that learning outcomes are a systematic procedure for obtaining information used in decision making expressed in values or numbers based on the results achieved through the learning process. The learning outcomes achieved by students through the learning process are as follows: (1) Increasing confidence in their own abilities in learning, (2) Motivating students intrinsically, (3) Realizing that the learning outcomes achieved are very meaningful for themselves, (4) Their ability to be able to control or assess and control themselves, especially in assessing the results they have obtained, and (5) Learning outcomes are obtained comprehensively (*comprehensive*).

D. CONCLUSION

One of the most effective ways to improve the learning process is by evaluating the learning outcome tests obtained from the learning process itself. In other words, we process the test results in such a way that we can identify which components of the learning process are weak.

Learning Outcomes Tests are said to be good if they have reliability or are reliable. If the term is associated with the function of the test as a measuring tool for student learning success, then a test can be said to be reliable if the measurement results carried out by using the test repeatedly on the same subject always show the same results or are steady and stable.

Learning outcomes are a systematic procedure for obtaining information used in decision making expressed in values or numbers based on the results achieved through the learning process. The learning outcomes achieved by students through the learning process are as follows: (1) Increasing confidence in their own ability to learn, (2) Motivating students intrinsically, (3) Realizing that the learning outcomes achieved are very meaningful for themselves, (4) Their ability to control or assess and control themselves, especially in assessing the results they have obtained, and (5) Learning outcomes are obtained comprehensively (*comprehensive*).

REFERENCES

- Arikunto, Suharsimi. (2002). *Basics of Educational Evaluation*. Bumi Aksara: Jakarta.
- Azwar, Saifuddin. (2000). *Achievement Tests: Functions and Development of Learning Achievement Measurement*. Student Library: Yogyakarta.
- Benjamin, S. Bloom. Thomas Hastings and George F. Madaus. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. McGraw-Hill Book Co: New York.
- Lislie, J. Briggs. (1993). *Instructional Design Principles and Applications*. Englewood Cliffs: New Jersey.

- Budi, Wono Setya. (2003). *First Steps Towards the Math Olympiad*. Ricardo: Jakarta.
- Danim, S. (2010). *Pedagogy, Andragogy, and Heutagogy*. Alfabeta: Bandung.
- Drajat, M. and Effendi, R. (2014). *Professional Ethics of Teachers*. Alfabeta: Bandung.
- Dick, Walter and Robert A. Reiser. (1988). *Planing Effective Instruction*. Allyn and Bacon: Boston.
- Gagne, Robert M. and Leslie J. Briggs. (1974). *Principles of Instructional Design*. Holt Rinehart and Winston: New York.
- Kundandar. (2007). *Professional Teachers Implement the School Level Curriculum (KTSP) and Preparation for Teacher Certification*. Raja Grafindo Persada: Jakarta.
- Majid, Abdul. (2004). *Competency-Based Islamic Religious Education: Concept and Curriculum Implementation*. Rosdakarya Youth: Bandung.
- Ma'mur, A. J. (2011). *PLPG Success Tips*. Diva Press: Yogyakarta.
- Mulyasa. E. (2007). *Becoming a Professional Teacher*. Rosdakarya Youth: Bandung.
- Nurdiyanto, Burhan. (1995). *Applied Statistics: For Social Sciences Research*. Rosdakarya Youth: Yogyakarta.
- Purwanto, Ngalim. (1994). *Principles and Techniques of Teaching Evaluation*. Rosdakarya Youth: Bandung.
- Rusman. (2013). *Learning Models*. Raja Grafindo Persada: Jakarta.
- Saifuddin, Azwar. (2000). *Development of Psychological Scales*. Student Library: Yogyakarta.
- Soetjipto and Kosasi. R. (1999). *Teaching Profession*. Rineka Cipta: Jakarta.
- Sudjana, Nana. (1991). *Assessment of the Results of the Teaching and Learning Process*. Rosdakarya Youth: Bandung.
- Sudijono, Anas. (1991). *Introduction to Educational Statistics*. Eagle: Jakarta.
- Sumarna, Surapranata. (2005). *Analysis, Validity, Reliability, and Interpretation of Test Results: Curriculum Implementation*. Rosda Karya: Bandung.
- Syafruddin, N. and Usman, B. (2002). *Professional Teachers and Curriculum Implementation*. Ciputat Press: Jakarta.
- Trianto. (2010). *Integrated Learning Model, Draft, Strategy and its Implementation in KTSP*. Bumi Aksara: Jakarta.
- Thoha, H. M. Chabib. (1996). *Selected Chapters on Islamic Education*. Student Library: Bandung.
- Usman. U. (2006). *Becoming a Professional Teacher*. Rosda Youth Works: Bandung.