DEVELOPING ASSESSMENT INSTRUMENTS FOR LEARNING OUTCOMES

Ina Magdalena[1], Suci Arwindi[2*], Saka Nayla Hasan[3]
[1-3]Faculty of Teacher Training and Education, Elementary School Teacher Education Study Program, Universitas Muhammadiyah Tangerang
E-mail: [1)] inapgsd@gmail.com, [2)] suciarwindi3030@gmail.com, [3)] sakanayla029@gmail.com

Abstract

This paper aims to provide understanding of assessment instruments for learning outcomes, focusing on benchmark learning outcome tools, norm-referenced and benchmark-referenced tests, validity, reliability, and practicality considerations, various types of tests, and the process of preparing assessment tools. Additionally, it explores performance-based assessment tools, emphasizing the significance of authentic assessment in an educational context. The research adopts a qualitative descriptive paradigm, oriented towards understanding nuances, perspectives, and subjective experiences related to learning outcomes assessment tools. This approach allows exploration of perceptions from teachers, students, and other stakeholders. Assessment is a crucial process in education, involving the collection of data to make informed decisions. This paper underscores the importance of various assessment types and techniques, ranging from daily quizzes to school exams. To ensure accurate and meaningful data, educators must use diverse assessment methods aligned with students' learning processes. The development of assessment tools is guided by determining types, forms, and variations based on specific needs. Effective assessment adheres to established principles, ensuring accuracy and reliability. The paper acknowledges potential shortcomings and errors, recommending the creation of a framework, use of relevant sources, effective collaboration, and time management when undertaking similar writing tasks.

Keywords: *Benchmark Learning Outcomes, Norm-Referenced Tests, Benchmark-Referenced Tests, Validity, Reliability*

1. INTRODUCTION

Assessment is a form of teacher activity related to making decisions about the achievement of competence or learning outcomes of students who follow a particular learning process. For this reason, data is needed as reliable information as a basis for decision making. In this case, the decision relates to whether or not students have succeeded in achieving a competency. So, class assessment is one of the pillars in the implementation of the competency-based Education Unit Level Curriculum (KTSP) (Muhson, 2014).

Assessment of learning outcomes listed in the Regulation of the Minister of Education and Culture (Permendikbud) No. 53 of 2015 Article 1 paragraph 1 states that: Assessment of learning outcomes by educators is the process of collecting information / data about the learning achievements of students in the aspects of attitude, knowledge aspects, and skills aspects carried out in a planned and systematic manner which is carried out to monitor the process, learning progress, and improvement of learning outcomes through assignments and evaluation of learning outcomes (Amrulloh, 2015).

The data obtained by the teacher during the learning process can be netted and collected through assessment procedures and tools that are in accordance with the

competencies or learning outcomes to be assessed. Therefore, classroom assessment is more a process of collecting and using information by teachers to make decisions, in this case the value of students' learning outcomes based on their learning stages.

From this process, a portrait/profile of the ability of students to achieve a number of Competency Standards and basic competencies listed in the curriculum is obtained. Class assessment is a process carried out through planning steps, preparing assessment tools, collecting information through a number of evidence that shows the achievement of student learning outcomes, processing, and using information about student learning outcomes. Classroom assessment is carried out in various ways, such as performance, attitude assessment, written assessment (paper and pencil test), project assessment, product assessment, assessment through a collection of work / work of learners (portfolio), and self-assessment (Sudjana, 2010).

Assessment of learning outcomes, both formal and informal, is conducted in a pleasant atmosphere, allowing learners to demonstrate what they understand and are able to do. A learner's learning results are not recommended to be compared with other learners, but with the learner's previous results. Thus learners do not feel judged by the teacher but are helped to achieve what is expected.

Based on the above background, this paper aims to provide an in-depth understanding of learning outcome assessment tools by focusing on several important aspects. First, the author wants to explain the meaning of benchmark learning outcome assessment tools. Furthermore, the purpose of the writing includes an understanding of norm-referenced tests and benchmark-referenced tests along with their similarities and differences. In addition, the author wants to provide insight into the validity, reliability, and practicality of learning outcome assessment tools, as well as explain the various types of tests that can be used. The process of preparing learning outcomes assessment tools is also a focus, by describing the steps. In addition, this paper aims to provide an understanding of performance learning outcome assessment tools and the importance of authentic assessment in an educational context. Through these objectives, readers are expected to gain a comprehensive understanding of learning outcomes assessment tools.

2. RESEARCH METHODS

This research method uses a qualitative descriptive paradigm. In a qualitative paradigm, this research is likely to be oriented towards understanding the nuances, perspectives and subjective experiences relating to learning outcomes assessment tools. This approach allows researchers to explore various aspects such as the perceptions of teachers, students and other stakeholders towards the assessment tools used.

3. RESULTS AND DISCUSSION

3.1. Definition of Benchmark Learning Outcome Assessment Tool

Benchmark Assessment (PAP) is an assessment approach model that refers to a predetermined goal achievement criteria (TKP). PAP is a way of determining student graduation by using a number of benchmarks. If the student has met the benchmark then it is declared successful. But if the student has not met the benchmark then it is said to have failed or has not mastered the learning material. The values obtained by students are

related to the level of achievement of student mastery of learning material in accordance with predetermined objectives (Helwig et al., 2011).

Any learning outcome assessment tool that is relevant to the ICT is valid to use. If at the end of the instructional process all learners are found to have mastered 100% of the behaviors in the ICT, it can be interpreted that the instructional process has been effective. It is not the instructional designer's place to assume that the learning outcome assessment tool for the ICT is too easy and should be changed or discarded. Such inference is incorrect. The conclusion from the data is that the instructional process was effective, because it succeeded in making all learners who initially did not master the competency, at the end of the instructional process can master it well.

The learning outcome assessment tool itself does not need to be changed, because it has measured what should be measured. Learning outcome assessment tools that refer to instructional objectives are called benchmark assessment tools. Furthermore, the learning outcomes of each learner are compared with the level of competency achievement in the instructional objectives. The term benchmark assessment is a translation of criterion-referenced assessment. Foreign literature refers to it as criterion-referenced test, objective- referenced assessment, content referenced assessment, domain referenced assessment, and universe referenced assessment. These terms are used interchangeably.

Students who have exceeded or are equal to the success criteria or benchmarks are declared to have passed or met the requirements. Teachers do not conduct assessments as they are but based on the success criteria that have been determined since learning began. Teachers who use this PAP approach model are required to always direct, assist and guide students towards minimal mastery since learning begins, is ongoing and until the end of learning. minimum competence.

3.2. Norm-referenced Tests and Benchmark Tests

1. Norm-referenced Tests

Norm-referenced assessment is an assessment carried out with reference to group norms. This means that the scores obtained by students are compared with the scores of other students included in the group. In this case, the norm is the capacity or achievement of the group, while the group is all students who take the test. It can be concluded that the word group means a number of students in one class, school, rayon and province or region. (Pangastuti & MunfaAti, 2018).

Another type of test that will not be used in instructional design, but needs to be known to distinguish it from the benchmark test type is the norm-referenced test. Norm-referenced assessment (PAN) is a classic approach, because the appearance of students' learning achievement on a test is compared with the performance of other students who take the same test. This measurement is used as a measurement method that uses the principle of competitive learning (Suparman, 2012).

Norm-referenced assessment (PAN) is an assessment that compares learners' learning outcomes to those of their group. The purpose of using norm-referenced tests is usually more general and comprehensive and covers a large content area and learning tasks. Norm-referenced tests are intended to determine the status of test takers in relation to others who have taken the test.

Examples of assessments that have been carried out to determine the graduation

(pass or fail) of a student in the UAS (End of Semester Examination) for junior and senior high schools at the end of the school year. From the UAS results, the UAS score is obtained, which comes from the results of the examination committee's assessment using a percentage benchmark, which shows the level of ability or mastery of students about the teaching material tested. In other words, the UAS score is the result of an assessment using the PAP method. However, after the UAS scores, generally very low so that they do not meet the requirements to be declared a pass, then the scores are processed into PAN using certain formulas with the intention that these values can be enlarged.

Formula used: PAN = (p + q + nR)/(2 + n)
Note: p = Odd semester report value
q = Even semester subsumptive average value
R = Final exam value
n = Coefficient of UAS/R coefficient

With the value range of n ranging from 2 to 0.5, so that each region can adjust to its local conditions (R coefficient). If a junior high school student in the Semarang Municipality, where the R coefficient (n) for the Semarang region is 0.75, obtains the values p = 5, q = 8, and the result of the final exam (R) = 4, according to the applicable formula, in Semarang, the student's value becomes:

N = (p + q + nR) / (2 + n)
N = (5 + 8 + (0.75 * 4)) / (2 + 0.75)
N = 16 / 2.75
N = 5.82
The value 5.82 is the one recorded in the report.

2. Reference Benchmark Test

Learning outcome assessment tools that refer to instructional objectives are called benchmark reference assessments. The learning outcomes of each learner are compared with the level of competency achievement in the instructional goal (Suparman, 2012). Benchmark reference assessment (PAP) usually also called criterion evaluation is a measurement that uses different references. In this measurement, students are compared with criteria that have been determined in advance in the learning objectives, not with the performance of other students. Success in the benchmark procedure depends on the material's mastery of the criteria that have been outlined in the question items to support the learning objectives.

With PAP each individual can know what he has and has not mastered. Individualized guidance to improve students' mastery of the subject matter can be designed, as well as to solidify what they have mastered can be developed. Through this benchmark-based assessment we can develop a measuring tool for the success or failure of a learning process by conducting a measuring tool for the success or failure of a learning process by conducting a test at the beginning of learning from the two tests we can find out how much material students can receive in learning activities In the approach with reference to criteria, the determination of levels is based on predetermined scores in the form of percentages.

To get an A or B grade, a learner must get a certain score within the specified limit without being affected by the performance (score) obtained by other learners in the class.

One of the disadvantages of using absolute standards is that learners' scores depend on the difficulty level of the test they receive. For example, like the problem above if we use PAP it will be like this: The first step is to get the criteria, for example as follows:

Table 1. Score range value

Score Range	Value
90 - 100	10
80 - 89	9
70 - 79	8
60 - 69	7
50 - 59	6
40 - 49	5
30 - 39	4
20 - 29	3
10 - 19	2
0 - 9	1

The next step is to convert the scores into grades. For scores: 50 converted to grade 6, 45 converted to grade 5, 40 converted to grade 5, 35: converted to grade 4, 30 converted to grade 4. If compare the above problems, then each grade will have a different meaning.

Table 2. Raw Score Criteria

Raw Score	Values Based on Normal Approximation	Criteria
50	10	6
45	9	5
40	8	5
35	7	4
30	6	4

3.3. Similarities and Differences of Benchmark Tests and Norm-referenced Tests

The similarities between PAN (Norm-referenced Assessment) and PAP (Benchmark Referenced Assessment) are that they both require specific evaluation objectives to determine the focus of the data needed. Both require a sample as a relevant subject when it will be the target of evaluation. Both have special formulas regarding the things to be measured, both are used in the field of education, although with different purposes, The quality of both values is seen in terms of validity and reliability.

The difference between PAN (Norm-referenced Assessment) and PAP (Benchmark Referenced Assessment) is that PAN is used for surveys while PAP is used for assignments. PAN measures specific behavior with few test items, while PAP measures specific behavior with many test items. PAN emphasizes differences among test takers in terms of relative levels of learning achievement, while PAP emphasizes explanations of behaviors that each test taker can and cannot perform.

The differences between Norm-referenced Assessment (PAN) and Benchmark Assessment (PAP) are as follows:

Table 3. Differences between Norm-referenced and Benchmark-referenced Assessments

No	Benchmark Assessment (PAP)	Norm-referenced Assessment (PAN)
1.	PAP tends to emphasize what learners can do, rather than comparing learners with their peers.	PAN is more focused on comparing students' learning outcomes with their peers.
2.	The criteria in this assessment is the level of learning experience that has been achieved by students after learning activities or related to the Basic Competencies that have been determined before starting learning activities.	This assessment is usually used for selection or final learning assessment, so that the teacher can find out the abilities possessed by each student.
3.	The aim is to measure the achievement of goals or competencies that have been set as criteria for success in the learning process.	The aim is to classify students according to the high and low levels of ability they have.
4.	Used to explain the Formative Test results	Used to explain Summative Test results
5.	Measures a large number of competencies with few test items.	Measures a limited number of competencies with many test items.

Normative Benchmark Assessment (PAN) and Benchmark Assessment (PAP): 1. PAP and PAN require specific evaluation objectives to determine the focus of the data needed 2. Both require samples as relevant subjects when they will be targeted for evaluation 3. Both have specific formulas regarding the things to be measured 4. Both are used in the field of education, although with different purposes 5. The quality of their scores is seen in terms of validity and reliability

3.4. Validity, Reliability, and Practicality of Learning Outcome Assessment Tools

Evaluation comes from the word evaluation which means an attempt to determine value or amount. The words contained in the definition also indicate that evaluation activities must be carried out carefully, responsibly, using strategies, and can be accounted for. Evaluation is carried out to provide information about whether the process and results of the activity are good or bad (Suparmin et al., 2013).

1) Validity

The various definitions of validity below may help clarify the notion of validity. "Validity is an evaluation of the adequacy and appropriateness of the interpretations and uses of assessment results" (Miller & Linn, 2000). They explained that validity concerns the adequacy and appropriateness of the interpretation and use of assessment results. Another opinion was expressed by (Wiersma & Jurs, 1990) who stated that "Validity is the term that refers to whether the test measures what it was designed to measure".

Validity is a term that refers to the concept of whether the test measures what it was designed to measure. In educational testing and measurement, the concept of validity is an absolute requirement that a good test must fulfill. A test only has a "price" if it is valid. If a test is verified as valid, then its reliability is proven. In the concept of validity, several notions are known, namely content validity, criterion validity, and construction validity (Suparmin, 2012). Content validity is characterized by two main points as follows.

1. The assessment tool must reflect the taxonomy of areas intended by the

instructional objectives. This statement implies that when learners answer the test, there is an internal process in thinking, moving and or behaving towards the competencies intended in the instructional objectives.

2. On the basis of expert judgment, the learning outcomes assessment tool is declared to be in accordance with the instructional objectives it is intended to measure.

2) Reliability

Reliability is about consistency and assessment results. It includes two notions, namely stability reliability and internal consistency reliability. Stability reliability is usually measured by test-retest, that is, the test whose reliability is determined is used twice on the group of test takers and then the correlation is sought. Internal consistency reliability is determined by being used once on one group of test takers dos then the correlation is sought using the Spearman - Brown formula (sple half), KR-20 or KR-21. When test items are not scored dichotomously, the Cronbach Alpha formula is used.

3)Practicality of Use

In terms of practicality, of course, taking turns in the exam has the potential for question leakage, although it can actually be overcome by different questions, but lecturers will have difficulty because they have to make different questions for a number of students. Many lecturers who give non-test exams teach three classes together. From the point of view of practicality, of course this is not practical if the existing exam takes five to fifteen minutes per student to be impractical (Suparmin, 2012). Revealed by students, the time required for non-written exams varies greatly, ranging from 2 minutes to even 3 months. However, about 11 students took the non-written exam for only 1 hour. This is a non-written exam that follows the academic schedule. When viewed in terms of practicality, 2 minutes per student would certainly be very impractical, especially if the number of students is large with the assumption that the lecturer teaches 3 classes.

Based on the number of classes, from the point of view of practicality, a non-written exam is certainly not practical if the implementation of the existing exam requires 5 minutes per student. Some students hope that the exam should use only written exams, no need for non-written exams and some other students also want practical exams to be held in an appropriate place. If non-written exams are still used, it is hoped that there will be special time, more time than written exams. In addition, students also hope that the questions made are not too difficult so that students can do the questions well.

3.5. Types of Tests

Test comes from the Latin (French) "testum" which means a plate or jar of clay used to select precious metals from other objects such as sand, stone, soil, etc. The term test was later used in the field of psychology. The term test was later used in the field of psychology (Inanna et al., 2021).

There are many ways to categorize test items. The most popular classification scheme for test items has two dimensions. The first dimension corresponds to the assessment method, and the second corresponds to the freedom of response for students. Tests that are objectively scored are called objective test items, while those that require a

more subjective scoring technique are called subjective test items. For example, true-false questions are classified as objective items because the answer is either correct or incorrect according to the pre-existing item. Essay tests are the most common example of subjectively scored items (TenBrink, 1974). In assessing an essay test, the teacher must make subjective judgments about how closely the student's answer aligns with the answer key. A test can be defined as a question or task, or a set of tasks, designed to obtain information about an educational or psychological trait or attribute. Each question or task has an answer or criterion considered correct.

Here are some definitions of a test, (Sax, 1980) suggest that test as a task or series of tasks. Hamid Hasan (1988) test is a specially designed data collection tool that can be seen from the construction of the items used

1. Essay tests are very popular, especially in higher education. Its use began in. 2300 BC in China, and only that form of test has been used for centuries. However, there are other reasons for its popularity. First, because it is easy. In contrast to objective tests, essay tests are relatively easy to prepare, but difficult to check students' answers. Another is the security for the examiner. Essay test writers are rarely asked to defend the 'right' answer or show that no 'wrong' answer is better than the 'right' one. Students are not given readily available answers, but must master a wealth of knowledge, which enables them to develop facts and principles, to organize them coherently and logically, and then to apply those ideas by writing down expressions.

   Answers given to essay test questions can often provide clues about the form and quality of students' thought processes. Many of the traits that essay tests can measure are not clearly defined, such as critical thinking, originality and the ability to organize and combine. The nature of the answers presented, to indicate which students have more/less ability, is rarely explicitly specified. Odel's (1927) scale in (Ebel, 2016) for scoring essay test answers, shows that the length of a student's answer is related to his/her score. The longer the answer, the higher the score. Sample Essay Test: *List the benefits of learning math for everyday life....*
2. Objective tests are tests that are arranged in such a way and have alternative answers provided. Objective tests can be sorted into true-false tests, matching tests, and multiple choice tests. Objective description tests are often used in the fields of science and technology or social fields where the answer is certain and only one answer is correct. Objective tests are tests whose scoring system is objective (Lumbantobing LR., 2021). Objective Test Examples: *Fill in the dots below correctly and precisely! The prime factor of the number 15 is.....*
3. The third type of test after essay tests and objective tests is a performance test, which is a test used to measure the successful implementation of learner tasks When instructional designers plan to assess learner competencies listed in instructional objectives that contain activities using laboratory equipment, operating photocopiers, constructing minimalist house models, and objective tests.

Performance Test Examples:

No	Caturwulan I	Kinerja				
Perilaku		A	B	C	D	E
	Disiplin (kehadiran tepat waktu)		v			
	Partisipasi (ikut serta memberi masukan)			v		
	Argumentasi (rasionalitas)				v	
	Tanggung jawab (misalnya, mengisi rekam medis)		v			
	Kerjasama (tenggang rasa, tolong-menolong, tanggap)			V		
Klinis (dapat dinilai melalui wawancara dan atau presentasi kasus)						
	Ilmu pengetahuan (mempunyai ilmu yang memadai dan mampu menerapkannya, dinilai melalui presentasi kasus dan atau protofolio)		V			
	Keterampilan medis klinis (Keterampilan klinis yang memadai termasuk anamnesis dan pemeriksaan jasmani, dinilai melalui audit medis)		V			
	Kemampuan membuat keputusan klinis ("Clinical reasoning" dinilai melalui presentasi kasus)		V			
	Kemampuan mengatasi kegawatan medis (kemampuan bertindak cepat dan tepat mengatasi kedaruratan sekalgus menyadari keterbatasannya)			v		
	Keterampilan prosedural (kemampuan menyelesaikan tindak medis secara "lege artis", sesuai dengan SOP, dinilai melalui laporan periodik.		V			
Komunikasi						
	Kemampuan berkomunikasi secara efektif (dengan pasien, keluarganya, sejawat, dan staf klinik)			V		
	Kemampuan bekeja dalam tim (kerjasama dengan semua unsur di dalam maupun di luar klinik)			V		
Kepribadian dan profesionalisme						
	Tanggung jawab profesional (kejujuran, keandalan)		V			

Figure 1. Performance Test

3.6. Procedure for developing learning outcome assessment tools

As stated at the beginning of this chapter, the assessment tool to be developed by instructional designers is a benchmark assessment tool because it is intended to measure the level of mastery of each learner of the competencies listed in the instructional objectives. To develop such an assessment tool, instructional designers need to take the following steps:

Step one: Determine the purpose of the assessment The assessment tool that the instructional designer will create will be used for the following two main purposes.

Provide feedback for students about the learning outcomes of students at each stage of the learning process. For this reason, instructional designers need to develop benchmark assessment tools to carefully measure learners' achievement of each

competency in the instructional objectives. This assessment tool functions as a formative test. The results of this test can also be used as clues about learners' difficulties in certain parts of the instructional materials used if followed by a questionnaire or interview.

Assessing the effectiveness of the instructional system as a whole. For this, instructional designers need to develop benchmark assessment tools that can measure students' learning outcomes in mastering all competencies in TIU and a sample of competencies in ICT. This assessment tool will be used as the initial and final tests in the pilot test of the instructional system that has been developed.

The second step is to create a specification table, commonly called a test blue print, for items 1a and 1b above. The simplest test grid consists of four columns: list of competencies, competency weights, percentage of test types, and number of test items. Determine the assessment tool for each competency:

a) Make a test grid b) Determine the test items to be tested c) Write test items in accordance with the grids that have been made; d) Assemble tests e) Write instructions f) Write answer keys; g) Test the technical quality of the test h) Analyze the test results i) Revise the test, and j) Test the validity and reliability of the questions.

3.7. Performance Learning Outcome Assessment Tool

In terms of tools, the assessment of learning outcomes can be divided into two, namely test techniques and non-test techniques. The following is an explanation:

1. Technical Facility

These tests are given orally (demanding answers orally), there are written tests (demanding answers in writing), and there are action tests (demanding answers in the form of actions). Some test questions are arranged in objective form, some are in the form of essays or descriptions.

A test is an official data collection tool because it is full of restrictions. Tests are generally used to assess and measure student learning outcomes, especially cognitive learning outcomes regarding mastery of subject matter in accordance with educational and teaching objectives. However, tests can also be used to assess learning outcomes in the affective and psychomotor fields.

There are two types of tests that will be discussed, namely description tests or essay tests and objective tests. The description test consists of free description, limited description and structured description. While objective tests consist of several forms, namely the form of true-false choices, multiple choices with various variations, matching, and short fillings.

2. Description test (subjective test)

In general, a description test is a question that requires students to answer it in the form of describing, explaining, discussing, comparing, giving reasons, and other similar forms according to the demands of the question using their own words and language. The form of description tests can be divided into three, namely:

a) *free essay*

In the free essay the student's answer is not limited, depending on the student's own view because the question is general. The weakness of this test is that the teacher is difficult to assess because the student's answers vary, it is difficult to determine the

assessment criteria, very subjective because it depends on the teacher as an assessor. Example: Try to mention the benefits of math in everyday life.

b) Limited description

In this form the question has been directed to certain things or there are certain restrictions. The question is more specific to a particular object. Example: Tina will put 21 red marbles and 28 white marbles into a box. Each box contains the same number of red marbles and the same number of white marbles. How many boxes are needed? How many red marbles and white marbles are in each box. The advantages of a descriptive test: a) Easy to prepare and compile b) Does not provide much opportunity for speculation or conjecture c) Encourages students to express their opinions and arrange them in good sentences d) Provides opportunities for students to express their intentions in their own language and style e) Can be known to what extent students explore a problem that is tested.

Meawhile, the weaknesses of the description test: a) The level of validity and reliability is low because it is difficult to know which aspects of the student's knowledge are truly mastered. b) Less representative of the entire subject matter because the questions are only a few. c) How to check is much influenced by subjective elements. d) The examination is more difficult because it requires more individual consideration from the assessor. e) The time for correction is longer and cannot be represented by others.

3. Objective test

Is a test in which the examination can be done objectively. In the use of objective tests, the number of questions submitted is much greater than essay tests. Various types of objective tests:

a) True-false test, one form of test where there is true and false. Example: (B)- (S) Rasullulah was born in 571 AH coinciding with the year of the Elephant.
b) Multiple choice tests are objective tests where each test provides more than the possibility of several answers and only one correct choice. Example; The change from solid to liquid is called: a) Evaporating b) Melting c) Freezing d) Condensing.
c) Matching test, this form of test is matching, looking for views. So that the question has the correct answer

The advantages of objective tests: a) More representative of the teaching material because there are more questions b) It is easier and faster to read because the answers are provided, just choose c) The examination can be handed over to someone else d) In the examination, there is no subjective element that affects. Weaknesses of objective tests: a) Preparation to compile it is much more difficult than essay tests b) The questions tend to reveal memory and recognition only, and it is difficult to measure high mental processes c) Many opportunities to play luck d) Cooperation between students when working on test questions is more open.

4. Non-test technique

Learning outcomes and processes. Teachers in schools generally use more tests than non-tests because the tools are easy to make, their use is more practical and what is assessed is limited to cognitive aspects based on the results obtained by students after

completing their learning experience. The following is an explanation of non-test or non-test tools:

a) Interview: Interview is a method or method used to obtain answers from respondents by means of one-sided questions and answers. Interviews can be conducted in two ways, namely free interviews and guided interviews. Example: What are your goals?

b) Questionnaire: The questionnaire is often called a questionnaire. A questionnaire is a list of questions that must be filled in by the person to be measured (respondent). The questionnaire can be viewed from several aspects: In terms of who answers, there are: a) Direct Questionnaire b) Indirect Questionnaire In terms of how to answer, it can be divided into: a) Closed Questionnaire b) Open Questionnaire.

c) Scale: A scale is a tool for measuring values, attitudes, interests, and concerns arranged in the form of statements to be assessed by respondents and the results are in the form of a range of values according to specified criteria. The scale is divided into two, namely: 1) Rating Scale: Rating scales measure the performance or behavior of others by an individual through a statement of the individual's behavior at a point on a continuum or a value-meaningful category; and 2) Attitude Scale: An attitude scale is used to measure a person's attitude towards a particular object. The results are in the form of attitude categories, namely supporting (positive), rejecting (negative), and neutral.

d) Cheklist: A checklist is a series of (usually short) statements where the respondent being evaluated simply marks the appropriate place.

e) Observation: Observation is a technique carried out by making careful observations and recording systematically. There are 3 types of observation, namely: Direct Observation: Direct observation at school can be done by using a special notebook about events related to students during school. Observation with Tools (Indirect): Observation can be done indirectly in other ways, such as reading the news. In this case, observation can be done indirectly by getting news from various sources, such as television, radio, and through the internet. Participant Observation: A method in which the researcher is involved in the daily activities of the person being observed or used as a source of research data. With this participant observation, the data obtained will be sharper and get to know at the level of meaning of every behavior that appears.

5. Sociometry

Sociometry is to determine students' ability to adjust themselves, especially students' social relationships with their classmates. Sociometry can be done by assigning all students in the class to choose one or two of their closest or most familiar friends. Try to make sure that no student tries to compromise to choose each other so that the choice is neutral, not pre-arranged. Write the name of the choice on a small piece of paper, then roll it up and collect it from the teacher. After all the choices are collected, the teacher processes them in two ways. The first way is to draw the flow of choices from each student in the form of a sociogram so that the relationship between students based on their choices can be seen. The second way is to score the students' choices.

3.8. Authentic Assessment

1. Authentic Assessment

According to Pokey & Siders in Santrock, authentic assessment is the process of assessing students, especially the competencies that students have acquired or evaluating students' knowledge or skills in a context that is as close to the real world or real life as possible. While Mueller argues authentic assessment is "a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills." So, authentic assessment is a form of task that requires learners to demonstrate meaningful real-world performance that is an application of essential knowledge and skills (Nisrokha, 2018).

Meanwhile, according to (Nurgiyantoro, 2001) authentic assessment emphasizes the ability of learners to demonstrate their knowledge in a real and meaningful way. Assessment activities do not simply ask or tap the knowledge that the learner already knows, but rather perform significantly from the knowledge and skills that have been mastered.

So it can be concluded that Authentic Assessment is an assessment of learning outcomes that refer to meaningful "real world" situations or contexts that are the application of the essence of knowledge and skills that require a variety of approaches to solving problems that provide the possibility that one problem can have more than one type of solution.

In other words, authentic assessment monitors and measures students' abilities in a variety of possible problem solving encountered in real-world situations or contexts. In a learning process, authentic assessment measures, monitors and assesses all aspects of learning outcomes (which are covered in the cognitive, affective and psychomotor domains), both those that appear as the end result of a learning process, as well as in the form of changes and development of activities, and learning gains during the learning process in the classroom and outside the classroom.

2. Characteristics of Authentic Assessment

A well-conducted assessment of learning outcomes will provide useful information in improving the quality of the teaching and learning process that has been carried out. The characteristics of authentic assessment are as follows:

Nurhadi suggests that the characteristics of authentic assessment are as follows: 1. Involves real-world experience (involves real-world experience.) 2. Implemented during and after the learning process takes place. 3. Includes self-assessment and reflection that measures skills and performance, not remembering facts. 4. Continuous. 5. Integrated. 6. Can be used as feedback. 7. The criteria for success and failure are clearly known to students.

Meanwhile, according to (Kunandar, 2011) the characteristics of authentic assessment are: a. Must measure all aspects of learning, namely performance and results or products. b. Carried out during and after the learning process takes place. c. Using various methods and sources. d. The tasks given to students reflect parts of students' lives that are real every day, they must be able to tell the experiences or activities they do every day. e. Assessment must emphasize the depth of knowledge and expertise of students, not the breadth (quantity).

3. Principles of Authentic Assessment

At the primary and secondary education levels in the 2013 Curriculum, the assessment of students is based on the following principles: a. Objective, meaning that the assessment is based on standards and is not influenced by the subjectivity of the assessor. b. Integrated, meaning that the assessment by educators is carried out in a planned manner, integrated with learning activities, and continuous. c. Economical, meaning that the assessment is efficient and effective in planning, implementing, and reporting. d. Transparent, meaning that the assessment procedure, assessment criteria, and the basis for decision making are accessible to all parties. e. Accountable, meaning that the assessment can be justified. Economical, meaning that assessment is efficient and effective in planning, implementing, and reporting. d. Transparent, meaning that assessment procedures, assessment criteria, and the basis for decision making are accessible to all parties. e. Accountable, meaning that assessment can be accounted for to internal and external school parties for aspects of techniques, procedures, and results. f. Educative, meaning that it educates and motivates learners and teachers.

Above it has been mentioned that the 2013 curriculum assessment standards prioritize the principles of honesty and aspects of knowledge, skills, and attitude. The assessment techniques and instruments in the 2013 curriculum are as follows:

a. Assessment of Attitude Competencies Educators assess attitude competencies through observation, self-assessment, peer review by students and journals. The instrument used for observation, self-assessment, and peer assessment is a checklist or rating scale accompanied by a rubric, while the journal is in the form of educator notes.
b. Knowledge Competency Assessment Assesses knowledge competencies through written tests, oral tests and assignments.
c. Skills Competency Assessment Teachers assess skills competencies through performance assessment, which is an assessment that requires students to demonstrate certain competencies using practical tests, project tests, and portfolio assessments. The instrument used is a checklist or rating scale equipped with a rubric.

Based on the explanation above, it can be concluded that the instrument must meet the requirements for presenting the competencies being assessed, the structure of the assessment is in accordance with the applicable requirements and the language must be communicative and easy for students to understand.

4. Types of Authentic assessment

1) Performance assessment

Authentic assessment involves learner participation as much as possible, especially in the process and aspects to be assessed. Teachers can do this by asking learners to name the elements of the project or task that they will use to determine the criteria for completion. The example is as follows :

INSTRUMEN PENILAIAN KINERJA PRAKTIKUM

Percobaan : Pengukuran Dasar

Nama praktikan :

NIM :

Kelas :

No	Aspek Keterampilan yang diamati	Skor				Nilai
		1	2	3	4	
1	Siswa mengecek kesesuaian alat dan bahan yang disiapkan dengan panduan percobaan					
2	Siswa melakukan percobaan dengan prosedur yang benar					
3	Siswa mencatat dan mengorganisasi data percobaan dengan tepat dan rapi					
4	Siswa menganalisis data percobaan yang di peroleh dan membuat laporan sederhana hasil percobaan					
	Skor Total					

Berikan tanda check list (√) di dalam kolom skor dengan ketentuan:
4 = sangat tepat = Nilai 100
3 = tepat = Nilai 75
2 = tidak tepat = Nilai 50
1 = sangat tidak tepat = Nilai 25

Figure 2. Performance assessment

2) Project assessment

Project assessment is an assessment of tasks that must be completed by learners according to a certain period/time. While working on a learning project, learners get the opportunity to apply their attitudes, skills and knowledge. The example is as follows :

Project Assessment Format

Table 4. Project Assessment

Subjects	:
Grade/Semester	:
Time Allocation	:
Number of Questions	:
Standard Compentence	:
Basic Competition	:

No	Assessed Aspect	Score
1.		
2.		

3) Portfolio Assessment

Portfolio assessment departs from the work of learners individually or produced in groups, requires learner reflection, and is evaluated based on several dimensions. Portfolio assessment is a continuous assessment based on a collection of information that shows the development of learners' abilities in a certain period. The information can be in the form of learners' work from the learning process that is considered the best, test results (not grades), or other information relevant to the attitudes, skills, and knowledge required by a particular topic or subject. The example is as follows :

Table 5. Portfolio assessment

No	Bill Type/Date	Score	Description	KDI/Assessed Indicator
1.				
2.				

4) Written Assessment

Written tests consist of choosing or supplying answers and descriptions. Selecting answers consists of multiple choice, true-false, yes-no, matching, and cause-and-effect. Supplying answers consists of filling in or completing short or short answers, and descriptions. The example is as follows :

Table 6. Written Assessment

Subject: Islamic Religious Education Class: IV (four)
Competency Standard: Carry out Basic Survey Practices Basic Competencies: Performing / Practicing Prayer Readings

No	Rated Aspect	Score
1.	Students can practice prayer movements.	
2.	Students are able to recite prayer prayers	
3.	Students can mention the various kinds of prayer movements.	
	Average Score	

Written Test Score:
Information :

1. Scores are scored from 40-100
2. The lowest score is 40 and the highest is 100.

4. CONCLUSION

Assessment is the process of collecting data from what has been done to create an assessment result that will later become a decision. In the assessment tool, there are types of assessments, and these types of assessments become the information/data for determining the scores of students. Assessments can range from daily quizzes to school

exams. With these assessments, it is hoped that educators will provide assessment results that align with what students have processed. To obtain accurate, authentic, and meaningful data, educators can use various techniques and forms of assessment to achieve accurate results. Good assessment involves applying established rules and principles. The development of assessment tools can be done by determining the types, forms, and variations to be used according to needs. Once determined, the assessment instrument can be created. Effective and accurate assessment should be based on assessment principles.

In writing this paper, there may still be shortcomings and errors. Therefore, the following suggestions can be conveyed to readers when they intend to create a written work: create a framework for the writing to be poured into the paper, seek relevant and accurate sources and references, collaborate effectively with group members and use time wisely.

REFERENCES

Amrulloh, M. A. (2015). Sistem Penilaian dalam Pembelajaran. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, *7*(2), 125–148.

Ebel, A. (2016). *Two essays on determinants of firm-level employee job satisfaction: Organization capital and within firm pay gap*. Rensselaer Polytechnic Institute, Troy, NY.

Helwig, N. E., Hong, S., Hsiao-Wecksler, E. T., & Polk, J. D. (2011). Methods to temporally align gait cycle data. *Journal of Biomechanics*, *44*(3), 561–566.

Inanna, I., Rahmatullah, R., & Hasan, M. (2021). *Evaluasi Pembelajaran: Teori dan Praktek*. Tahta Media Group.

Kunandar. (2011). Evaluasi program pengembangan dan implementasi kurikulum tingkat satuan pendidikan (KTSP). *Jurnal Evaluasi Pendidikan*, *2*(2), 171–181.

Lumbantobing LR. (2021). *Kelebihan dan Keterbatasan dari Bentuk-Bentuk Penilaian Tes Tertulis*. Pusdiklat Keuangan Umum. https://klc2.kemenkeu.go.id/kms/knowledge/kelebihan-keterbatasan-penilaian-tes-tertulis-0f81178c/detail/

Miller, D. M., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement*, *24*(4), 367–378.

Muhson, A. (2014). *Penyusunan Alat Penilaian Hasil Belajar Dan Analisis Butir Soalnya*. http://staffnew.uny.ac.id/upload/132232818/pengabdian/ali-muhson-2012-penyusunan-alat-penilaian-hasil-belajar-sma-1-bantul.pdf

Nisrokha, N. (2018). Penilaian Otentik. *Madaniyah*, *8*(2), 209–229.

Nurgiyantoro, B. (2001). *Penilaian dalam pengajaran bahasa dan sastra*. BPFE-Yogyakarta.

Pangastuti, R., & MunfaAti, K. (2018). Penilaian Acuan Norma, Penilaian Acuan Patokan, Kriteria Ketuntasan Minimal Di Madrasah Ibtidaiah an-Nur Plus Junwangi Krian Sidorajo Jawa Timur. *Tarbiyah Al-Awlad: Jurnal Kependidikan Islam Tingkat Dasar*, *8*(2), 202–217.

Sax, G. (1980). *Principles of Educational and Psychological Measurement and Evaluation. Belmont, California: Wadsworth*. Inc.

Sudjana, N. (2010). *Penilaian hasil proses belajar mengajar*. Remaja Rosdakarya.

Suparman, M. A. (2012). Desain Instruksional Modern: Panduan Para Pengajar dan Inovator Pendidikan, Jakarta: PT. *Gelora Aksara Pratama*.
Suparmin, U. A. (2012). Validitas, reliabilitas, dan kepraktisan ujian melalui observasi dan bentuk lisan bagi kelas besar pada prodi PAI Fakultas Tarbiyah dan Bahasa Iain Surakarta. *Kodifikasia*, *6*(1), 1–12.
TenBrink, T. D. (1974). Evaluation: a practical guide for teachers. *(No Title)*.
Wiersma, W., & Jurs, S. G. (1990). Educational measurement and testing. *(No Title)*.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).