

The Implementation of Criterion Referenced Test Approach in Analysis of Student Learning Outcomes at Madrasah Ibtidaiyah

M Akmal Binashril Hadi¹, Prayoga Saputra^{2*}, Anisul Imamah³ Imas Maisarah⁴

^{1,2,3,4} Pendidikan Bahasa Arab, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

*Email: saputraprayoga33@gmail.com

ABSTRACT

Criterion Referenced Test (PAP) establishes specific standards that students must meet, providing objective evaluations and helping teachers identify which students need remedial support. PAP ensures that all students achieve the required basic competencies. This research describes the application of PAP in analysing Arabic language scores at the Madrasah Ibtidaiyah level. Utilizing a qualitative approach with descriptive research, data was collected through unstructured interviews with the school principal and Arabic teacher, as well as observations of student grades. Data analysis involved organizing data by research objectives and interpreting findings accordingly. The study found that PAP facilitates teachers in conducting grade corrections or remedial actions when students fail to meet the set criteria, which is a minimum of 33% or a grade C. The student's raw Arabic scores are converted to a letter scale, revealing cases where students did not meet the minimum score, necessitating remediation. At MI Miftahul Huda, PAP is used with a letter scale conversion to evaluate grade 4 students' Arabic language performance, with a minimum criterion set at 55% (grade C). This approach simplifies the implementation of remedial measures for students not achieving passing grades and proves effective in tracking student progress.

Keywords: *Criterion Referenced Test (PAP), Scale Four, Remedial*

ABSTRAK

Criterion Referenced Test (PAP) menetapkan standar khusus yang harus dipenuhi siswa, memberikan evaluasi objektif dan membantu guru mengidentifikasi siswa mana yang membutuhkan dukungan perbaikan. PAP memastikan bahwa semua siswa mencapai kompetensi dasar yang dibutuhkan. Penelitian ini menjelaskan penerapan PAP dalam menganalisis skor bahasa Arab pada jenjang Madrasah Ibtidaiyah. Memanfaatkan pendekatan kualitatif dengan penelitian deskriptif, data dikumpulkan melalui wawancara tidak terstruktur dengan kepala sekolah dan guru bahasa Arab, serta pengamatan nilai siswa. Analisis data melibatkan pengorganisasian data berdasarkan tujuan penelitian dan menafsirkan temuan yang sesuai. Studi tersebut menemukan bahwa PAP memfasilitasi guru dalam melakukan koreksi nilai atau tindakan perbaikan ketika siswa gagal memenuhi kriteria yang ditetapkan, yaitu minimal 33% atau nilai C. Skor bahasa Arab mentah siswa dikonversi ke skala huruf, mengungkapkan kasus di mana siswa tidak memenuhi skor minimum, yang memerlukan perbaikan. Di MI Miftahul Huda, PAP digunakan dengan konversi skala huruf untuk mengevaluasi kinerja bahasa Arab siswa kelas 4, dengan kriteria minimal ditetapkan sebesar 55% (kelas C). Pendekatan ini menyederhanakan penerapan langkah-langkah perbaikan bagi siswa yang tidak mencapai nilai kelulusan dan terbukti efektif dalam melacak kemajuan siswa.

Kata-kata Kunci: *Penilaian Acuan Patokan, Skala Empat, Remedial*

INTRODUCTION

Recently, the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia (Kemendikbud Ristek) created new learning and assessment process policy standards, which are expected to be able to monitor the continuous improvement in the quality of education. This regulation is not a new regulation but a simplified form of Permendikbudristek regulation no. 62 of 2016, which was then simplified and became Permendikbudristek regulation no. 53 of 2023. The new learning process standards include three things, namely; a) Planning the learning process. b) Implementation of the learning process. c) Assessment of the learning process. Meanwhile, the learning assessment process is divided into two ways, namely; formative assessments carried out during the learning process, with the aim of improving the learning process. Then, summative assessment, this assessment is carried out at the end of the learning process by referring to Graduate Learning Achievements (CPL) (Efrianti & Zakir, 2024)

Evaluation is the final step that can describe the results of learning activities (Devi dkk., 2022) Through evaluation, obstacles, constraints and achievements experienced during learning can be identified (Sholiha & Rizal, 2023). This evaluation is carried out at various levels of education, with the aim of assessing and knowing the extent to which learning objectives have been achieved. The results of this evaluation can be represented in the form of an assessment. Apart from that, the choice of assessment technique also plays an important role in explaining the role of evaluation. The assessment technique has two approaches, namely normal reference assessment and benchmark assessment. (Yulianto, 2024).

Norm Referenced Evaluation can be called a relative standard or group norm, namely an assessment carried out by interpreting the test results obtained by students by comparing them with the test results of other students in their group (Intiana dkk., 2023). This comparison is determined based on the scores achieved by students in that group. Therefore, in contrast to PAP, passing criteria can only be determined after students receive their grades. According to Zaenal Arifin, PAN is also defined as an assessment that compares student learning outcomes with the learning outcomes of other students in their group. PAN is comparing the scores obtained by students with relative standards or norms (Arifin, 2009). Because if a student who falls into group A is considered "Great, maybe if they move to another group they will only be of "Mediocre" quality. (Arikunto, 2010). This assessment approach can be said to be an "as is" approach in the sense that the benchmarks are only facts obtained during measurement/evaluation, namely student learning outcomes that are measured, PAN has nothing to do with benchmarks other than the results measured by a group of students. The score obtained from a new test will be meaningful if it is interpreted based on a benchmark or based on a norm. This is what is called assessment. Processing grades into a student's final grade can be done by referring to certain criteria or benchmarks.

Based on the definition above, we can conclude that normative reference assessment (PAN) includes assessments carried out by referring to the norms contained in the group; The grades achieved by a student are compared with other grades in the group. This measurement can be called a competitive measurement method in learning principles (Sukardi, 2008). Meanwhile, assessments that refer to criteria or are called absolute standards are assessments of the level of achievement of learning outcomes based on the scores achieved by students compared to the ideal (maximum) score. In other words, PAP is an assessment carried out by interpreting the test results obtained by students by comparing them with predetermined standards. The criteria used to determine the graduation limit are determined before the test results are available, even before teaching and learning activities are carried out (Intiana dkk., 2023). In this paper, value processing techniques are presented in a simple structure which includes understanding, characteristics, objectives and benefits, usage procedures, advantages and disadvantages, as well as the differences between the two approaches. There is one prominent figure who does not agree with determining standards or criteria in assessment that can be drawn from assessment standards in learning, namely Einstein (Sinulingga, 2022), he believes that a person's abilities cannot be determined by criteria set in a group or institution, because such things are usually done based on abilities or standards from a certain point of view, so that not everyone can reach these standards because everyone has criteria and standards of ability. certain in him.

According to Alfath, these two techniques are commonly used to process raw scores of student learning outcomes into meaningful values (Alfath, 2019). It can be seen that the score is the sum of each question answered by the testee in the form of a number. Meanwhile, the value is the result of the accumulation of various scores in the form of numbers or letters which have been adjusted to certain standard settings (Khalim, 2019). Based on Sukiman's statement, it can be concluded that the score will be meaningful when interpreted through benchmarks or norms. Meanwhile, processing scores into grades can be done by comparing the scores obtained by a student with other students or by comparing them with

established standards. These two references have different assumptions, resulting in different information.

Madrasah Ibtidaiyah, hereinafter abbreviated as MI, is one level of formal education unit under the auspices of the Ministry of Religion. MI is a formal education unit that provides general education with Islamic religious characteristics consisting of 6 levels (Manalu dkk., 2020). MI is at the same level as elementary school (SD) education which is a continuation of Raudhatul Athfal. At this level students are introduced to Arabic and it is one of the compulsory subjects. As Hanani said, Arabic in Indonesia is studied or taught from elementary to university level, especially Islamic universities. This means that Arabic language learning can be found at various levels of educational units in Indonesia, both formal and non-formal. In practice, learning Arabic is the same as other subjects, which go through a learning and evaluation process (Rahman & Maisurah, 2023).

MI Miftahul Huda is a madrasah located in Kediri district. This madrasah conducts learning evaluations twice a year, namely formative and summative. This is done to measure compliance with standards which aim to monitor students' learning processes and progress and to increase learning effectiveness. In this evaluation activity, MI Miftahul Huda made previous preparations, in the form of a teacher meeting to determine the test model and scores for each test item. Then the next stage is the evaluation or test implementation stage. Then the analysis or assessment stage. At this stage the teacher analyzes the student test results in the form of raw scores to be interpreted into grades using a benchmark assessment approach. The aim of this research is to describe PAP and its application in analyzing students' Arabic language scores at MI Miftahul Huda Kediri.

Research on the use of this benchmark assessment approach has been widely researched, including research conducted by Khairuddin Alfath et al on techniques for processing assessment results using PAN and PAP, from the research results obtained he explained that measuring the level of student mastery through PAP refers to the objectives instructional (Alfath, 2019), while PAN aims to determine the position of students in their group. Then, research conducted by Ina Magdalen et al regarding the analysis of PAP at Sukasari 3 Elementary School, the results of the research were in the form of assessments for students in the middle of the odd semester in all subjects, with the highest average being 92 and the lowest average being 61. (Magdalena dkk., 2020, hlm. 3). Furthermore, research conducted by Ratna Pangastuti et al regarding PAP, PAN, KKM at MI An-nur plus Junwangi Krian Sidoarjo, East Java (Rahman & Maisurah, 2023). The results of this research state that PAP refers to the criteria for achieving learning objectives or learning indicators that were previously formulated, while PAN refers to group norms, namely the score obtained by a student compared with the scores of other students who are included in that group. These studies are different from research conducted by researchers at MI Miftahul Huda Kediri. MI Miftahul Huda Kediri has a minimum standard that students must achieve in learning Arabic, namely 55% with a grade of 33 or the equivalent of C, so that students who do not achieve this minimum grade must take remedial measures to improve their grades.

This research aims to describe the use of a benchmark assessment approach in analyzing student learning outcomes at the Madrasah Ibtidaiyah level using a letter scale carried out at MI Miftahul Huda Kediri. The use of the benchmark assessment approach at MI Miftahulk Huda has undergone changes before, so the researcher intends to find out and describe the use of this benchmark assessment approach through the following question: how is the benchmark assessment approach used in analyzing student learning outcomes in Arabic language subjects at MI Miftahul Huda? So why does MI Miftahul Huda use a benchmark assessment approach? This benchmark assessment approach is easier to use in measuring students' learning achievement levels, so it is important to carry out research.

Researchers assume that the use of the benchmark assessment approach with letter scale conversion can only be used in universities, because in general it is very common for universities to use the approach. The benchmark assessment is especially true when converted using a letter scale, so the benchmark assessment approach using a letter scale is very suitable for determining a student's GPA.

RESEARCH METHODS

This research uses a qualitative approach with descriptive research type. Descriptive research is research that is used to explain a situation without manipulating (Sugiyono, 2010). The research object was carried out at MI Miftahul Huda Kediri, Indonesia, 2023. The researcher chose this object because this Madrasah is one of the A-accredited private madrasahs in Kediri and has made several changes to the assessment approach.

The data sources for this research are the school principal and Arabic language subject teacher at MI Miftahul Huda. The school principal is the party who handles and directs and monitors the most accurate learning activities, while the subject teachers are the implementers of the learning process, so researchers hope that the data collected will be more accurate. This research was carried out using a data collection process in the form of observation. The use of observation techniques is carried out in a naturalistic manner without any manipulation of sources or field conditions. The second technique uses unstructured interviews, namely an interview guide that only contains an outline (Arikunto, 2010). The use of unstructured interviews is due to deeper curiosity and ease of accessing the desired information according to research objectives. This process is carried out by researchers with informants who are considered to know the details or key informants of learning evaluation activities. Apart from interviews, documentation is also used in the form of student grades in Arabic subjects.

Meanwhile, the data analysis used uses Miles and Huberman's theory, which includes data collection, in the form of interviews and documentation, data organization, namely grouping data based on research objectives, as well as interpretation of the results, namely in the form of meaning of data, this meaning is in the form of raw data that researchers obtain in interviews and documentation, then processed and interpreted for meaning so that appropriate research findings are obtained. In the final stage, the researcher presents the data that has undergone the reduction process in narrative form (Rahmah dkk., 2023)

RESULTS AND DISCUSSION

Benchmark Assessment at MI Miftahul Huda

MI Miftahul Huda is a madrasa that has adopted a systematic approach to student assessment. This institution utilizes the LAP (Laporan Analisis Penilaian) approach, a framework that aligns closely with the Criterion Referenced Test (PAP) method. The LAP approach is particularly advantageous in that it precisely measures the extent to which students achieve the set goals or competencies defined as success criteria. This precision is crucial for several reasons.

Firstly, the LAP approach facilitates the development of targeted remedial programs. By clearly identifying which specific competencies students have not yet mastered, educators can design interventions that address these gaps effectively. This targeted approach ensures that remedial efforts are not only relevant but also efficient, thereby optimizing the learning process.

Secondly, the LAP approach provides valuable feedback on the achievement of learning objectives. It allows educators to evaluate whether students are meeting the predefined benchmarks and to assess the overall effectiveness of the teaching strategies employed. This feedback loop is essential for continuous improvement in both teaching practices and student performance.

As noted by Suryadi, one of the key benefits of benchmark assessments like PAP is their ability to aid teachers in crafting appropriate remedial programs. Benchmark assessments provide clear standards against which student performance can be measured, thus enabling educators to determine the exact areas where students need additional support. Furthermore, these assessments offer a reliable measure of learning outcomes, helping educators gauge whether educational goals are being met and making adjustments as necessary. (Suryadi, 2019).

The Criterion Referenced Test (PAP) is employed as a pivotal tool in educational settings to measure adherence to established standards. Its primary functions include monitoring student learning processes and progress, as well as enhancing the effectiveness of learning outcomes. This assessment approach is integral to ensuring that educational objectives are met and that students achieve the competencies required by the curriculum.

At MI Miftahul Huda, the implementation of PAP is guided by a well-defined assessment policy. This policy delineates two key types of assessment: educator-based assessments and educational unit assessments. Educator-based assessments are conducted by teachers to evaluate individual student performance, while educational unit assessments involve broader evaluations conducted at the level of the entire educational institution. Both types of assessments are crucial in providing a comprehensive overview of student learning and development.

The madrasah employs various forms of assessments to gauge student performance effectively. These include oral assessments, written tests, assignments, and project evaluations. Each type of assessment serves a distinct purpose and provides different insights into student understanding and capabilities. Oral assessments typically measure students' verbal communication skills and immediate recall of information, while written tests assess their ability to apply knowledge in a structured format. Assignments and projects,

on the other hand, evaluate students' ability to synthesize and apply their learning to practical or creative tasks.

MI Miftahul Huda adheres to specific criteria for assessing student performance, encompassing a range of evaluative dimensions. These criteria include assessments of knowledge, attitudes, and social skills. Knowledge assessments focus on evaluating students' understanding of the subject matter and their ability to recall and apply information. Attitude assessments gauge students' behavioral dispositions, such as their motivation, engagement, and responsiveness to learning activities. Social assessments, meanwhile, evaluate students' interactions with peers, teachers, and their overall integration into the learning community.

Marta further elaborates on the assessment framework by highlighting that it encompasses three critical aspects: knowledge assessment, attitude assessment, and social assessment. Knowledge assessment is concerned with the cognitive domain, measuring students' grasp of academic content. Attitude assessment addresses the affective domain, evaluating students' emotional responses and attitudes towards learning. Social assessment pertains to the behavioral domain, focusing on students' social interactions and their contribution to the classroom environment (Safitri & Harjono, 2021).

MI Miftahul Huda has periodically updated its assessment types to ensure alignment with the evolving needs of students, changes in the educational environment, and shifts in the curriculum. This dynamic approach reflects the institution's commitment to maintaining relevant and effective evaluation methods that address both current educational standards and the specific context of its students. To systematically assess student learning outcomes, MI Miftahul Huda follows a structured evaluation process. This process consists of several distinct stages: First, preparation stage. This initial stage involves designing the assessment framework and planning the various assessment activities. It includes setting clear objectives for what the assessments aim to measure, determining the appropriate assessment tools and methods, and ensuring that all logistical aspects are in place. This stage also involves preparing any necessary materials and informing students and educators about the upcoming assessments. Proper preparation is crucial for ensuring that the assessment process runs smoothly and meets its intended goals.

Second, implementation stage: During this phase, the planned assessments are carried out according to the schedule. This stage involves administering oral, written, assignment, and project assessments as outlined in the preparation phase. It also includes monitoring the administration of assessments to ensure that they are conducted fairly and consistently. Implementation must be executed with attention to detail to maintain the integrity of the assessment process and to ensure accurate measurement of student performance.

Third, analysis/processing stage. After the assessments are completed, the next step involves analyzing and processing the results. This stage includes grading and evaluating student work, compiling assessment data, and interpreting the findings. It is essential to use appropriate methods for analyzing the data to ensure that the results accurately reflect student performance and learning outcomes. This stage also includes identifying any patterns or trends in the data that may indicate areas where students may need additional support or where the curriculum may need adjustment.

Fourth, follow-up stage. Based on the analysis of the assessment results, follow-up actions are taken to address any identified issues. This may involve providing additional support or remedial programs for students who did not meet the expected standards. It also includes making any necessary adjustments to the curriculum or teaching strategies based on the insights gained from the assessment data. Effective follow-up ensures that the assessment results are used constructively to enhance the learning experience and address any gaps in student performance.

Fifth, reporting stage. The final stage involves reporting the results of the assessments to relevant stakeholders. This includes providing feedback to students, informing parents or guardians about student progress, and sharing summary reports with educational authorities or other relevant parties. Reporting should be clear, transparent, and provide actionable insights based on the assessment results. It is crucial for ensuring that all stakeholders are informed about student performance and the outcomes of the assessment process.

Moli elaborates on the evaluation process by emphasizing that it encompasses several key components: planning, implementation, monitoring, data processing, data reporting, and the utilization of evaluation achievements. Planning involves setting objectives and designing the assessment framework, while implementation focuses on carrying out the assessments. Monitoring ensures that the assessments are conducted as planned, and data processing involves analyzing the results. Data reporting provides feedback

to stakeholders, and the utilization of evaluation achievements involves applying the insights gained to improve educational practices and outcomes (Karnovi dkk., 2020).

According to Harun, the Criterion Referenced Test (PAP) serves as a method for comparing student learning outcomes against pre-established standards. This approach highlights the necessity of setting minimum standards prior to conducting assessments, which then act as benchmarks or reference points for evaluating student performance. These pre-determined standards are crucial as they provide a clear criterion against which students' grades can be measured.

In educational assessment, these standards are referred to as *Kriteria Ketuntasan Minimal (KKM)*, or Minimum Mastery Criteria. The KKM represents the minimum level of proficiency that students must achieve to be considered as having passed the assessment. Setting a KKM is fundamental because it establishes a baseline for acceptable student performance, ensuring that assessments are aligned with educational goals and standards.

Students who meet or exceed the KKM are deemed to have passed the assessment. Achieving or surpassing this minimum threshold signifies that the student has reached the required level of competency for the subject matter. In some cases, students may exceed the KKM, reflecting higher-than-expected levels of proficiency. Such students are recognized for their exceptional performance and are typically permitted to advance to higher levels of study or more advanced coursework.

Conversely, students who do not meet the KKM are considered to have failed to reach the required standard. For these students, additional support and remediation are necessary to help them strengthen their understanding and skills. These students are generally encouraged to engage in further learning activities designed to help them achieve the minimum standards set by the KKM. This process ensures that all students have the opportunity to reach the "pass threshold" and attain the necessary level of mastery.

Therefore, the KKM serves as a critical benchmark in the assessment process. Graduation criteria are thus based on students' ability to achieve scores that meet or exceed the KKM. This approach ensures that assessments are not only standardized but also aligned with the goal of achieving consistent educational outcomes across different student cohorts. By establishing clear and measurable standards, the PAP method helps to maintain the integrity of the assessment process and supports the objective evaluation of student performance (Sriwahyuni, 2021).

When employing the Criterion Referenced Test (PAP) as an assessment tool, several critical considerations must be addressed to ensure its effectiveness and accuracy. These considerations are essential for maintaining the integrity of the assessment process and ensuring that student evaluations are fair and aligned with educational objectives. The key considerations include:

1. Determining Scores Based on Predetermined Criteria

The first consideration involves establishing clear criteria or standards that will serve as the reference points for grading students' learning tests. The PAP approach necessitates that students' scores be evaluated against these predetermined standards or criteria. This means that the assessment must be designed so that student performance is measured relative to specific, pre-set benchmarks rather than in comparison to the performance of other students. By using these criteria, educators can ensure that grades reflect whether students meet the expected levels of proficiency and understanding. The predetermined standards provide a consistent and objective basis for evaluating student performance, thereby enhancing the reliability and validity of the assessment.

2. Aligning Grades with Predetermined Benchmarks

The second consideration emphasizes the importance of aligning the grades awarded to students with the predetermined criteria or benchmarks set forth in the PAP. The grading process must accurately reflect the students' achievements relative to these benchmarks, whether the outcomes are high or low. This alignment ensures that grades are not arbitrary but are based on the student's actual performance in relation to the established standards. The PAP approach helps in maintaining transparency in grading by making sure that the evaluation is directly tied to the specific goals and standards outlined at the outset of the assessment. This practice helps in providing a clear understanding of how well students have met the educational objectives and ensures consistency in grading.

3. Determining Grades According to PAP Guidelines

The third consideration involves the comparison of raw scores obtained by students with the ideal maximum scores, as specified in the PAP guidelines. This process requires that the determination of student grades be based on a comparative analysis of the raw scores against the ideal or maximum

possible scores for the assessment. By following these guidelines, educators can systematically evaluate how well students have performed in relation to the highest standards of achievement. This comparison helps in determining the level of mastery demonstrated by the students and in assigning grades that accurately reflect their performance. The PAP framework provides a structured approach to grading that ensures consistency and fairness by aligning students' scores with the established ideal benchmarks.

Interpreting test results using the Criterion Referenced Test (PAP) approach involves a detailed comparison between students' actual scores and the predetermined minimum standards, known as *Kriteria Ketuntasan Minimal (KKM)*. This process ensures that students' performance is evaluated relative to specific, established benchmarks rather than in comparison to their peers. The interpretation of results using the PAP approach requires a thorough and structured methodology that includes several critical steps and considerations.

1. Establishing the Minimum Limit (KKM)

Before interpretation can occur, the minimum limit or *KKM* must be determined through a collaborative process involving all relevant stakeholders. This agreement encompasses not only the minimum criteria required for students to pass but also includes achievement criteria, material mastery criteria, and minimum graduation criteria. Achievement criteria refer to the standards that students must meet to demonstrate their competence in specific learning objectives. Material mastery criteria outline the level of understanding required for each topic or subject area, while minimum graduation criteria define the overall standards necessary for students to successfully complete their studies. Reaching a consensus on these criteria is essential for ensuring that the assessment process is fair, transparent, and aligned with educational goals.

2. Utilizing Conversion Standards or Scales

In addition to setting the *KKM*, it is crucial to establish conversion standards or scales for interpreting test scores. These scales determine how raw scores are translated into meaningful grades or evaluations. The conversion scale used can vary depending on the agreement reached by the educational institution. Commonly used scales include:

- The five-point scale (0 – 4 or E – A), which provides a range of grades from the lowest to the highest, with corresponding letter grades.
- The ten-point scale (1 – 10), which offers a more granular assessment by providing ten distinct levels of performance.
- The eleven-point scale (0 – 10), which extends the ten-point scale by including an additional level for more precise differentiation.
- The one hundred-point scale (1 – 100), which allows for detailed grading with a wide range of possible scores.

The choice of scale depends on the specific needs and agreements of the educational institution. The scale used must be consistently applied to ensure that grading is fair and comparable across different assessments.

3. Comparing Scores to Ideal Maximum Scores

Determining the passing grade limit in the PAP approach involves comparing students' raw scores to the maximum ideal score. For instance, if the ideal score for a test is set at 100, a student's performance is evaluated relative to this maximum score. If a student scores 65 out of 100, this can be translated to a score of 6.5 on a ten-point scale. This method of comparison ensures that student's achievements are assessed in relation to the highest possible score, providing a clear indication of their proficiency and areas for improvement.

In summary, the PAP approach to interpreting test results requires a systematic process that includes establishing *KKM* through stakeholder agreement, using appropriate conversion scales, and comparing students' scores to ideal maximum scores. This approach ensures that student performance is evaluated against well-defined criteria, providing a consistent and objective measure of academic achievement. Adhering to these steps helps maintain the integrity of the assessment process and supports the fair evaluation of student learning outcomes.

Implementation of the Benchmark Approach at MI Mifathul Huda

The LAP approach can be carried out in the following ways:

1. Looking for the ideal score, namely the score that students might achieve if all the questions could be answered correctly.
2. Find the ideal average (X) with the formula:
$$X = \frac{1}{2} \times \text{ideal score}$$
3. Find the ideal standard deviation (s) using the formula:
$$s \text{ ideal} = \frac{1}{3} \times X \text{ ideal}$$
4. Develop conversion guidelines.

Example: The ideal score determined by an educational institution based on the answer key = 80.

Based on this data, processing can be carried out:

1. Ideal score = 80
2. Average (X) is ideal:
$$X = \frac{1}{2} \times 80$$
$$= 40$$
3. Simpangan baku (s) ideal :
$$s \text{ ideal} = \frac{1}{3} \times 40$$
$$= 13,33$$
4. Conversion guidelines:
Scale four
$$X + (1.5 \text{ SD}) \text{ up} = A$$
$$X + (0.5 \text{ SD}) \text{ up} = B$$
$$X - (0.5 \text{ SD}) \text{ up} = C$$
$$X - (1.5 \text{ SD}) \text{ down} = D$$

Then we get:
$$40 + (1.5 \times 13.33) = 60$$
$$40 + (0.5 \times 13.33) = 47$$
$$40 - (0.5 \times 13.33) = 33$$
$$40 - (1.5 \times 13.33) = 20$$

After conducting an evaluation of the Arabic language subject, the subsequent analysis of students' scores involves a structured approach using benchmark assessment methods and conversion to a 4-point or letter grading scale. This process ensures that the assessment results are systematically interpreted and converted into a standardized format for clear communication of student performance. The evaluation consists of a comprehensive set of 30 questions, which are categorized into three distinct types: multiple choice questions, fill-in-the-blank questions, and descriptive questions. Each question type is assigned specific scoring values to reflect the varying levels of complexity and cognitive demand.

1. Scoring Breakdown

Multiple Choice Questions. The evaluation includes 15 multiple choice questions. Each multiple choice question is assigned a score of 2 points. This scoring reflects the fundamental nature of multiple choice questions, which typically require students to select the correct answer from a set of options. The total score for multiple choice questions is therefore 30 points (15 questions x 2 points each).

Fill-in-the-Blank Questions. There are 10 fill-in-the-blank questions included in the evaluation. Each of these questions is allocated a score of 3 points. Fill-in-the-blank questions assess students' ability to recall and apply specific information, requiring a higher level of cognitive engagement compared to multiple choice questions. The total score for fill-in-the-blank questions amounts to 30 points (10 questions x 3 points each).

Descriptive Questions. The evaluation also includes 5 descriptive questions, with each question worth 4 points. Descriptive questions are designed to evaluate students' ability to articulate their understanding and provide detailed responses, reflecting a deeper level of comprehension and analysis. The total score for descriptive questions is 20 points (5 questions x 4 points each).

2. Calculation of Ideal Maximum Score

The total possible score for the Arabic language evaluation is calculated by summing the scores for all question types. The ideal maximum score is thus determined to be 80 points, which is the aggregate of the individual scores for multiple choice questions (30 points), fill-in-the-blank questions (30 points), and descriptive questions (20 points).

3. Benchmark Assessment and Conversion

Following the scoring, students' raw scores are analyzed using benchmark assessment methods. This analysis involves comparing the obtained scores against pre-established criteria or benchmarks to evaluate student performance relative to set standards. Additionally, the raw scores are converted into a standardized 4-point or letter grading scale. This conversion facilitates the interpretation of results and provides a clear representation of students' performance levels.

Categories :

Table 1. Assessment Categories

Score range	Predicate	Categories
60 - 80	A	Very Good
44 - 59	B	Good
33 - 43	C	Fair
20 - 32	D	Poor

Explanation of Criteria:

To be declared COMPLETE, each student must meet passing grade 33 or C. This means that a student who has not mastered at least 55% of the knowledge and skills as stated in the learning objectives has not been declared complete and must take remedial tests.

Raw score processing

The following 10 students took the Arabic test with the following results:

Table 2. Assessment Results

No	Score	Predicate
	60	A
	62	A
	45	B
	55	B
	55	B
	50	B
	46	B
	44	B
	34	C
	30	D

Based on the table above, there is one student who does not meet the predetermined completion criteria, namely 55%, and needs to take remedial action. According to Erlina, remedial is carried out by teachers for students in an effort to help improve the grades of students who have not met the complete criteria (Rismanda & Sri Hartini, 2017). This remedial is one of the advantages of PAP, apart from that PAP has several advantages, such as the following; can help teachers design remedial programs; does not require complex statistical calculations; can measure whether learning objectives have been achieved; the value remains as long as the standard used is the same; and the results of the assessment can be used for feedback or to find out whether the learning objectives have been achieved or not. This approach is also widely used in classes in providing learning materials (Suryadi, 2019). So, based on the explanation above, it is hoped that the choice of the PAP approach can provide feedback on achieving the learning objectives carried out at MI Miftahul Huda.

Apart from the advantages, there are also disadvantages to the PAP, including: a) it is relatively complicated because it takes time to agree on standards and criteria; b) Ratings may be falsely stated positively or negatively; c) PAP is not used to process or determine summative test scores, such as in general exams to fill in report cards, final exams to fill in diploma grades, or to determine graduation. d) Students will be given low marks if the questions given are too difficult, and e) the stupidest students will be given high marks if the questions given are too easy. f) Academics sometimes do not have the ability or confidence necessary to make professional judgments (Pangastuti & MunfaAti, 2018).

Based on the results of the analysis above, it can be seen that the use of a benchmark assessment approach using letter scale conversion is not only used at the higher education level, but MI Miftahul Huda also uses this assessment approach. This means that the benchmark assessment approach can also be used at various levels of education. The choice of assessment approach depends on each educational institution. These institutions have the right to determine which assessment they wish to use according to the situation and conditions in their respective institutions.

CONCLUSIONS

Based on the results of the analysis, it can be concluded that MI Miftahul Huda uses a benchmark assessment approach in analyzing or processing grade 4 student learning outcomes scores in Arabic subjects and converting them using a letter scale. The aim of using the PAP approach is to make it easier for teachers to improve scores or remedial measures when students do not reach the predetermined criteria, namely at least 55% or the equivalent of 33 or predicate C. Based on the processing of scores into grades that has been carried out by teachers in Arabic subjects. class 4, there is one student who has not achieved a passing grade of 55% so that a grade correction or remedial needs to be carried out to achieve these criteria.

Based on the conclusions above, the use of the benchmark assessment approach (PAP) with letter scale conversion when used at the Madrasah Ibtidaiyah level can be a more practical assessment approach, apart from being easy to use, this assessment approach can also monitor the level of student development, making it easy to determine the level of development from time to time. Apart from that, this assessment approach also makes it easier for teachers to carry out remedial measures for students who do not reach the minimum threshold.

BIBLIOGRAPHY

- Alfath, K. (2019). Teknik pengolahan hasil asesmen: Teknik pengolahan dengan menggunakan pendekatan acuan norma (PAN) dan pendekatan acuan patokan (PAP). *AL-MANAR: Jurnal Komunikasi dan Pendidikan Islam*, 8(1), 1–28.
- Arifin, Z. (2009). *Evaluasi pembelajaran* (Vol. 2). Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (2010). Metode penelitian. *Jakarta: Rineka Cipta*, 173.
- Devi, M. Y., Hidayanthi, R., & Fitria, Y. (2022). Model-model evaluasi pendidikan dan model sepuluh langkah dalam penilaian. *Jurnal Basicedu*, 6(1), 675–683.
- Efrianti, L., & Zakir, S. (2024). PERSEPSI GURU TERHADAP PEMBERLAKUAN KURIKULUM MERDEKA DI SMAN KOTA PAYAKUMBUH. *Jurnal Inovasi Pendidikan*, 7(3).
- Intiana, H. S. R. ., Syaiful Musaddat, S., & Muh. Khairussibyan, R. H. (2023). *PEMBELAJARAN BERBASIS LESSON STUDY FOR LEARNING COMMUNITY DI PRODI PENDIDIKAN BAHASA DAN SASTRA INDONESIA FKIP UNIVERSITAS MATARAM*. 4(2), 179–187.
- Karnovi, R., Habibi, R., & Fauzan, M. N. (2020). *Tutorial membuat aplikasi sistem monitoring progres pekerjaan dan evaluasi pekerjaan pada job desk operational human capital menggunakan metode naive bayes*. Kreatif.
- Khalim, A. D. N. (2019). Pola Manajemen Pembelajaran Akidah Akhlaq Di MTs Negeri 6 Sleman. *Al-fahim: Jurnal Manajemen Pendidikan Islam*, 1(2), 36–57.
- Magdalena, I., Oktavia, A., & Arlita, N. D. (2020). Analisis Penilaian Acuan Patokan Di Sd Sukasari 3. *EDISI*, 2(2), 244–251.
- Manalu, S. B., Kadir, A., & Siregar, N. S. S. (2020). Implementasi Peraturan Menteri Agama Tentang Pendidikan Agama Islam Pada Sekolah Dalam Kinerja Pengawas Pendidikan Agama Islam. *Strukturasi: Jurnal Ilmiah Magister Administrasi Publik*, 2(2), 168–177.

- Pangastuti, R., & MunfaAti, K. (2018). Penilaian Acuan Norma, Penilaian Acuan Patokan, Kriteria Ketuntasan Minimal Di Madrasah Ibtidaiyah an-Nur Plus Junwangi Krian Sidorajo Jawa Timur. *Tarbiyah Al-Awlad: Jurnal Kependidikan Islam Tingkat Dasar*, 8(2), 202–217.
- Rahmah, L. A., Saputra, P., & Mufidah, N. (2023). Management of Arabic Language Competence: Insiders' Awareness and Experiences/ Manajemen Kompetensi Bahasa Arab: Kedasaran dan Pengalaman Anggota. *Loghat Arabi: Jurnal Bahasa Arab dan Pendidikan Bahasa Arab*, 4(2), 267. <https://doi.org/10.36915/la.v4i2.154>
- Rahman, H., & Maisurah, F. L. (2023). Pengajaran Bahasa Arab dan Sosiologi Masyarakat Islam Indonesia: Kajian Fenomenologi-Sosiolinguistik. *Cognitive: Jurnal Pendidikan dan Pembelajaran*, 1(2), 31–46.
- Rismanda, I., & Sri Hartini, S. (2017). *Peran Guru Dalam Membimbing Anak Berkesulitan Belajar Matematika Pada Siswa Kelas III SDN Dlepih I Kecamatan Tirtomoyo Kabupaten Wonogiri*.
- Safitri, K., & Harjono, N. (2021). Pengembangan Instrumen Penilaian Sikap Sosial Aspek Tanggung Jawab Pembelajaran Tematik Terpadu Siswa Kelas 4 SD. *Jurnal Pedagogi Dan Pembelajaran*, 4(1), 111–121.
- Sholiha, R., & Rizal, M. S. (2023). Pelaksanaan dan Hambatan Evaluasi Formatif dalam Pembelajaran Menulis Teks Laporan Hasil Observasi di SMK PGRI 3 Malang. *Jurnal Pendidikan Bahasa*, 12(1), 192–209.
- Sinulingga, E. D. B. (2022). Urgensi Integrasi Ilmu Sebagai Upaya Menumbuhkan Karakter Islami dalam Proses Pembelajaran. *Book Chapter of Proceedings Journey-Liaison Academia and Society*, 1(1), 189–195.
- Sriwahyuni, E. (2021). *Pengaruh Model Pembelajaran Snowball Throwing Terhadap Hasil Belajar Pada Pembelajaran IPS Kelas VIII SMP Islam YPI 3 Way Jepara*.
- Sugiyono, S. (2010). Metode penelitian kuantitatif dan kualitatif dan R&D. *Alfabeta Bandung*, 170–182.
- Sukardi, H. (2008). Evaluasi pendidikan prinsip dan operasionalnya. *Jakarta: Bumi Aksara*, 2.
- Suryadi, E. (2019). *Metode Penelitian Komunikasi (dengan pendekatan Kuantitatif)*. PT. Rosdakarya.
- Yulianto, I. (2024). Analisis Classrooms Assessment: Remedial, Pengayaan, Pendekatan Acuan Patokan (PAP) dan Pendekatan Acuan Normatif (PAN). *Grata: Jurnal Inovasi Pendidikan*, 1(1), 27–36.