Application of Norm-Referenced Assessment at Vocational High School

Nurul Syahrayni[1*], Hasna'i Masruhani[2] , Supriyansyah[3], Nisaurrahmah[4]
[1, 2, 3, 4] Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia7

Abstract:
Assessment is a critical component of the educational process, serving as a means of evaluating student learning and ensuring the overall quality and continuity of instruction. Two major approaches commonly used in educational assessment are norm-referenced assessment (NRA) and criterion-referenced assessment (CRA). This study aims to investigate the implementation of norm-referenced assessment in the context of Arabic language instruction at SMK Muhammadiyah 2 Malang, Indonesia. Employing a qualitative research design, the study utilized the Miles and Huberman interactive model for data collection and analysis. The participants consisted of 12th-grade students and Arabic language teachers. The findings reveal that although teachers are generally familiar with the concept of NRA, both norm-referenced and criterion-referenced assessment approaches have not been explicitly or systematically applied in the classroom. These results suggest the need for increased awareness and training in assessment methodologies to improve instructional effectiveness and student evaluation.
Keywords: *Norm-Referenced Assessment, Criterion-Referenced Assessment, Arabic Language Education,*

Abstrak:
Asesmen merupakan komponen krusial dalam proses pendidikan yang berfungsi untuk mengevaluasi hasil belajar peserta didik serta menjamin kualitas dan keberlanjutan pembelajaran. Dua pendekatan utama yang umum digunakan dalam evaluasi pembelajaran adalah penilaian acuan norma (PAN) dan penilaian acuan patokan (PAP). Penelitian ini bertujuan untuk mengkaji implementasi asesmen acuan norma dalam konteks pembelajaran bahasa Arab di SMK Muhammadiyah 2 Malang, Indonesia. Penelitian ini menggunakan pendekatan kualitatif dengan model analisis interaktif Miles dan Huberman untuk pengumpulan dan analisis data. Partisipan dalam penelitian ini terdiri dari siswa kelas XII dan guru bahasa Arab. Hasil penelitian menunjukkan bahwa meskipun para guru umumnya memahami konsep PAN, baik pendekatan asesmen acuan norma maupun asesmen acuan kriteria belum diterapkan secara eksplisit dan sistematis dalam pembelajaran. Temuan ini menunjukkan perlunya peningkatan pemahaman dan pelatihan terkait metode asesmen guna mendukung efektivitas pembelajaran dan evaluasi peserta didik.
Kata Kunci: *penilaian acuan norma, penilaian acuan patokan , pembelajaran bahasa arab*

INTRODUCTION / المقدمة

Designing the learning process is a fundamental task that educators must undertake before initiating instruction. This includes the selection of media, methods,

models, and particularly, the formulation of assessments to be administered at the conclusion of the learning process. Assessment is one of the most critical components of the learning evaluation cycle. It allows teachers to determine whether learning activities have been conducted effectively and efficiently. Generally, the higher the students' achievement scores, the more apparent the quality of the teaching and learning process becomes(Aliyan & Dayanti, 2021).

Teacher-conducted assessments serve not only as feedback for students but also provide valuable information for teachers to monitor student progress and mastery of learning content(Rapono, Safrial, & Wijaya, 2019). Each educational institution often employs distinct assessment approaches tailored to the needs of teachers, students, and the institution itself. As such, assessments must align with clearly defined standards. In contrast, when assessments are carried out arbitrarily without specific criteria or benchmarks, the resulting data may lead to inaccurate interpretations of students' abilities and competencies(Billah & PAP, 2021). Therefore, institutions must adopt appropriate reference frameworks to determine the extent to which students have achieved learning success or experienced failure.

Assessment models applied by teachers are generally divided into two categories: norm-referenced assessment (NRA) and criterion-referenced assessment (CRA). These two models are based on differing assumptions regarding student capabilities. One institution that applies the norm-referenced approach is SMK Muhammadiyah 2 Malang, a vocational high school located in Malang, Indonesia. This school uses NRA to evaluate individual students in comparison with the performance of other students or groups(Vioreza, 2017). The implementation of such a reference model can lead to more definitive assessment outcomes, as demonstrated in SMK Muhammadiyah 2 Malang, where the quality of student learning following a given instructional period can be clearly observed. The use of norm-referenced assessment in specific educational settings thus facilitates accurate, reliable, and objective evaluation of learning implementation and student performance(Magdalena, Fauzi, & Putri, 2020).

Previous studies have provided foundational insights for this research. One such study by Mohammad Nasir explored the philosophical concept of perennialism in the context of learning outcome assessment and its application in schools. The findings indicated that perennialist-oriented evaluation had a measurable impact on students' academic achievement(Nasir, 2010). Another study by Khaeruddin Al-Fath examined assessment processing techniques using both norm-referenced and criterion-referenced models, offering practical insights into how each approach can be implemented in educational contexts(Alfath, 2019). Similarly, research by Ratna Pangastuti investigated the use of NRA and CRA in meeting minimum competency standards at an elementary school in East Java, finding that NRA refers to group-based standards while CRA relates to the achievement of predefined learning objectives and indicators(Pangastuti & MunfaAti, 2018).

The main distinction between those previous studies and the present research lies in their focus. While earlier studies concentrated on the technical aspects of processing assessment results, this study investigates the practical implementation of norm-referenced assessment (NRA) specifically at SMK Muhammadiyah 2 Malang, aiming to examine how NRA is applied within the school's learning environment.

RESEARCH METHOD / المنهجية

This study employs a qualitative approach, as the data required are not in the form of statistical or numerical figures but rather descriptive and detailed information related to the research subject. Bogdan and Taylor define qualitative research as an approach that focuses on obtaining textual data derived from spoken or written words, as well as observed behaviors(Moleong, 2017). The type of research conducted is descriptive-analytical, as the researcher aims to describe and analyze the phenomenon systematically to inform others about the application of norm-referenced assessment (NRA) at SMK Muhammadiyah 2 Malang, which presents a unique and compelling case for investigation.

Through this descriptive design, the researcher seeks to present, describe, analyze, and interpret facts as they are, without the intention of testing hypotheses(S. Bolster Arthur, 1983). The focus of this research is on ninth-grade students at SMK Muhammadiyah 2 Malang. Respondents were selected based on their representational relevance to the application of the norm-referenced assessment approach implemented in the school.

Given the qualitative nature of the study, the primary research instrument is the researcher himself, who directly enters the field to collect data using observation, interviews, and documentation techniques.

After data collection, the researcher analyzes the information using the data analysis framework developed by Miles and Huberman(Miles, Huberman, & Saldaña, 2014), which consists of the following steps: 1) Data reduction, which involves summarizing, selecting essential elements, focusing on relevant points, identifying core themes, and simplifying field data; 2) Data display, which refers to organizing and presenting the data in a way that allows for drawing conclusions and facilitating further action; 3) Conclusion drawing and verification, which involves interpreting the meaning of observed phenomena and validating the findings based on the field data.

FINDINGS AND DISCUSSION / نتائج البحث و المناقشة

The scores obtained by students are compared with those of other students within the same group. In this context, the term "norm" refers to the performance or achievement of the group, while the "group" itself includes all students who take the test. Thus, the group may refer to a class, school, region, district, or province.

Interpreting test scores based on norms provides insight into how an individual student compares to their peers taking the same assessment. The simplest and most commonly used method of comparison in norm-referenced assessments is ranking the scores from highest to lowest and then determining the student's position relative to others.

One important consideration when applying norm-referenced standards is that the reference group should accurately represent all students who are taking the test. To effectively compare student abilities, it is essential to consider students with similar age, grade level, and educational background.

An example of the application of a norm-referenced assessment is when a teacher evaluates a student's performance by comparing it to that of other students. Here, the standard for determining success or failure is based on the achievements of peers, not on predefined learning criteria. For instance, a student named Tara is considered successful because her score exceeds that of her classmates or surpasses the class average. This approach is typical in ranking systems such as end-of-semester report cards or promotion reports, where standards may fluctuate depending on the group average. Thus, if students

generally perform well, the norm or benchmark rises accordingly.

At SMK Muhammadiyah 2 Malang, various forms of assessment are employed by teachers and the school to measure students' academic activities, skill development in the teaching-learning process, and their industrial field practice. Students' cognitive competencies are typically assessed through daily quizzes, mid-term exams, and final examinations. For psychomotor assessment, teachers assign additional tasks outside of regular school hours. For example, students may be encouraged to develop entrepreneurial skills by selling school-made products to the community, thereby nurturing business acumen while still in school. Affective assessment is observed through students' participation in school-designed religious programs aimed at deepening their spiritual commitment.

In cognitive assessments, teachers measure final learning outcomes based on daily tests, mid-semester, and final exams. These scores are recorded on report cards coordinated by the Vice Principal of Curriculum. These reports enable teachers to identify students who may still struggle in cognitive, psychomotor, or affective domains.

Each student's report card contains multiple columns that teachers must complete for each subject. The "Knowledge" column includes categories such as daily test scores, homework or assignment scores, midterm exam scores, and final exam scores. The "Skills" column includes performance/practical tests, project assessments, and portfolios. The final column presents a composite report that integrates both cognitive and psychomotor elements, providing a comprehensive overview of the student's performance in each subject.

Table of Arabic Language Scores for Grade 11 Students

Class Category	Scores
Accounting	78, 80, 66, 64, 71, 68, 82, 77, 71, 70, 66, 64, 80, 80
Multimedia	51, 53, 51, 65, 68, 68, 48, 48, 51, 60, 71, 62, 71, 51
Office Management	83, 76, 50, 56, 74, 78, 75, 81, 72, 79, 64, 68, 62, 60
Network Engineering	56, 40, 55, 78, 47, 67, 50, 54, 57, 43, 67, 60, 55, 60
Marketing	60, 65, 70, 81, 80, 47, 68, 63, 58, 79, 78, 61, 50, 40

From the above data, it is evident that the score ranges differ across class categories. In each class, the teacher sets a reference value or benchmark according to the students' performance in that specific group. For instance, in the Accounting class, the minimum score is 64 and the maximum is 82, while in the Multimedia class, the minimum score is 48 and the maximum is 71.

This comparison indicates that the minimum score in Accounting (64) becomes approximately the average score in Multimedia, while the maximum score in Multimedia (71) corresponds to the average score in Accounting. Such variations reflect that each class has distinct achievement characteristics, and therefore, a uniform standard across all classes is not appropriate.

Teachers cannot equate performance standards across categories such as Accounting and Multimedia due to the differing score distributions and student capabilities in each group. The assessment must be contextualized within each specific class or group, aligning with the principles of norm-referenced assessment, which evaluates students' learning outcomes relative to their peers within the same group(Nasir, 2010). Students are

grouped based on similar learning outcome levels, allowing for a clearer picture of each student's relative position within the class context.

Table of Grade 11 Students' Scores in Computer and Network Engineering Class

Subject: Arabic Language

Class: 11

No.	Name	Knowledge	Skills	Final Score	Grade
1	Achmad Hildan Nur Zian	90	83	85	B+
2	Agung Prasetyo Putra	84	80	81	B
3	Akhmad Zacky Ramadhan	84	80	81	B
4	Andreano Ferdiansyah	95	83	87	B+
5	Atika	90	83	85	B+
6	Desti Ayu Nurhidayah	90	83	85	B+
7	Dewi Ramadani	95	83	87	B+
8	Dewi Ratu Asriani Rahayu	84	80	81	B
9	Hikmal Aditya Pradana	77	82	80	B
10	Joe Farozal Namana Sitompul	90	83	85	B+
11	Moh. Adi Setiawan	96	83	87	B+
12	Muhammad Yusuf Asrofi	77	83	81	B
13	Nofa Fitria	92	83	86	B+
14	Revaldo Radhitya Dhio Ramadani	96	83	87	B+
15	Wahyuning Tyas	90	83	85	B+
16	Ad Dinda Sholehati	90	83	85	B+

Based on the table above, it is evident that the norm-referenced evaluation in Arabic language learning for the Computer and Network Engineering class operates on a performance scale where the highest score is 87 with a B+ (very good) grade, and the lowest score is 80–81, corresponding to a B (good) grade.

This pattern indicates that the norm-referenced approach is used to differentiate students based on their relative performance within the same class. The grading thresholds observed are approximately:

- 85–90: Grade B+ (Very Good)
- 80–84: Grade B (Good)

However, interviews conducted with several teachers at Muhammadiyah Vocational Senior High School 2 Malang reveal that the implementation of norm-referenced assessment remains relatively basic and superficial, lacking rigorous theoretical and methodological foundations.

This limited implementation stems from teachers' lack of familiarity with the concept and correct calculation techniques of norm-referenced assessment. In practice, it is used primarily to rank students within the same class from the highest to the lowest

performer. Test items are developed with varying levels of difficulty to enable comparison among students. The evaluation is typically conducted at the end of the learning process to determine the students' learning rank based on their final semester exam results.

This practice confirms that the school utilizes a relative performance scale, rather than criterion-based standards, to assess student achievement.

CONCLUSION / الخلاصة

From the above discussion, it can be concluded that the implementation of norm-referenced assessment at Muhammadiyah Vocational Senior High School 2 Malang has not been deeply or systematically applied in determining student learning outcomes.

This conclusion is supported by interviews with several teachers at the school, who explained that the grading process carried out by most teachers does not prioritize the principles of norm-referenced assessment. The limited application is primarily due to the teachers' lack of in-depth understanding of the concept and procedures of norm-referenced assessment. As a result, student evaluations are often based on general performance without rigorous reference to comparative group analysis, which undermines the full potential of this assessment approach in reflecting relative student achievement accurately.

This research is limited to the application of the norm-referenced assessment approach at Muhammadiyah Vocational Senior High School 2 Malang. The scope was adjusted by the researchers to focus only on variables relevant to the norm-referenced assessment, based on data that were considered saturated after conducting the study within this specific context.

Therefore, future researchers are encouraged to explore the implementation of norm-referenced assessment more broadly—either by expanding the investigation within the same institution or by conducting studies at other educational levels that are more complex and widely implemented. Such future work could provide deeper insights and comparative analyses that may contribute to the development of more comprehensive assessment practices in vocational or general education settings.

REFERENCES / المراجع

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.

Aliyan, D. N. A., & Dayanti, F. (2021). Implementasi Pendekatan Penilaian Acuan Normatif (PAN) Dan Pendekatan Penilaian Acuan Patokan (PAP) Dalam Mengevaluasi Hasil Belajar Siswa (Studi Kasus Siswa Kelas XI Sma Gema 45 Surabaya). *At-Ta'lim: Media Informasi Pendidikan Islam*, *20*(2), 183–191.

Billah, M. F., & PAP, P. A. P. (2021). Analisis Penilaian Pada Mata Pelajaran Alquran Hadist Mengacu Pendekatan Acuan Patokan (Pap) Kelas 2C Di Mi Al Hidayah Tarik Sidoarjo. *No. July*.

Magdalena, I., Fauzi, H. N., & Putri, R. (2020). *Pentingnya evaluasi dalam pembelajaran dan akibat memanipulasinya*.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook. 3rd*. Thousand Oaks, CA: Sage.

Moleong, L. J. (2017). *Metodologi penelitian kualitatif/Lexy J. Moleong*.

Nasir, M. (2010). Konsep Penilaian Hasil Belajar Perenialisme Dan Implementasinya Di Madrasah. *Dinamika Ilmu*, *10*(2).

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. *Jurnal Tarbiyah Al-Awlad-Jurnal Kependidikan Islam Tingkat Dasar*, *8*(2), 202–217.

Rapono, M., Safrial, S., & Wijaya, C. (2019). Urgensi Penyusunan Tes Hasil Belajar: Upaya Menemukan Formulasi Tes Yang Baik dan Benar. *Jupiis: Jurnal Pendidikan Ilmu-Ilmu Sosial*, *11*(1), 95–104.

S. Bolster Arthur, J. (1983). Toward a more effective model of research on teaching. *Harvard Educational Review*, *53*(3), 294–308.

Vioreza, N. (2017). Pengaruh Pendekatan Penilaian dan Gaya Belajar Terhadap Motivasi Berprestasi Mahasiswa. *Visipena*, *8*(1), 1–16.