Comparative Study of Norm Based Assessment and Benchmark Based Assessment in Arabic Learning Evaluation

Retisfa Khairanis¹ Ahmad Rifai², Nur Qomari³
¹,²,³Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia
E-mail: retisfakhairanis182001@gmail.com, rifaialtair@gmail.com , qomari@uin-malang.ac.id

Article Info	Abstract
	Assessment in Arabic language learning plays a crucial role in measuring students' learning achievement objectively and fairly. This article comparatively examines two main approaches to learning evaluation, namely Norm-Based Assessment (PAN) and Benchmark-Based Assessment (PB). PAN assesses student achievement based on comparison with the group, while PAP measures achievement based on predetermined standards. This research uses a qualitative approach with a library research method, comparative analysis of the characteristics, advantages and limitations of each approach, as well as its impact on the fairness and accuracy of assessment results. Data were obtained through a literature study that included scientific journals, textbooks, and evaluation policy documents. The results show that PAN tends to be competitive and relative, while PAP emphasizes individual achievement in line with learning objectives. The implications of these findings emphasize the importance of choosing an evaluation model that is appropriate to the context and objectives of Arabic language learning to ensure fairness of assessment and improve the quality of education.

Artikel Info	Abstrak
	Penilaian dalam pembelajaran Bahasa Arab memegang peran krusial dalam mengukur pencapaian belajar siswa secara objektif dan adil. Artikel ini mengkaji secara komparatif dua pendekatan utama dalam evaluasi pembelajaran, yakni Penilaian Acuan Norma (PAN) dan Penilaian Acuan Patokan (PAP). PAN menilai capaian siswa berdasarkan perbandingan dengan kelompoknya, sementara PAP mengukur pencapaian berdasarkan standar yang telah ditetapkan sebelumnya. Penelitian ini menggunakan pendekatan kualitatif dengan metode studi kepustakaan (library research), analisis komparatif terhadap karakteristik, kelebihan, dan keterbatasan masing-masing pendekatan, serta dampaknya terhadap keadilan dan keakuratan hasil asesmen. Data diperoleh melalui studi pustaka yang mencakup jurnal ilmiah, buku teks, dan dokumen kebijakan evaluasi. Hasil kajian menunjukkan bahwa PAN cenderung bersifat kompetitif dan relatif, sementara PAP lebih menekankan pada pencapaian individual yang sesuai dengan tujuan pembelajaran. Implikasi dari temuan ini menekankan pentingnya pemilihan model evaluasi yang sesuai dengan konteks dan tujuan pembelajaran Bahasa Arab untuk menjamin keadilan asesmen dan peningkatan kualitas pendidikan*.*

I. INTRODUCTION

Learning evaluation is a crucial component in the education system that aims to assess the extent to which instructional goals have been achieved. In the context of Arabic language learning, evaluation not only functions as a measuring tool for student achievement, but also as a means of reflecting on teaching effectiveness and curriculum quality. Evaluation becomes a barometer of the success of pedagogical intervention as well as an instrument of accountability in the formal education process. (Irawan, Hanip, & Jasnur, 2025).

One of the main issues in learning evaluation is the approach used to assess learning outcomes. The two most widely used approaches in educational practice are norm-referenced assessment (PAN) and benchmark-referenced assessment (PAP). Each has different basic assumptions, methods of interpretation, and psychopedagogical impact on learners. In Arabic language education, the choice between PAN and PAP is often made without deep consideration of the consequences that come with it (IRAWAN, 2025).

The study of the effectiveness and relevance of both approaches is important because the choice of evaluation method will affect the validity and reliability of assessment results. PAN, which compares students' results with their peer group, can encourage competition but often ignores absolute individual achievement. PAP, on the other hand, refers to specific standards that students must achieve, so it focuses more on the

achievement of competencies (Busnawir et al., n.d.).

The importance of evaluating and comparing PAN and PAP in the context of Arabic language learning is based on the need for an evaluation system that is fair, accurate and relevant to learning objectives. Arabic as a subject has complex linguistic and cultural characteristics, thus requiring an evaluation model that is able to capture these complexities objectively and systematically (Rofii, Yuniarti, & Wahyuni, 2025).

A number of previous studies have discussed the effectiveness of PAN and PAP in various educational contexts. For example, Nitko (2004) showed that PAN tends to produce a centralized distribution of grades, while PAP emphasizes the achievement of minimum standards. However, most of these studies are still general in nature and have not been specifically applied to the context of evaluating Arabic as a school subject (Raynaldi, 2025).

On the other hand, research by Brown and Abeywickrama (2019) states that the selection of evaluation methods should be tailored to learning objectives and learner characteristics. They also emphasize the need for empirical testing of the effectiveness of evaluation approaches in specific subject contexts. However, the study has not explicitly examined the context of Arabic language learning in the Indonesian educational environment or countries with similar curriculum systems (Lestari, 2025).

There is a gap in the literature regarding the selection of evaluation methods in Arabic language learning. Most studies focus more on teaching strategies or the development of teaching materials, while evaluative approaches receive less attention. In fact, the fairness and accuracy of assessment determine the success of the teaching and learning process as a whole (ريتسفا خيرانس &محمد ألدي, 2025).

In addition, there is still controversy over the effectiveness of PAN in providing a true picture of students' abilities. In some cases, it creates unhealthy competitive pressures, especially for students with varied learning backgrounds. On the other hand, PAP is often criticized for being too rigid and not taking into account the relative development of students in their group (Atin, 2024).

This study seeks to fill this gap by systematically comparing PAN and PAP in the context of Arabic language learning evaluation. Using a comparative approach, this study aims to analyze the characteristics of each method, as well as its impact on the fairness and accuracy of evaluation results (Amal, 2024).

The main objective of this research is to determine the most appropriate, fair and accurate evaluation method in measuring Arabic learning outcomes. Specifically, this study aims to answer the question: "Which evaluation method is more suitable to be used in Arabic language learning at school, PAN or PAP, in terms of fairness of assessment and accuracy of measurement of learning outcomes?"

In conducting this review, the research refers to a conceptual framework that makes a clear distinction between PAN and PAP. PAN is defined as an evaluation system that compares an individual's performance with that of his or her group, while PAP refers to an evaluation based on specific criteria or pre-determined learning objectives.

Theoretically, PAN is based on classical measurement theory, which emphasizes normal distribution and inter-individual reliability within groups, while PAP is rooted in behavioristic and objectivistic approaches, which emphasize the achievement of predetermined learning outcomes. These two approaches have different psychometric foundations, which directly affect the validity of the interpretation of scores. (Rosalinda, 2024).

The literature on evaluation theory as presented by Airasian and Russell (2008) underscores the importance of the fit between instructional objectives and evaluative approaches. Choosing a model that is not aligned with the characteristics of the subject can obscure the understanding of student achievement. This is particularly relevant in Arabic language learning which demands mastery of grammatical, phonological and semantic aspects (Rinantanti, n.d.).

In the context of Arabic, some studies such as Al-Mekhlafi (2010) emphasize that learning success is strongly influenced by the assessment strategies applied. When evaluation is too normative, students tend to focus on comparisons between individuals rather than the development of personal competencies. In contrast, PAP is considered to encourage a constructive and individualistic approach to learning (Aldi, Supriyatno, Trinova, & Muslimin, 2025).

However, the main challenge in implementing PAP is determining valid and reliable criteria. Teachers often struggle to design assessment indicators that truly reflect Arabic learning objectives. This is where the role of this research

becomes significant-critically evaluating the effectiveness of PAN and PAP based on empirical data and conceptual reflection.

Thus, this study not only makes a theoretical contribution to the debate on evaluation approaches, but also provides practical implications for teachers, curriculum designers and educational policy makers. It is hoped that the results of this study can be a reference in choosing an evaluation approach that is more fair, accurate, and in line with the principles of pedagogy in Arabic language learning.

II. METHOD

This research uses a descriptive qualitative approach with a library research method to analyze and compare two learning evaluation approaches, namely Normative Assessment (PAN) and Benchmark Assessment (PAP), in the context of Arabic language learning (Pringgar & Sujatmiko, 2020). This approach was chosen because the main objective of the research was to conduct theoretical exploration and critical analysis of the concepts, advantages, limitations, and implications of each assessment model. The data analysis technique was carried out qualitatively by reviewing, categorizing, and interpreting various sources of scientific literature relevant to the focus of the study.

The research stages were conducted systematically through three main steps: (1) collecting literature data from various sources such as scientific journals, textbooks, curriculum documents, and education policies; (2) identifying and classifying information related to the characteristics of PAN and PAP, both conceptually and applicatively; and (3) comparative analysis of the strengths and weaknesses of each approach based on the dimensions of assessment fairness and accuracy of evaluation results. The research flow can be described as follows:

Problem Focus Identification → Literature Collection → Classification of PAN and PAP Concepts → Comparative Analysis → Conclusions and Implications.
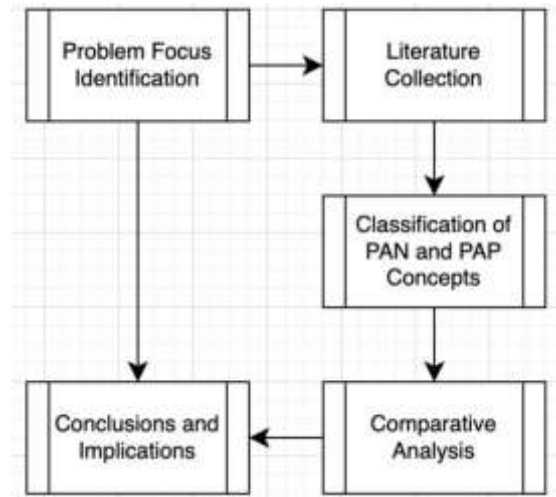

Figure 1. Research framework

Although this research is non-empirical in nature and does not involve direct field data collection, the target population is conceptually the evaluation system applied in Arabic language learning in formal educational settings, particularly the secondary school level. Therefore, the research location is conceptually directed at the context of Arabic language education in Indonesia, given the urgency of implementing a fair and accurate assessment system in the national education system. The results of this study are expected to provide theoretical contributions as well as practical guidance for Arabic language teachers, curriculum developers, and policy makers in designing an appropriate evaluation system.

III. RESULT AND DISCUSSION
1. Evaluation Fairness in the Context of Learner Heterogeneity

The results of the study show that Benchmark Assessment (PB) significantly guarantees evaluation fairness more than Norm-Based Assessment (PAN), especially in the context of heterogeneous classes. In an Arabic language education environment in Indonesia, learners have diverse backgrounds, ranging from students with experience of learning Arabic in Islamic boarding schools to students who have never had an introduction to it before. PAN, which assesses students based on their performance relative to the group, tends to be unfair to students with learning delays, as their achievement is measured not against the learning objectives that have been set, but rather their position in the group score distribution (Lamusiah et al., 2025). In contrast, PAP allows all students to be measured against mutually agreed objective criteria, without competitive

pressure. As such, it is more inclusive and promotes evaluative fairness, which is in line with the principles of humanistic and socially just education currently being promoted in the Merdeka Curriculum .

Table 1. Evaluation Equity in the Context of Learner Heterogeneity

No	Evaluation Aspect	Norm-referenced Assessment (PAN)	Benchmark Assessment (PAP)	Critical Analysis
1	Basis of Assessment	Student performance compared to the group	Student performance compared to specific standards/criteria	PAN is competitive, PAP is objective and individualized.
2	Fairness to heterogeneous students	Unfair to students with weak or minority backgrounds	Fairer because it measures the absolute achievement of each individual	PAP is more inclusive and relevant in a diverse classroom context.
3	Reflects competency attainment	Does not always reflect actual student competence	Reflects actual achievement of learning objectives	PAN can be misleading in assessing learning quality; PAP is more honest about learning outcomes.
4	Psychological impact	Increases negative competition and stratification	Encourages internal motivation and growth	PAP supports growth mindset, PAN reinforces fixed mindset.
5	Relationship to curriculum	Less aligned with differentiation approach (Merdeka Curriculum)	Supports differentiated and personalized learning	PAP is compatible with inclusive and humanistic principles in education today.
6	Diagnostic ability	Difficult to identify specific learning needs	Allows mapping of individual student needs	LAP serves as an effective diagnostic and instructional improvement tool.
7	Digital technology adaptability	Difficult to implement in automated or online evaluation systems	Easy to implement in rubric and indicator-based e-learning platforms	PAP is more compatible with technology-based learning and digital evaluation.
8	Teacher's role in evaluation	Teacher as a classifier of students	Teacher as facilitator of competency achievement	PAP encourages a more educative and reflective role for teachers in the assessment process.

From the table above, we can draw a red thread that a critical comparison between Norm-referenced Assessment (PAN) and Benchmark Assessment (PAP), it can be concluded that PAP is superior in ensuring evaluation fairness, especially in the context of heterogeneous classes such as in Arabic language learning in Indonesia. PAP allows each student to be assessed objectively based on predetermined achievement standards, so it is more in line with the principles of differentiated, inclusive and humanistic learning as promoted in the Merdeka Curriculum. It is also more adaptive to the development of digital evaluation technology and encourages mastery-based learning motivation rather than competition between individuals. In contrast, PAN tends to reinforce academic inequality, create social stratification in the classroom, and risks being misleading in measuring actual learning achievement. Therefore, in the framework of strengthening the quality of Arabic language education that is equitable and transformative, the PAP approach is a more relevant and progressive choice.

In the context of Arabic language learning in Indonesia, the inequality of learners' backgrounds is a big challenge in designing fair and proportional evaluations. Some students come from Islamic boarding school backgrounds with mastery of the basics of nahwu and sharaf, while others are just getting to know the hijaiyah letters for the first time. In this reality, the use of Norm-referenced Assessment (PAN), which compares achievement between individuals in a group, creates a significant bias. Students who in absolute terms have not achieved the minimum competencies can be considered successful simply because their relative performance is higher than the majority of other students who are also less capable (Mubarak & Rahman, 2025).

In contrast, Benchmark Assessment (PB) offers a fairer approach because it is oriented towards objective achievements based on predetermined standards. PB requires teachers to set clear and measurable indicators of competency achievement, such as "students are able to translate simple texts correctly" or "students can write nominal and verbal sentences according to the right structure". This makes it easier to evaluate each student based on individual development and not solely in the context of group competition.

From an educational equity perspective, PB is more relevant in supporting the principle of inclusiveness, especially in an education system that emphasizes quality equity. In the context of implementing the Merdeka Curriculum, the government emphasizes the importance of differentiated learning-that is, providing space for students to learn according to their readiness, interests, and learning styles. PAN contradicts this approach, as it does not make room for differences in students' backgrounds and makes competition the measure of success (Aldi & Khairanis, 2025).

Furthermore, PAN often does not objectively reflect the success of the learning process. For example, if all students in a class score low because learning is not effective, PAN will still produce a distribution of scores with the "best" and "worst" students even though no one reaches the standard in absolute terms. This runs the risk of masking instructional failure because it appears as if there are always students who "succeed", even though success is purely relative

(Nurmalina, Raudatussa'adah, & Hadijaya, 2025).

On the other hand, PB demands clarity of instructional goals and accountability for teaching. Teachers using the PAP approach must develop question grids based on learning objectives, design explicit assessment rubrics, and conduct remedial work for students who have not reached the target. In this way, evaluation becomes a tool for improving the learning process, not just a means of classifying students as is common under the PAN approach.

Evaluation fairness is also closely linked to the principle of assessment for learning, not just assessment of learning outcomes. PAP supports this paradigm because it can be used to map the learning needs of individual students, allowing teachers to design more personalized follow-up learning strategies. This is particularly important in Arabic language learning, which demands gradual skills from phonology recognition to complex sentence construction.

In addition, socially, PAN can reinforce stratification within the classroom, creating a stigma between "smart" and "weak" students. In the long run, this can reduce the confidence of students who are always in the lower group and increase psychological pressure in learning. PAP instead encourages each student to compete with themselves, achieve the targets that have been set, and strengthen the principle of growth mindset in the learning process.

Finally, in the context of Arabic language education, which is now starting to integrate digital technology, the use of PAP is also more easily adapted to online evaluation systems. E-learning platforms can automatically score answers based on criterion-referenced criteria, allowing teachers to monitor each student's progress in real-time. In contrast, PAN requires group distribution data that is not always accurately available in online systems. As such, the LAP is not only fairer, but also more compatible with contemporary developments in education technology.

2. Conceptual and Practical Accuracy of Arabic Language Competency Measurement

From the point of view of evaluation validity, PAP is considered more accurate in measuring Arabic learning outcomes, especially in aspects of linguistic competence (nahwu, sharaf), communicative, and

receptive-productive skills (Sofa, Muarrifah, Hanafi, Parawansah, & Nurhamsalim, 2025). PAN, which relies on the distribution of group scores, does not give a true picture of the absolute extent to which students have mastered the subject matter. For example, in an Arabic text reading exam (qirā'ah), a student can score highly on the PAN even if he or she has not reached the nationally set comprehension standard, as long as he or she is better than his or her classmates. In contrast, PAP demands a minimum standard of achievement (criterion-referenced), so that teachers and students have a more measurable and clear learning direction. This is in line with the demands of the Industrial Revolution 4.0 era, which emphasizes mastery of competencies based on learning achievement indicators, not just relative performance.

Table 2. Accuracy of Measuring Arabic Language Competence Conceptually and Practically

No	Evaluation Aspect	Norm-referenced Assessment (PAN)	Benchmark Assessment (PAP)	Critical Analysis
1	Assessment Orientation	Comparative between students	Based on defined achievement standards	PAN assesses who is the best among students; PAP assesses who has achieved the learning objectives.
2	Construct Validity	Weak - results do not directly represent actual ability	Strong - tailored to language-specific competencies	PB ensures assessment results are relevant to the skills being measured such as qirā'ah, kitābah.
3	Reliability of Assessment	Fluctuates depending on student group composition	Consistent because it is based on fixed indicators	PAP allows for stable and standardized assessment across time.
4	Individualized Ability Measurement	Reduced because it relies on the class average	Focuses on individual achievement against specific targets	PAP is more in-depth in analyzing the needs and development of each student.
5	Strengths of Formative Assessment	Weak, more summative and classificatory in nature	Strong, supporting continuous evaluation for learning improvement	PAP serves as a diagnostic tool for more targeted and personalized learning.
6	Benefits in Instructional Planning	Limited - does not provide competency specific information	High - provides detailed data on achievement of each indicator	PAP is compatible with integrative and skill-level approaches to language learning.
7	Compatibility with Arabic Language Learning	Inflexible, obscures real competence	Very suitable as it can be used for every language skill	PAP is more compatible with technology-based learning and digital evaluation.

From the table above we can draw a red thread that Benchmark Assessment (PAP) is more effective and accurate in measuring Arabic language competencies conceptually and practically than Norm Based Assessment (PAN). PAP provides assessment that is oriented towards predetermined learning outcomes, allows for high construct validity, and supports formative assessment that is

able to reflect individual student progress. With a more consistent and standardized evaluation design, PB not only captures students' abilities objectively, but also provides actionable data for learning improvement. In contrast, PAN, which focuses on comparisons between individuals, tends to make in-depth interpretation of students' real competency achievements difficult, and is less appropriate in the context of Arabic language learning, which emphasizes tiered and integrative understanding.

Assessment in Arabic language learning does not only function as a tool for classifying students, but more than that, it is an important instrument to capture the accuracy of achieving language competence as a whole. Arabic language competencies include cognitive (such as grammar), affective (interest and motivation to learn the language), and psychomotor (speaking and writing skills) dimensions. Within this framework, Benchmark Assessment (PB) offers a more accurate evaluation approach as it is directly oriented towards the achievement of specific learning objectives. Each achievement indicator designed in PB must be in accordance with objectively verifiable competency standards, so that evaluation results truly reflect students' actual abilities (Azhar et al., 2025).

In contrast, norm-referenced assessments tend to obscure the true picture of students' mastery of competencies. For example, a student can be considered "successful" because his score is above the class average, even though he has not achieved full mastery of nominal sentence structure or has not been able to understand informative texts in Arabic well. In PAN evaluations, student achievement is highly dependent on the performance of their classmates, not on the achievement of objective learning indicators. This is pedagogically dangerous because it can create an illusion of success or failure that is not substantively accurate.

The application of PAP also better supports the principle of construct validity, which is the relationship between the measuring instrument and the construct of the ability being measured. In the context of Arabic language learning, construct validity is very important because each language skill has a specific structure and mastery objectives, such as distinguishing between receptive (listening and reading) and productive (speaking and writing) skills. With the PAP approach, teachers can develop clear and detailed rubrics for each of these skills, such as clarity of syntactic structure, mastery of mufradat, and fluency of articulation. Thus, the evaluation results can really be used to make pedagogical decisions appropriately (Marsila, Sosri, Rahayu, Vreticia, & KOTO, 2025)

In addition to validity, PB also has the advantage of reliability, which is the consistency of measurement results. Because PAP uses fixed and standardized assessment criteria, evaluation results can be more stable and unaffected by variations in student group performance. Teachers can assess students from year to year using the same benchmarks, allowing for meaningful longitudinal evaluation. This is important in Arabic language learning which is progressive in nature and demands the strengthening of competencies over time, especially in the mastery of qawā'id and the ability to read complex texts (Haqiqi, 2025).

Furthermore, the PAP approach strongly supports the implementation of formative assessment, where evaluation results are not only used to give final grades, but also as a tool for reflection and learning improvement. Teachers can identify specific aspects of competence that students have not achieved and design targeted learning interventions such as remedial, enrichment or individualized assistance. In PAN, this is difficult to do because the focus of the evaluation is comparison between individuals, not analysis of learning success indicators (Aldi, Khairanis, Lestari, & Trinova, 2025).

In a world of education that increasingly emphasizes accountability and data-driven instruction, PB makes a major contribution in providing data that is accurate, actionable and appropriate to the context of learning objectives. PAN-based evaluation may still be useful in selection or ranking, but in terms of developing the quality of Arabic language learning that is fair, progressive and purposeful, PB is much more strategic. It not only measures ability, but also strengthens the teaching system itself as it provides constructive feedback for teachers and students.

3. Psychopedagogical Implications for Student Motivation and Learning Orientation

The analysis also found that PAN and PAP approaches have different psychopedagogical impacts on learners. PAN often fosters a negatively competitive learning climate, where students are motivated to beat their classmates rather than develop themselves intrinsically (Purnamaningsih & Purbangkara, 2022). This can demotivate students who feel 'left behind' and reinforce academic inequality. In contrast, PAP promotes mastery-goal-based learning motivation, as students are encouraged to achieve certain criteria that can be pursued through personal effort. Recent research in educational psychology confirms that evaluation approaches that focus on individual development are more effective in building academic resilience and long-term motivation. In the context of Arabic language learning, which requires patience and continuous practice, PAP is more supportive of students' long-term success.

Table 3. Psychopedagogical Implications for Student Motivation and Learning Orientation

No	Evaluation Aspect	Norm-referenced Assessment (PAN)	Benchmark Assessment (PAP)	Critical Analysis
1	Learning Motivation	Dominated by extrinsic motivation (competition, ranking)	Encourage intrinsic motivation (personal progress, goal achievement)	PAP fosters growth-based learning rather than competition between individuals.
2	Psychological Impact	Prone to anxiety, distress and a sense of failure	Fosters self-confidence and a sense of competence	PAN has the potential to create academic pressure; LAP builds a mentally healthy learning experience.
3	Perception of Failure	Failure is perceived as falling behind others	Failure is perceived as feedback for improvement	PAP changes the paradigm of failure to become part of the learning process. Teacher Intervention
4	Teacher Intervention	Limited because results are classificatory	Can be adaptive and personalized	PAP provides more specific diagnostic data for learning decisions.
5	Self-esteem building	Academic Tends to be damaging for low-achieving students	Improves as it focuses on effort and process	LAP is more inclusive of students with diverse abilities. Learning Orientation
6	Learning Orientation	Outcome- and grade-oriented	Process, mastery and continuous improvement oriented	PB is in line with 21st century learning approaches and the Merdeka Curriculum.
7	Compatibility with Arabic Language Learning	Less suitable as comparative nature does not accommodate gradual progress	Very suitable as it supports tiered and continuous learning	PB facilitates thematic, integrative and process-based approaches to Arabic language learning.

From the table above we can draw a red thread that Benchmark Assessment (PAP) is significantly superior to Norm-based Assessment (PAN) in supporting the motivational and psychological aspects of

student learning. PAP encourages intrinsic motivation, self-confidence, and positive perceptions of the learning process through an orientation towards achieving objective and measurable standards. Meanwhile, PAN tends to create psychological pressure, anxiety and social stratification in the classroom due to the emphasis on comparisons between individuals. The pedagogical implications of these findings confirm that PAP is not only more equitable and inclusive, but also more effective in creating an emotionally healthy learning environment and supporting long-term academic growth, particularly in the context of Arabic language learning which demands perseverance and gradual progress.

One important dimension of learning evaluation that is often overlooked is the psychological aspect of students, especially related to learning motivation. The benchmark assessment approach has been shown to be more constructive in shaping intrinsic motivation as students are oriented towards achieving specific and realistic learning goals rather than competing with classmates. When students know that their success is measured against a predetermined standard, rather than compared to the performance of others, self-confidence and learning autonomy increase. This is in line with Deci & Ryan's autodetermination theory of motivation which emphasizes the importance of autonomy, competence and relatedness in creating healthy motivation.

Norm-referenced assessment (PAN), on the other hand, can lead to academic anxiety and unhealthy psychological pressure. In the context of PAN, students who are below average will feel like failures even though they may have made significant individual improvements. PAN reinforces hierarchy and social stratification within the classroom, creating labels of "smart" and "not smart" students, which in turn erodes the self-esteem and interest in learning of lower-middle-achieving students. Pedagogically, this contradicts the principles of inclusivity and differentiation of learning promoted in modern education, especially in the implementation of the Merdeka Curriculum (Anwar, 2023).

PAP also allows teachers to conduct formative assessments that are adaptive to students' psychological needs. When assessment results show that students have

not achieved certain targets, teachers can provide specific feedback and alternative learning strategies without making students feel defeated or inferior. This approach fosters a culture of improvement-based learning and lifelong learning, not merely achieving the best rank. In the context of Arabic language learning, which demands consistency and perseverance in skill acquisition, stable and positive motivation is indispensable.

From an educational psychology perspective, PAN can create a negative effect in the form of learned helplessness - a condition where students feel unable to catch up because they are always at the bottom of the rankings. This not only hampers academic performance, but also damages students' mental health in the long run. Students can feel that their efforts are in vain because the grading system does not allow for recognition of the learning process. PAP, on the other hand, values the process of student progress and rewards effort, not just competitive results.

Motivation to learn through PAP also tends to be more sustainable because it is underpinned by clear and achievable learning objectives. In recent studies on mastery-oriented assessment, students who are evaluated based on achievement standards tend to have higher learning endurance and show a willingness to improve continuously. This is particularly crucial in Arabic language learning, which requires gradual mastery of phonetics, sentence structure, and understanding of complex classical texts. With PAP, students do not feel chased by social comparison, but rather by the drive to be better than they were before.

The most important pedagogical implication of these findings is that teachers need to design evaluation systems that not only measure learning outcomes, but also build emotionally healthy and academically motivating learning experiences. PB supports this through its emphasis on growth, process, and objectively measurable outcomes. As such, it is not only more academically fair, but also more psychologically ethical. This makes it a relevant and strategic approach in contemporary Arabic education, particularly in building a generation of learners who are not only intelligent, but also resilient and mentally healthy.

IV. CONCLUSION AND SUGGESTION

A. Conclusion

Based on the results of the study and critical analysis of Norm-Based Assessment (PAN) and Benchmark-Based Assessment (PB), it can be concluded that the evaluation approach has a significant influence on the fairness, accuracy and quality of student learning outcomes, especially in the context of Arabic language learning. PB proves to be conceptually and implementatively superior because it assesses students based on the achievement of predetermined learning objectives in an objective and standardized manner. This approach is able to guarantee construct validity, consistency between indicators, and provide meaningful feedback that can be followed up in the learning process. In contrast, relative and comparative PAN is prone to classification bias, does not reflect students' real competencies, and is less adaptive to differentiated learning needs.

Furthermore, from a pedagogical and psychological perspective, PB has a positive impact on students' motivation to learn and mental health. By emphasizing individual progress and standards-based achievement, PB supports the development of intrinsic motivation, self-confidence, and a desire for continuous improvement. It is also more inclusive and ethical as it avoids the social labeling common in the PAN system. In the era of 21st century learning and the implementation of the Merdeka Curriculum, PAP offers a more contextual, transformative, and relevant evaluative approach to encourage meaningful, sustainable, and mastery-oriented Arabic language learning. Thus, the transition from PAN to PAP is an urgent need to create an evaluation system that is fair, humanist, and based on learning objectives.

B. Suggestion

This research makes an important contribution by critically revealing the fundamental differences between Norm-Based Assessment (PAN) and Benchmark-Based Assessment (PAP) in the context of Arabic learning evaluation, especially in terms of fairness, accuracy, and psychopedagogical impact. The strength of the research lies in the systematic comparative approach and the integration of motivational aspects and learning psychology, which have so far received less attention in educational

evaluation. However, this study also has limitations, such as the focus on a relatively limited population and location, and the use of instruments that may not cover all complex psychological variables. For the future, further studies should involve a more diverse sample and adopt mixed-methods methods to enrich the quantitative data with in-depth qualitative findings. In addition, the development of a more holistic and contextualized evaluation instrument will strengthen the validity of the results as well as the relevance of the application in various Arabic language education settings.

REFERENCES

Aldi, M., & Khairanis, R. (2025). History and Culture of Minangkabau in Educational Perspective: Integrating Traditional Values for Character Development. *Jurnal Penelitian Ilmu-Ilmu Sosial*, *6*(1), 52–67.

Aldi, M., Khairanis, R., Lestari, A. D., & Trinova, Z. (2025). Effectiveness of 5E Cycle Learning Model Assisted by Wordwall Media in Improving Critical Thinking Ability. *JIIP-Jurnal Ilmiah Ilmu Pendidikan*, *8*(5), 5373–5382.

Aldi, M., Supriyatno, T., Trinova, Z., & Muslimin, I. (2025). Resistance and Adaptation of Islamic Education in the Discourse of Islamic Ideology Language: A Socio-Historical Study of Old Order and New Order Politics: Resistensi dan Adaptasi Pendidikan Islam dalam Wacana Bahasa Ideologi Islam: Kajian Sosio-Historis Politik Orde Lama dan Orde Baru. *ELOQUENCE: Journal of Foreign Language*, *4*(1), 29–42.

Amal, I. (2024). Integrasi Pendekatan Acuan Patokan (PAP) dan Pendekatan Acuan Normatif (PAN) dalam Konteks Penilaian Pembelajaran Sekolah. *Ideguru: Jurnal Karya Ilmiah Guru*, *9*(3), 1373–1380.

Anwar, R. (2023). *Bimbingan Klasikal Hots Dan Tpack Dalam Kurikulum Merdeka: Suatu Pendekatan Best Practice*. Feniks Muda Sejahtera.

Atin, S. (2024). BAB 14 PENILAIAN ACUAN PATOKAN DAN PENILAIAN ACUAN NORMA. *Evaluasi Pembelajaran*, 222.

Azhar, M., Rahmawati, M., Saputra, M. R., Mulyani, R., Nurdinah, S., Frananda, A., & Fitri, L. (2025). Evaluasi Pembelajaran Bahasa Arab Berbasis Teknologi Digital. *Ukazh: Journal of Arabic Studies*, *6*(1), 78–99.

Busnawir, M. S., Judijanto, L., Abdullah, G., Abdurahman, A., Lumbu, M. D. A., Pd, Z. M., … Mahliatussikah, H. (n.d.). *EVALUASI PEMBELAJARAN*.

Haqiqi, I. H. (2025). *PENGARUH LITERASI DIGITAL DAN PENGUASAAN INFORMASI KEISLAMAN TERHADAP HASIL BELAJAR KOGNITIF PENDIDIKAN AGAMA ISLAM PADA SISWA SD ISLAM DARUL HUDA SEMARANG*.

IRAWAN, A. N. B. (2025). *MODEL EVALUASI PEMBELAJARAN TAHFIZ SDIT AL-BAROKAH TAHFIZ SCHOOL JAKARTA SELATAN*.

Irawan, F., Hanip, R., & Jasnur, P. (2025). *EVALUASI PEMBELAJARAN*. Penerbit Widina.

Lamusiah, S., Al Absin, A., Muslimin, I., Nufus, H., Nurfaidah, N., Jenab, S. H., … Akbar, R. M. (2025). Pendampingan literasi Al-Qur'an dalam kurikulum pendidikan dan pemahaman antarbudaya di SMAN 1 Labuapi. *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, *9*(2), 763–768.

Lestari, S. (2025). *Pengembangan Modul Interaktif Pada Materi Akhlak Mapel Pai Kelas V di Sekolah Dasar Negeri Sungapan Kokap Kulon Progo*.

Marsila, S., Sosri, L., Rahayu, P. W., Vreticia, E., & KOTO, I. (2025). ANALISIS PERBANDINGAN SISTEM PENDIDIKAN DI INDONESIA, UNI EMIRAT ARAB, DAN SINGAPURA DALAM INOVASI PEMBELAJARAN SAINS. *TEACHING: Jurnal Inovasi Keguruan Dan Ilmu Pendidikan*, *5*(1), 54–64.

Mubarak, A. F., & Rahman, F. (2025). IMPLEMENTASI PENDIDIKAN MULTIKULTURAL DALAM MENINGKATKAN SIKAP TOLERANSI SISWA DI MAN 2 PROBOLINGGO. *AL-MUADDIB: Jurnal Kajian Ilmu Kependidikan*, *7*(2), 527–542.

Nurmalina, N., Raudatussa'adah, R., & Hadijaya, Y. (2025). PELATIHAN MANAJEMEN PENDIDIKAN MULTIKULTURAL UNTUK PENDIDIK DALAM MEMBANGUN HARMONI DI SEKOLAH. *JPMA-Jurnal*

Pengabdian Masyarakat As-Salam, *5*(1), 12–21.

Pringgar, R. F., & Sujatmiko, B. (2020). Penelitian kepustakaan (library research) modul pembelajaran berbasis augmented reality pada pembelajaran siswa. *IT-Edu: Jurnal Information Technology and Education*, *5*(01), 317–329.

Purnamaningsih, I. R., & Purbangkara, T. (2022). *Pengelolaan sarana dan prasarana pendidikan dalam meningkatkan kualitas pembelajaran*. uwais inspirasi indonesia.

Raynaldi, R. (2025). *Penerapan Pembelajaran Model Quantum Teaching; Dalam Meningkatkan Hasil Belajar IPS Kelas VIII SMP; Negeri 2 Mattiro Sompe Kabupaten Pinrang*.

Rinantanti, Y. (n.d.). BAB 5 EVALUASI HASIL BELAJAR: TEKNIK, ALAT DAN PENDEKATANNYA. *EVALUASI PENDIDIKAN*, *9*(1), 85.

Rofii, A., Yuniarti, L., & Wahyuni, U. (2025). PENGEMBANGAN BUKU TEKS PADA MATERI PROSES MORFOLOGIS BAHASA INDONESIA BERBASIS CONTEXTUAL TEACHING AND LEARNING (CTL). *Aksara: Jurnal Ilmiah Pendidikan Bahasa Dan Sastra Indonesia*, *9*(1), 1–13.

Rosalinda, S. P. I. (2024). BAB 18 TEKNIK PENGOLAHAN HASIL EVALUASI. *Evaluasi Pembelajaran*, 289.

Sofa, A. R., Muarrifah, S., Hanafi, H., Parawansah, S. H., & Nurhamsalim, M. (2025). Penilaian Pembelajaran PBA Berbasis Google Data Studio Power Bi, Table AI dan Phyiton AI di MTs Thoyyib Hasyim Jorongan Probolinggo. *Indonesian Research Journal on Education*, *5*(3), 107–116.

ريتسفا خيرانس &محمد ألدي) .2025 .(العلاقة بين إتقان المفردات وفهم القراءة في معهد التربية الإسلامية بايور. *Shibghoh: Prosiding Ilmu Kependidikan UNIDA Gontor*, *3*(1), 1209–1222.