



## Interpreting Test Outcomes: A Comparative Study of Norm-Referenced and Criterion-Referenced Assessments

**Herdah<sup>1\*</sup>; Ainul Uyun<sup>2</sup>; Raodhatul Jannah<sup>3</sup>**

*Department of Tarbiyah, IAIN Parepare, Indonesia<sup>1\*</sup>, MA Nurul As'adiyah Callaccu, Indonesia<sup>2</sup>,*

*Department of Tarbiyah IAIN Parepare, Indonesia<sup>3</sup>*

\*Corresponding Author. E-mail: [herdah@iainpare.ac.id](mailto:herdah@iainpare.ac.id)<sup>1</sup>

**Receive: 11/08/2024**

**Accepted: 11/09/2024**

**Published: 01/10/2024**

### Abstract

*Learning is a dynamic process involving interaction among students, educators, and learning resources within a structured environment aimed at enhancing understanding and competence. Beyond knowledge transfer, educators play a crucial role in preparing students for learning by assessing situational readiness and optimizing learning conditions. Post-learning, evaluations are imperative to gauge goal attainment and guide improvement strategies, employing measurement and assessment techniques. Student capabilities and teaching efficacy benefit from evaluation, with assessments extending to norm-referenced (NRA) and criterion-referenced (CRA) methodologies. The NRA compares individual achievements against group norms, facilitating performance ranking within a cohort. Conversely, CRA utilizes predetermined benchmarks to determine academic success. This literature review synthesizes prior studies on NRA and CRA methodologies. Noteworthy studies include processing techniques discussed by Alfath and Raharjo, implementation insights by Aliyam, Dayanti, and Mukaffa, and completeness criteria exploration by Pangastuti and Munfa'ati. The review employs a descriptive analysis approach, utilizing educational and curriculum literature to inform its findings. The study concludes that NRA offers comparative insights into student performance within a group context, while CRA sets clear benchmarks for academic achievement. This study underscores the significance of evaluation in educational settings, advocating for effective evaluation tools to optimize learning strategies and enhance student engagement.*

**Keywords:** *test interpretation, evaluation, norm-referenced assessment, criterion-referenced assessment.*

### Introduction

Learning is a process that involves the interaction of students, educators, and learning resources within the learning environment. This process aims to transform students' understanding from a state of incapacity to one of ability, facilitating an exchange of information between them and their educators. In addition to conveying knowledge, educators also play a role in conditioning students to be ready to learn, in this case by paying attention to the situation and conditions before learning activities

begin, as well as the facilities and infrastructure used. (Syah & Pertiwi, 2024)

After carrying out the learning, educators certainly do not stop; there needs to be a learning evaluation that is carried out to see how far the learning goals that have been set are achieved, and then follow-up is carried out in order to improve the quality of learning. Evaluation activities can be in the form of measurement and assessment. (Zeinoun et al., 2022)

The ability of students and the effectiveness of learning methods used by educators can be improved through evaluation. In addition to measurement and

assessment, another evaluation technique that can be used is assessment. (Juryatina & Amrin, 2021) Educators can also assess the success of a curriculum and learning techniques based on the results of the evaluation.

Learning evaluation is one of the components that is no less important than the learning process. In this case, evaluation is a process to collect, analyze and interpret information to determine the level of achievement of learning objectives. A good evaluation system will show an overview of the quality of learning so that the results of learning evaluation can help educators plan learning strategies. (Rahman et al., 2022) In addition, a good evaluation system can also motivate students to further improve their abilities and be more active in learning.

Teachers must be responsible for the learning implementation process, which means that the learning outcomes of students are also the responsibility of educators. This responsibility can be demonstrated through evaluation activities. The results of the learning evaluation must be accountable. Therefore, an educator must know and understand matters related to evaluation.

Implementing an evaluation requires a tool. A tool, in a general sense, is something that can be used to make it easier for a person to carry out a task or achieve a goal to be more effective and efficient. (Tes et al., 2020) This tool in learning evaluation is better known as an evaluation instrument.

An evaluation tool can be said to be good if it is able to evaluate something with results such as the actual state of the evaluation object. The evaluation tool is a measurement tool to assess by evaluating the extent to which the learning process is conveyed to students. (Yul et al., 2022) Evaluation tools or instruments that can be used in learning evaluation, namely in the form of tests and non-tests.

A test is a tool or procedure used to find out or measure something in various ways and rules that have been determined. Meanwhile, when associated with learning evaluation, a test is a method (that can be

used) or a procedure (that can be taken) in the context of measurement and assessment that can be in the form of assignments either in the form of questions or commands.

Tests that are often used in schools to measure learning outcomes are created by educators. Test questions used to monitor the success rate of learning activities are required to be able to provide careful (Mahmudi, 2018) and accurate information related to students' absorption and mastery of the learning material presented. (Ridho, 2018)

The test results are in the form of different scores for each student. Test scores are the key for educators to understand students' abilities, so it is very important to be able to interpret student test scores. (Poerwanti, 2015) The calculation of test scores is usually done as is. For example, in an Arabic language learning test, a student gets a score of 50. So, the question is what the meaning of the score of is 60. Whether the score indicates that the ability is below average, average, or above average. Therefore, there needs to be a reference when interpreting test scores. The reference will be a comparison of the abilities of each student. (Febriana, 2021) The reference is known as the norm reference assessment (NRA) or the criterion reference assessment (CRA).

Research on norm reference assessment and criterion reference assessment is not the first time it has been conducted. There have been previous studies examining this issue. There are at least three parts of the study conducted by previous researchers. First, a study that discusses assessment processing techniques, such as in the article *Assessment Processing Techniques: Processing Techniques Using the Norm Reference Approach (NRA) and the Criterion Reference Approach (CRA)* written by Khairuddin Alfath and Fajar Fauzi Raharjo. This article was published in the *Al-Manar Journal: Journal of Islamic Communication and Education* in 2019 whose focus is on the techniques used in

processing test results, both NRA and CRA.(Alfath, 2019)

The second part is a study that discusses the implementation of the CRA and NRA approaches, such as in the article Implementation of the Normative Reference Assessment Approach (NRA) and the Criterion Reference Assessment Approach (CRA) written by Dewi Nur Annisa Aliyam, Fifin Dayanti, and Zumrotul Mukaffa. This article was published in Jurnal At-Ta'lim: Journal of Islamic Education Information Media in 2021 whose focus is on the implementation of approaches in assessment, both norm reference assessments and criterion references.(Annisa et al., 2021) The third part is an article that discusses the assessment of norms and benchmarks, as well as the criteria for completeness, such as in the article entitled Norm Reference Assessment, Benchmark Reference Assessment, Minimum Completeness Criteria at Madrasah Ibtidaiyah An-Nur Plus Junwangi Krian Sidoarjo East Java written by Ratna Pangastuti and Kusnul Munfa'ati. This article was published in Jurnal Tarbiyah Al-Awlad: Journal of Islamic Education in 2018. The focus of this article is to interpret the level of students' knowledge of learning. (Pangastuti & Munfa'ati, 2018)

This study seeks to present a norm reference assessment and a criterion reference assessment, as well as the interpretation of the two results using a library research approach. The study carried out is to process the data that has been collected in the form of books, journals, texts related to the research theme as reference materials.

## Method

This research uses the method of literature review. Library research is research that uses literature and reading sources in the form of books, texts, encyclopedias, journals, and other research results in the field of education and curriculum as the source. The literature

studied consists of references in the field of evaluation, higher education curriculum, and literature on learning Arabic. This research method is also often referred to as descriptive analysis, which is a research method that leads to a conclusion from the results of a study based on document analysis. The descriptive method is research that seeks to describe systematically, factually and accurately about a phenomenon, symptom, event, and event that is happening now. The descriptive method of analysis with a qualitative approach is a method that aims to make a descriptive, systematic, factual and accurate description of a phenomenon, event, event that occurs today In other words, the descriptive method of analysis is a method that formulates itself rather than solving problems that exist in the present where data is compiled and explained then analyzed. Based on the description above, the method used in this study is a descriptive analysis method, namely by collecting data, analyzing and drawing conclusions to solve the problems that exist at the time of the research.

## Result and Discussion

### 1. Norm Reference Assessment (NRA)

Norm reference assessment is an assessment that compares the learning outcomes of each student in a certain group. In this case, it compares the scores obtained by students with relative standards or norms. The norm in question shows the capacity or achievement of the group of students.(Sa'i & Anwar, 2023)

For example, a student if he is in group A may get a high predicate, but if he is in group B, he may only get a moderate predicate. The meaning of the value in the form of the number obtained has a relative nature, which means that the assessment reference is only useful for certain groups.

The characteristics of the Norm Reference Assessment (NRA) are as follows:

- a. The norm reference assessment is used in determining the status of each student against the abilities of other students. Therefore, this assessment is used to determine the ability of students in a group such as in class, school, and so on.
- b. The assessment of norm reference uses relative criteria, which means that the assessment is always changing according to the conditions and or needs at that time.
- c. The value of the results of the norm reference assessment does not reflect the level of ability and mastery of students related to the learning material being studied, but only shows the position of the students, in this case their ranking in their group or a certain class.
- d. Norm reference assessments have a tendency to use a range of levels of mastery over a person's group, ranging from those who are very special to those who experience quite serious difficulties.
- e. The norm reference assessment provides a score that describes the mastery of the group.(Rohmah, 2024)

The assessment of norm reference is based on several important assumptions. First, in a heterogeneous student population—including differences in gender, educational background, social status of parents, and environment—groups with varying qualifications will be formed, ranging from good, moderate, to poor. This assumption indicates that in any learning outcome assessment activity, the majority of students will have scores that are close to average or in the middle of the distribution, while only a few participants will have very high or very low scores.

Second, the purpose of learning outcome evaluation is to determine the relative position of each test taker, whether they are above, in the middle, or below. This assessment is known as a relative standard-based assessment, which places each

participant's score in the context of the group's overall performance.(Kadir, 2015) This approach is often used in summative tests, such as end-of-semester exams, to provide a more accurate picture of the relative position of learners in the assessed population.

Norm-Referenced Assessment (NRA) possesses several advantages and disadvantages that are crucial to consider within the educational context. One significant advantage is its ability to assign grades maximally, as it evaluates student performance by comparing it to the performance of a specific group. This ensures that each student receives a grade that most accurately reflects their performance relative to their peers. Furthermore, NRA is effective in differentiating student abilities, using relative standards to clearly identify those who have mastered the learning material and those who have not. This differentiation is instrumental in designing more targeted educational interventions.

Another strength of NRA is its flexibility, as it can be adapted to various conditions and educational contexts, making its application feasible across different learning settings and curricula. The assessment process is simplified as it does not adhere to rigid benchmarks, allowing educators to easily allocate grades based on the performance distribution within the group. Additionally, NRA is comprehensive in its ability to evaluate cognitive, affective, and psychomotor domains, making it a robust tool for assessing various aspects of student development and achievement beyond mere academic capability.

However, NRA also has notable drawbacks. One limitation is its insufficient emphasis on specific competencies that students have mastered, focusing more on relative positioning rather than the attainment of predetermined competencies outlined in the curriculum.(Yusuf, 2017) Moreover, student rankings in NRA are influenced not only by individual performance but also by the performance of

other students in the same group. This can result in situations where high-performing students are ranked lower if the majority of the group also performs well, and vice versa. Additionally, in the context of NRA, students who fail often find it challenging to receive opportunities for improvement, as the relative assessment against their peers can hinder the provision of constructive and beneficial feedback for future improvement.

By understanding these advantages and disadvantages, educators and policymakers can make more informed decisions in implementing NRA to optimize student learning outcomes.

## 2. Criterion Reference Assessment (CRA)

Criterion Reference Assessment (CRA) is often also called Benchmark Reference Assessment (BRA) which refers to a certain criterion that has been determined beforehand. CRA is a way to determine student graduation using a number of benchmarks/criteria. Students who meet these benchmarks can be declared successful. Meanwhile, students who have not met the benchmarks or criteria can be declared failed or have not succeeded. The values obtained by students are related to the level of achievement of students' mastery of learning materials in accordance with the criteria that have been set. (Kaukab & Rahman, 2021)

The assessment of the reference criteria assesses the test results of individual students, not comparing a student with his group, but using a certain criterion. The criteria in question are a level of learning experience that is expected to be achieved after the completion of learning activities or a number of basic competencies that have been determined in advance before learning activities. For example, the criteria used are 70 or 80, students whose abilities are below these criteria are declared unsuccessful and must do remedial.

Criterion-Referenced Assessment (CRA) is characterized by several distinct features that set it apart from other forms of evaluation. Firstly, CRA compares an

individual's performance against specific, absolute criteria rather than against the performance of other individuals. (Fadhilah, 2021) This approach ensures that the assessment is focused on whether each student has achieved the predefined learning objectives, providing a clear measure of mastery.

A key aspect of CRA is the careful definition of knowledge or content areas to ensure the interpretation of results is meaningful and beneficial. This meticulous delineation of content helps to ensure that assessments are aligned with educational goals and standards, making the outcomes more reliable and valid for educational decision-making.

Typically, CRA employs a narrow and limited scope of material. By focusing on specific areas of content, the assessment can delve deeper into those areas, allowing for a more thorough evaluation of student understanding and skills within the defined domain. This specificity ensures that the assessment is directly relevant to the instructional content.

Moreover, CRA assessments usually contain a large number of test items. This extensive item pool allows for comprehensive coverage of the content area, increasing the reliability of the assessment by reducing the impact of any single item on the overall score.

The test items in CRA are designed to encompass a range of difficulty levels appropriate to the content being assessed. This ensures that the assessment can accurately measure different levels of student proficiency, from basic understanding to advanced application. (JANNAH et al., 2021)(Jannah et al., 2024)

For instance, if a test result indicates that a student has an 80% correct score, it means that the student has successfully answered 80% of the test items correctly, demonstrating a specific level of mastery over the tested material. This straightforward interpretation helps educators identify which students have met

the learning objectives and which areas may require further instruction.

By focusing on these characteristics, CRA provides a detailed and precise measure of student achievement in relation to specific educational standards, making it a valuable tool for informing instruction and improving learning outcomes.

CRA offers several advantages and disadvantages that are crucial for educators to consider. One significant advantage is its ability to aid educators in designing remedial programs. By clearly identifying which specific learning objectives students have not mastered, teachers can create targeted interventions to address those gaps. Additionally, CRA does not require complex statistical methods, making it simpler to implement compared to Norm-Referenced Assessment (NRA). This simplicity is beneficial for educators who may not have extensive training in statistical analysis.

CRA is also effective in measuring the achievement of learning objectives, providing a direct assessment of whether students have met specific educational goals. The scores in CRA are fixed and absolute, ensuring that they remain consistent over time and across different cohorts of students. This reliability makes CRA a valuable tool for evaluating student performance in subjects where the learning material consists primarily of concepts, as it allows for precise measurement of understanding. Moreover, the presence of clear criteria simplifies the grading process, as educators have a concrete standard against which to measure student performance.

However, CRA also has certain limitations. It is not typically used for processing or determining scores in summative assessments, where the goal is often to rank students rather than to evaluate specific competencies. Another potential drawback is that if the test items are too difficult, students may receive low scores despite having a reasonable understanding of the material. Conversely, if the test items are too easy, students may achieve high scores even if they have not fully mastered the

content. This can lead to an inaccurate representation of student learning and may fail to identify areas where further instruction is needed.

Understanding these advantages and disadvantages helps educators and policymakers make informed decisions about when and how to use CRA to effectively measure and support student learning outcomes.

### 3. Interpretation of NRA and CRA Test Results

Norm Reference Assessment (NRA) is an assessment that interprets the results of measurements by comparing the learning outcomes of students with the learning outcomes of other students in their group, so that the position or ranking of a student in their group can be known. In other words, the determination of the score refers to the acquisition of scores in that group as a benchmark and is relative.

An example of an Arabic subject test score in the table 1:

Table 1. Example of Arabic Subject Test Scores

No	Name	Multiple-choice	Fill in the blank	Essay
1	Amanda	80	78	81
2	Anugrah	79	80	82
3	Asmi	78	80	80
4	Asrul	77	78	80
5	Fajar	80	79	81
6	Farhan	81	80	80
7	Ismail	78	80	80
8	Meli	80	78	80
9	Sherina	78	79	78
10	Yusuf	78	80	81

The process involves several steps to analyze data using Microsoft Excel. First, the data is inputted into Microsoft Excel, typically organized in a table format for clarity and ease of manipulation. Once the data is entered, the next step is to calculate the average score, which provides a central measure of the data set's tendency. This

average score is computed using the mean function in Excel, which sums up all scores and divides by the number of observations.

After determining the mean, the next statistical measure often calculated is the standard deviation. This statistic measures the amount of variation or dispersion in the data set. In Excel, the standard deviation is computed using the STDEV function, providing insight into how spread out the scores are around the mean.

In educational settings, scores are often converted into letter grades to provide a clearer understanding of performance. This conversion process assigns letters (such as A, B, C, etc.) based on predefined score ranges, allowing educators and students to easily interpret and communicate academic achievement. As seen in the table 2:

Table 2. Conversion of Scores to Letter Grades in Arabic Subject Test

Name	Multiple-choice	Fill in the blank	Essay	Average	Grade
Amanda	80	78	81	79,67	C
Anugrah	79	80	82	80,33	B
Asmi	78	80	80	79,33	C
Asrul	77	78	80	78,33	C
Fajar	80	79	81	80,00	B
Farhan	81	80	80	80,33	B
Ismail	78	80	80	79,33	C
Meli	80	78	80	79,33	C
Sherina	78	79	78	78,33	C
Yusuf	78	80	81	79,67	C

Based on the results of the test mentioned above, it can be concluded that 30% of students get a B score and the other 70% get a C score, besides that it can also be seen that there are two students who get the highest scores, namely Anugrah and Farhan with a score of 80.33 and there are two students who get the lowest scores, namely Asrul and Sherina with a score of 78.33.

The assessment of the reference criteria can be based on the criteria set in an educational unit, this assessment aims to measure exactly the competencies that are set as the criteria for success. The success or failure of a student is based on the criteria

that are not related to the competence of other students. Criterion-based assessments do not position rankings as important because the most important thing is that the student achieves minimum competency standards, for example 80.

### Conclusion

The norm reference assessment (NRA) is an assessment that compares the learning outcomes of each student in a certain group. In this case, it compares the scores obtained by students with relative standards or norms. The norm in question shows the capacity or achievement of the group of students.

CRA is a way to determine student graduation using a number of benchmarks/criteria. Students who meet these benchmarks can be declared successful. Meanwhile, students who have not met the benchmarks or criteria can be declared failed or have not succeeded.

Norm Reference Assessment (NRA) is an assessment that interprets the results of measurements by comparing the learning outcomes of students with the learning outcomes of other students in their group, so that the position or ranking of a student in their group can be known. The criterion reference assessment (CRA) assesses students based on the achievement of the minimum of completeness of mastery learning (KKM) in their educational unit.

Based on the conclusions of the research, teachers should select assessment methods according to learning objectives, using NRA to determine students' relative positions and CRA to ensure they meet minimum competency standards. Combining both methods offers a comprehensive view of student performance, informing tailored teaching strategies and interventions. Clear communication with students and parents, regular evaluation, and teacher training on assessment methods are essential for improving educational quality.

## References

- 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.
- Annisa, A. D. N., Amirudin, H., & Muzaki, I. A. (2022). PENGEMBANGAN KURIKULUM DALAM MATA PELAJARAN PENDIDIKAN AGAMA ISLAM PADA SEKOLAH MENENGAH KEJURUAN (SMK) MUHAMMADIYAH 2 CIKAMPEK. *PeTeKa*, 5(3), 521–530.
- Fadhilah, R. (2021). Pengembangan Penilaian Pengetahuan Bahasa Arab Berbasis Test Objektif di Pusat Pendidikan Islam Muharrirun Najaah. *Uktub: Journal of Arabic Studies*, 1(2), 134–142.
- Febriana, R. (2021). *Evaluasi pembelajaran*. Bumi Aksara.
- JANNAH, R., Darmawati, & Saepudin. (2021). Analisis Butir Soal Ujian Sekolah Berstandar Nasional (USBN) Mata Pelajaran Bahasa Arab Madrasah Tsanawiyah Kota Parepare. *International Conference on Islam, Law, and Society (INCOILS)*, 1(1), 130–137.
- Jannah, R., Herdah, H., & Muharram, S. (2024). *ARABIC MATERIAL DEVELOPMENT DESIGN BASED ON LOCAL*. 12(2), 6–9.
- Juryatina, J., & Amrin, A. (2021). Students' interest in Arabic language learning: the roles of teacher. *Journal of Educational Management and Instruction (JEMIN)*, 1(1), 40–49.
- Kadir, A. (2015). Menyusun dan menganalisis tes hasil belajar. *AL-TA'DIB: Jurnal Kajian Ilmu Kependidikan*, 8(2), 70–81.
- Kaukab, M. E., & Rahman, R. A. (2021). Penilaian Autentik dalam Pembelajaran Bahasa Arab. *NIVEDANA: Jurnal Komunikasi Dan Bahasa*, 2(1), 60–75.
- Mahmudi, A. (2018). The use of G-form as an assessment instrument in Arabic language teaching based on HOTS. *At-Tarbawi: Jurnal Kajian Kependidikan Islam*, 3(2).
- Pangastuti, R., & Munfa'ati, K. (2018). Penilaian Acuan Norma. *Penilaian Acuan Patokan, Kriteria Ketuntasan Minimal Di Madrasah Ibtidaiyah An-Nur Plus Junwangi Krian Sidoarjo Jawa Timur*, 210.
- Poerwanti, E. (2015). Konsep dasar asesmen pembelajaran. *PT. Remaja Rosdakarya*.
- Rahman, N. F., Masdar, M. I., Ilmiani, A. M., Habibie, F., & Samdouni, S. (2022). Transformation of Arabic Assessment in Indonesia: Conventional Assessment Toward Digital Assessment. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya*, 10(2), 199–212.
- Ridho, U. (2018). Evaluasi Dalam Pembelajaran Bahasa Arab. *An Nabighoh Jurnal Pendidikan Dan Pembelajaran Bahasa Arab*, 20(01), 19. <https://doi.org/10.32332/an-nabighoh.v20i01.1124>
- Rohmah, F. L. (2024). Pengolahan Hasil Belajar Peserta Didik. *Grata: Jurnal Inovasi Pendidikan*, 1(1), 37–43.
- Sa'i, M., & Anwar, C. (2023). Penerapan Penilaian Beracuan Norma Dan Penilaian Beracuan Kriteria Pada Mata Pelajaran Pendidikan Agama Islam Di SDN Ganding I Sumenep. *MUBTADI: Jurnal Pendidikan Ibtidaiyah*, 4(2), 177–185.
- Syah, M. E., & Pertiwi, D. S. (2024). *Psikologi belajar*. Feniks Muda Sejahtera.
- Tes, P., Dalam, D., & Arab, B. (2020). *AL-ISHLAH Jurnal Pendidikan Islam*.
- Yul, W., Andrian, R., Musthofa, A., & Rozianie, J. F. (2022). Assessing Arabic Speaking Skills: A Critical Study for Implementation of Final Examination in Indonesia. *Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 5(2), 151–166.
- Yusuf, A. M. (2017). *Asesmen dan evaluasi pendidikan*. Prenada Media.
- Zeinoun, P., Ilescu, D., & El Hakim, R. (2022). Psychological tests in Arabic: A review of methodological practices and recommendations for future use. *Neuropsychology Review*, 32(1), 1–19.