

**THE USE OF OBJECTIVE TEST INSTRUMENTS TO MEASURE ARABIC GRAMMAR MASTERY OF GRADE 8 AND 9 STUDENTS OF MTS AL-MADANIYAH JEMPONG**Annatul Islamiyah<sup>1</sup>, Safrina Dwiyani<sup>2</sup>, Siti Hajar<sup>3</sup>, Wirangga<sup>4</sup>, and Nasaruddin<sup>5</sup><sup>1</sup> Muhammadiyah University of Mataram, Mataram, Indonesia<sup>2</sup> Muhammadiyah University of Mataram, Mataram, Indonesia<sup>3</sup> Muhammadiyah University of Mataram, Mataram, Indonesia<sup>4</sup> Muhammadiyah University of Mataram, Mataram, Indonesia<sup>5</sup> Muhammadiyah University of Mataram, Mataram, Indonesia**Corresponding Author:**

Safrina Dwiyani,

Department of Arabic Language Education, Faculty of Islamic Studies, Muhammadiyah University of Mataram.

Email: [safriadwiani@gmail.com](mailto:safriadwiani@gmail.com)**Article Info**

Received: October 30, 2025

Revised: November 15, 2025

Accepted: November 30, 2025

Online Version: December 15, 2025

**Abstract**

This study aims to develop a valid and reliable objective test instrument to measure Nahwu (Arabic grammar) mastery in grades 8 and 9 of MTs Al-Madaniyah Jempong. Nahwu mastery is a fundamental competency in Arabic language learning, but the availability of standardized evaluation instruments is still limited. This study used the Research and Development (R&D) method which includes the stages of instrument development, expert validation, and empirical trials on 15 students. The instrument was a multiple-choice test consisting of 25 questions arranged based on the Nahwu material grid. Content and construct validity were evaluated by experts, while empirical validity was analyzed using Pearson Product Moment correlation. The results showed that 22 items were declared valid and 3 items were invalid. The reliability of the instrument calculated using Cronbach's Alpha produced a value of 0.852, which indicates very high internal consistency. In addition, the Independent Sample t-Test test showed that there was no significant difference between the average scores of 8th and 9th grade students ( $t_{\text{count}} = -1.497 < t_{\text{table}} = 2.160$ ), so the instrument was considered capable of measuring the abilities of both groups in a balanced and unbiased manner. Overall, the objective test instrument developed was proven to be suitable for use as a standard evaluation tool in measuring Nahwu mastery at the MTs level and is recommended to be retested on a wider sample to strengthen its psychometric evidence.

**Keywords:** Instrument Development, Nahwu, Reliability, T-Test, Validity Test

© 2025 by the author(s)

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).

Journal Homepage <https://journal.zmsadra.or.id/index.php/jqa>  
How to cite: Islamiyah, I., Dwiyani, S., Hajar, S., Wirangga, Wirangga., & Nasaruddin, Nasaruddin. (2025). The Use of Objective Test Instruments to Measure Arabic Grammar Mastery of Grade 8 and 9 Students of MTs Al-Madaniyah Jempong. *Qaul ‘Arabiy*, 1(3), 131–136. <https://doi.org/XX.XXXXX/jqa.v1i3.1420>  
Published by: Yayasan Zia Mulla Sadra

---

## INTRODUCTION

Arabic is a compulsory subject in Islamic Junior High Schools (MTs). It encompasses four core skills, one of which is grammatical skills, or Nahwu. Mastery of Nahwu is a crucial foundation for understanding Arabic texts, including the Quran, Hadith, and other academic literature. Failure to master Nahwu thoroughly will result in students having difficulty grasping the meaning and structure of complex sentences.

Efforts to improve the quality of Nahwu learning must be balanced with an accurate evaluation system. The test instruments used must be valid and reliable to accurately reflect students' level of Nahwu mastery. Based on initial observations at MTs Al-Madaniyah Jempong, the evaluation instruments used tend to be conventional and have not undergone adequate psychometric testing. This results in inaccurate measurement results and makes it difficult for teachers to determine appropriate follow-up learning strategies.

Therefore, this research focuses on the development of an objective test instrument to measure the mastery of Nahwu (grammar) in Arabic at grades 8 and 9 of MTs Al-Madaniyah Jempong. This study aims to produce an objective test instrument that is standardized and statistically tested for quality, so that it can be used effectively to assess students' Nahwu competency at the MTs level.

## RESEARCH METHOD

This study uses the Research and Development (R&D) method, which aims to produce new products or improve existing products, namely objective test instruments to measure the level of mastery of Nahwu material. The study was conducted at MTs Al-Madaniyah Jempong with limited trial subjects totaling 15 students consisting of students in grades 8 and 9. The research instrument was in the form of a multiple-choice objective test consisting of 25 questions arranged based on the Nahwu material grid taught at that level, and equipped with a questionnaire or expert validation sheet to obtain data regarding content and construct validity from material experts and evaluation experts. Data collection techniques include the use of expert validation sheets to obtain qualitative data regarding the feasibility of the instrument, as well as tests to obtain quantitative data in the form of student trial scores. All data were analyzed qualitatively and quantitatively, namely through qualitative descriptive analysis to process input and suggestions from experts as a basis for revising the instrument, and quantitative statistical analysis to empirically test the quality of the test items, which included a validity test using Pearson Product Moment correlation, where an item was declared valid if the calculated  $r_{\text{value}} \geq r_{\text{table}}$  (with  $\alpha = 0.05$  and  $N = 15$ , the  $r_{\text{table}}$  value = 0.514), a reliability test using the Cronbach Alpha coefficient to assess the internal consistency of the instrument, as well as an analysis of the discriminatory power (D) and level of difficulty (P) as part of an additional evaluation to determine the optimal quality of the test items.

## RESULTS AND DISCUSSION

This study employed a Research and Development (R&D) method, adapting a development model that included instrument design, expert validation, and empirical testing to ensure the instrument's overall quality. The trial subjects were 15 students from grades 8 and 9 at MTs Al-Madaniyah. The instrument was an objective test consisting of 25 multiple-choice questions, which were then analyzed using Descriptive Analysis of Student Test Results to assess the instrument's ability to measure Nahwu mastery and identify the quality of each item. Based on the trial results, the average Nahwu mastery score was 50.13 out of a maximum score of 100, with a maximum score of 80, a minimum score of 40, and an average score of 50.13.

Empirical validity testing was conducted using the Pearson Product Moment Correlation technique with a significance level of 0.05 and a sample size of 15 subjects (N). The  $r_{table}$  value used as a reference was 0.514. The results of the analysis showed that 22 items were declared valid, namely P1, P2, P3, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20, P21, P22, and P23, with correlation coefficient values ( $r_{count}$ ) ranging from 0.514 to 0.864; while 3 items, namely P4, P24, and P25, were declared invalid because they had  $r_{count}$  values below  $r_{table}$  and significance values (Sig.) greater than 0.05.

Table 1. Validity Test Results

Question Number	r count (Pearson Correlation)	Sig.(2-tailed)	r table (N=15, $\alpha=0.05$ )	Interpretation
P1	864	0	514	Valid
P2	864	0	514	Valid
P3	864	0	514	Valid
<b>P4</b>	<b>179</b>	<b>532</b>	<b>514</b>	<b>Invalid</b>
P5	650	9	514	Valid
P6	745	1	514	Valid
P7	514	50	514	Valid
P8	812	0	514	Valid
P9	745	1	514	Valid
P10	674	6	514	Valid
P11	701	3	514	Valid
P12	626	13	514	Valid
P13	799	0	514	Valid
P14	812	0	514	Valid
P15	842	0	514	Valid
P16	601	20	514	Valid
P17	799	0	514	Valid
P18	799	0	514	Valid
P19	760	1	514	Valid
P20	548	35	514	Valid
P21	674	6	514	Valid
P22	595	22	514	Valid
P23	734	2	514	Valid
<b>P24</b>	<b>37</b>	<b>896</b>	<b>514</b>	<b>Invalid</b>
<b>P25</b>	<b>369</b>	<b>183</b>	<b>514</b>	<b>Invalid</b>

The results of the reliability test calculated using the Cronbach's Alpha coefficient for all 25 questions showed a reliability value of 0.852, which indicates that the instrument has high internal consistency and is suitable for use as a measuring tool for mastery of Nahwu.

Table 1. Reliability Test Results

Question scale	Alpha Cronbach	Interpretation
25	0,852	The instrument has high internal consistency and is suitable for use.

Furthermore, an Independent Sample t-Test was conducted to determine the differences in learning outcomes between 8th and 9th grade students in the Nahwu mastery test instrument trial. The average score of 8th grade students was 61 with a standard deviation of 17.07 ( $n = 8$ ), while the average score of 9th grade students was 72 with a standard deviation of 11.07 ( $n = 7$ ). The t-test calculation using the two-sample independent t-test formula produced a  $t_{count}$  value of -1.497 with a degree of freedom (df) of 13. This value was then compared with the  $t_{table}$  at a significance level of 0.05 (two-tailed), which was 2.160. Based on the testing criteria, because  $|t_{count}| < t_{table}$ , it can be concluded that there is no significant difference between the average scores of 8th and 9th grade students.

### Discussion of Item Validity Test Results

The results of the empirical validity test are a crucial step in determining the extent to which the test items are able to measure the intended construct, namely mastery of Nahwu (literacy). Of the 25 test items, 22 (88%) were found to be statistically valid.

#### Implications of Valid Items

The high percentage of valid items indicates that the majority of the developed items met the criteria for internal alignment.

1. Measurement Consistency: The significant correlation coefficient  $r_{count}$  (even reaching 0.864 for items P1, P2, and P3) indicates that these items are consistent with the total score. This means that students with high Nahwu skills tend to answer these items correctly, and vice versa.
2. Construct Representativeness: These valid items successfully represent Nahwu material according to the grid and are able to measure students' ability to apply Nahwu principles in the context of objective questions.
3. Analysis of Invalid Items (P4, P24, P25) The discovery of three invalid items (P4, P24, P25) requires in-depth review. These items have a very low correlation with the total score (P4: 0.179; P24: 0.037; P25: 0.369).

The invalidity of these items can be caused by several factors:

1. Content and Construction Factors: The Nahwu material tested in these items may not have been fully mastered by the students (for example, the material is outside the scope of the curriculum being effectively studied). Alternatively, the question wording may be ambiguous, or the answer key and distractors may not function optimally. For item P24, which has an  $r_{count}$  close to zero (0.037), it can be interpreted that students' answers on this item are random or unrelated to their Nahwu ability.
2. Student Response Factors: Although the items have been theoretically revised based on expert validation, in an empirical context, there is a possibility that students misinterpreted the meaning of the questions or that the items measured knowledge that was too easy/difficult, thus failing to distinguish between high and low groups.
3. As a follow-up, these 3 items must be eliminated or completely revised to avoid reducing the overall measurement quality of the instrument in the final version.

### Discussion of Reliability Test Results

Reliability testing was conducted to ensure that the developed test instrument had good internal consistency. Testing using the Cronbach's Alpha coefficient yielded a value of 0.852.

In the context of educational evaluation, a reliability coefficient of 0.852 is categorized as Very High (because it generally exceeds 0.80). This figure has important implications:

1. **Instrument Reliability:** This objective test instrument has a very good level of reliability (consistency) in measuring Nahwu mastery in grades 8 and 9 of MTs Al-Madaniyah Jempong. If this test were re-administered to a group of students with similar characteristics and conditions, the results would likely be highly consistent.
2. **Measurement Stability:** High reliability ensures that variations in student scores are largely due to differences in their actual Nahwu abilities (true scores), rather than random measurement error.
3. **Impact of Item Quality:** Despite the presence of 3 invalid items, the Cronbach's Alpha value remains high (0.852) indicating that the 22 valid items have contributed significantly to the consistency of the measurement. This strong reliability supports the use of the instrument for standardized evaluation purposes.

### Discussion of t-Test Results

The results of the Independent Sample t-Test were used to determine whether there was a difference in Nahwu mastery between 8th and 9th grade students. Descriptively, 8th grade students obtained an average score of 61 with a standard deviation of 17.07 ( $n = 8$ ), while 9th grade students had an average score of 72 with a standard deviation of 11.07 ( $n = 7$ ). Although there was an 11-point difference in the averages, the inferential analysis showed that this difference was not statistically significant.

The t-test calculation yielded a t-value of  $-1.497$  with 13 degrees of freedom (df). This value is smaller than the t-table value at the 0.05 significance level (two-tailed), which is 2.160. Since  $|t_{\text{count}}| < t_{\text{table}}$ , it can be concluded that there is no significant difference between the scores of the two classes. In other words, the Nahwu mastery abilities of 8th and 9th grade students were relatively equivalent in this instrument trial.

These results indicate that the developed test instrument is able to consistently measure student abilities across two groups at different grade levels, without exhibiting bias toward one group or the other. Furthermore, the homogeneity of ability across grades indicates that the Nahwu material tested is evenly distributed and can be measured equally across both grade levels.

## CONCLUSION

This research successfully developed an objective test instrument to measure the mastery of Nahwu (literacy) in eighth and ninth grade students at MTs Al-Madaniyah Jempong through a Research and Development (R&D) approach. The resulting instrument consisted of 25 multiple-choice questions structured based on the Nahwu material outline. Through expert validation and empirical testing, the instrument demonstrated excellent quality.

The empirical validity test results indicated that 22 of the 25 items were valid, thus accurately measuring students' Nahwu skills. Three invalid items (P4, P24, and P25) required revision or elimination to ensure the instrument's quality. Furthermore, the reliability test using Cronbach's Alpha showed a value of 0.852, indicating that the instrument has very high internal consistency and is suitable for use as a learning evaluation tool. Furthermore, the results of the Independent Sample t-Test showed no significant difference between the average scores of 8th and 9th grade students ( $t_{\text{count}} = -1.497 < t_{\text{table}} = 2.160$ ). This finding indicates that the Nahwu mastery abilities of both classes were relatively equivalent and that

the developed instrument was able to measure students' abilities consistently without bias between groups.

Based on these findings, it can be concluded that the developed objective test instrument has met theoretical and empirical quality standards. This instrument is not only suitable for use as a measure of students' Nahwu competency but can also serve as a reference for teachers in conducting more accurate learning evaluations. For further development, revisions to invalid items and further trials with a larger sample size are needed to enhance the instrument's psychometric strength.

## REFERENCES

- Arikunto, S. (2019). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Azwar, S. (2016). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2019). *Penyusunan Skala Psikologi (Edisi Ke-3)*. Yogyakarta: Pustaka Pelajar.
- Borg, W. R., & Gall, M. D. (1983). *Educational Research: An Introduction (4th ed.)*. New York: Longman.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to Design and Evaluate Research in Education (8th ed.)*. New York: McGraw-Hill.
- Ghazali, I. (2011). *Aplikasi Analisis Multivariate dengan Program IBM SPSS*. Semarang: Badan Penerbit Universitas Diponegoro.
- Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Parama Publishing.
- Nurgiantoro, B., Gunawan, & Marzuki. (2015). *Penilaian Pembelajaran Bahasa*. Yogyakarta: BPFE.
- Purwanto. (2010). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya.
- Sudijono, A. (2012). *Pengantar Evaluasi Pendidikan*. Jakarta: RajaGrafindo Persada
- Sugiyono. (2018). *Metode Penelitian & Pengembangan (Research and Development)*. Bandung: Alfabeta.
- Sukardi. (2013). *Evaluasi Pendidikan: Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara.
- Umar, H. (2014). *Desain Penelitian Evaluatif*. Jakarta: Rajawali Pers.
- Widarjo, M. (2016). *Statistik untuk Penelitian Pendidikan*. Malang: Universitas Negeri Malang Press.
- Yasin, A. (2012). *Nahwu Praktis untuk Pemula*. Jakarta: Pustaka Amani.

---

**Copyright Holder :**

© Annatul Islamiyah et.al (2025).

**First Publication Right :**

© Qaul ‘Arabiy

**This article is under:**

