



Validity and Reliability Analysis of Reading Comprehension Skill Test Using Islamic English Texts

Asa Bani Putra^{1✉}, Rudi Hartono², Muhammad Khumaedi³

^{1,2,3}Universitas Negeri Semarang, Indonesia

Article Info

History Articles

Received:

07 May 2023

Accepted:

09 June 2023

Published:

30 August 2023

Keywords:

Test instrument, reading comprehension, Islamic text, validity, reliability

Abstract

One of the requirements for graduation in Islamic universities is the ability to master English. Students encounter obstacles in fulfilling these requirements, especially non-English study program students. This research aims to prove the validity and estimate the reliability of the reading comprehension skills test instrument using Islamic English texts. The method used in this study was the research and development method proposed by Mardapi. Proof of content validity used the Aiken's V formula while estimation of reliability between raters used a two-way ANOVA test which was then calculated using Hoyt's formula. The research data shows that the Aiken V coefficient value for each statement item is more than 0.3. It can be said that all aspects are valid. The estimated reliability of the two-way ANOVA test yields a score of 0.94. It can be stated that this instrument is consistent. Thus, it can be declared that the test instrument for reading comprehension skills of Islamic English texts has good validity and reliability among raters is strong. The instrument test is expected to be used as a recommendation for the test instrument makers and help the students improve their English proficiency test score, especially in reading comprehension skill.

✉Correspondence Address :

Postgraduate, Universitas Negeri Semarang
Jalan Kelud Utara III, Kel. Petompon Kec. Gajah Mungkur
Kota Semarang 50237
E-mail : asabanip@students.unnes.ac.id

p-ISSN 2252-6420

e-ISSN 2503-1732

INTRODUCTION

English is a worldwide language that functions as a language that connects a person or a group with other individuals or groups around the world. Much of science and technology is written in English (Maghfirotillah et al., 2015). It affects users of English sources, for example academics and students. Sources of knowledge written in English can be in the form of books, articles, and magazines. Academics and students are required to be fluent in English both orally and in writing.

In education, English is the language of instruction covered in multidisciplinary science. English is the only foreign language that must be learned in school. At the higher education level, all study programs offer English courses for one or two semesters even though the subjects taken are not related to English. In addition to English courses, they are also required to pass an English proficiency test with a certain score. The form of the English language proficiency test at each university is different depending on the policies at the university. Many universities use the TOEFL (Test of English as a Foreign Language) form as a reference. TOEFL tests three English skills. They are listening, structure and written expression, and reading skills (*TOEFL ITP Test Content*, n.d.). Based on the observation done at the Center for International Language Development in an Islamic university, the test takers got low scores in reading comprehension skills. Reading in a foreign language like English is not as easy as reading in Indonesian.

Nugroho et al., (2018) stated that reading comprehension is the ability to read and process a text, as well as comprehend its meaning. A person's ability to comprehend a text is affected by its nature and skills, one of which is the ability to reason. If word identification is difficult, students will use too much of their processing power to read separate words, which reduces their ability to understand what is being read. Sukarni et al., (2017) states that foreign language reading

activities are often hampered by limited vocabulary knowledge, lack of fluency, use of ineffective reading strategies, and reading themes that are difficult and unfamiliar. Texts presented in reading skill tests that are usually given at universities or colleges in America use academic theme such as education, social, geography, history, biology, economics, culture, and so on. Guntur and Pordanjani (2019) argue that students have difficulties in understanding texts and that these difficulties vary due to different levels of grammar and vocabulary proficiency and that some students are incapable of critical thinking with a certain text. As a non-cognitive factor, motivation also affects reading performance from a psychological perspective. Those statements show that the development of reading comprehension skill test instruments for English texts is needed to improve English proficiency test results.

The difficulties experienced by examinees can be overcome by adjusting the English proficiency instrument test, especially reading skills. Test instrument makers can choose genre and themes that are more familiar to test takers so that participants can more easily understand the English text presented. The test takers can choose factual genre for the text presented in the test instrument. The factual genre is a genre that deals with real-world information, and it differs from other genres in how that information is managed (Wiratno, 2018). The right strategy is needed in choosing the appropriate English text theme, especially for Islamic university. One of the right solutions is to insert Islamic elements into the English text presented.

Article 3 of Law No. 20/2003 quoted from Munir and Hartono (2016) stipulates that the ultimate goal of national education is not only to produce excellent Indonesian graduates (who are academic excellence), but also to develop students' potential to be people who have faith in and fear the Almighty God, citizens of dignified character, healthy, capable, creative, independent, democratic, and reliable. It can be concluded

from the law that character education, one of which is related to religion, is also a goal of national education.

Several studies related to the development of reading comprehension skill test instruments in English texts have been carried out. One of them is the research conducted by Fathurohman and Cahyaningsih (2021) which focuses on developing an instrument for authentic assessment of English reading skills through metacognitive strategies. The research produced a valid, reliable, attainable authentic assessment instrument of English reading skills through metacognitive strategies. The instrument produced was in essay form, while this research will produce an instrument in the form of multiple choice. Another research was conducted by Suaidi et al., (2019). The research aims to develop a model test of reading comprehension skills for Islamic studies. However, this research is only limited to the preparation of test instruments and has not been tested on students. In addition to conducting trials, the test instrument must also prove its validity and estimate its reliability.

The validity and reliability of the test instrument are used to see the suitability of the test instrument. Validity is a mean to ensure whether the instruments used by researchers are valid and also ensure that the findings in research have been measured appropriately. One type of validity used in the initial stages of preparing the test instrument is content validity which is used to determine the extent to which the instrument actually covers all the classifications you want to see (Suwito, 2011). Meanwhile, the reliability test is a test to see how far an instrument can be trusted to measure a test (Susanto et al., 2018). This study used inter-rater reliability which is one of the important things in the development of test instruments. Reliability is used when an assessor or rater is engaged between two or more raters (Kusumastuti et al., 2020).

Therefore, this research aims to analyze the validity and reliability of the reading comprehension skill instrument test that has

been developed and tested by using Islamic English texts. It is hoped that this research can be a guide for the test instrument development team and can improve the results of the English proficiency test, especially in reading comprehension skills.

METHODS

The method used was the development research method proposed by Mardapi (2018). The development research design consists of 10 steps. They are (1) defining instrument specifications, (2) writing the instrument, (3) deciding instrument scales, (4) deciding scoring systems, (5) reviewing the instrument, (6) doing trials, (7) analyzing the instrument, (8) revising the instrument, 9) making and perfecting the instrument, and 10) evaluating the instrument.

Development research was carried out through steps such as arranging instrument specifications, instrument feasibility tests, and instrument trials. Before testing the instrument in the field, the instrument must first be tested by experts to determine the validity and reliability of the contents of the instrument to be used. The validity and reliability of the contents of the instrument were obtained by giving questionnaires with a score from 1 to 5 to the experts. The results of the questionnaires filled out were then analyzed. This step consisted of proving the validity and estimating the reliability of the instrument. It is consistent with the objective of this study, which is to test the validity and estimate the reliability of the test instrument for reading comprehension skills of Islamic English texts.

In this study, content validity was examined using the formula of the Aiken's V assessment. Aiken formulates the Aiken formula for calculating content validity coefficients based on the results of a panel of n experts on a given system. It is based on the degree to which the item signifies the structure examined (Hendryadi, 2017). Aiken (1985) introduces the concept of content validity that detail can be seen from the validity standard. It is influenced by the number of experts and

the rating scale used. The coefficient that meets the useable criteria is 0.30 (Azwar, 2016). The formula can be seen below:

$$V = \frac{\sum s}{n(c - 1)}$$

$$s = r - lo$$

lo = lowest validity rating score

c = highest validity rating score

r = number given by a validator

n = number of validators

Interrater reliability was estimated using two-way ANOVA test. This two-way ANOVA test is used to check the similarity of the assessment results between validators on the reading comprehension skills test instrument of Islamic English texts. There were three experts from Sultan Agung Islamic University Semarang to examine the instrument. They are a lecturer of English Education Study Program, a staff of Center for International Language Development, and a staff of Sultan Agung Waqf Board Foundation. The results of the study were then compared between one expert and another. Instrument reliability can also be estimated through the analysis of variance approach, one of which was proposed by Hoyt. The item scores in this case belong to a two-way factorial design without replication or treatment x subject design (Mardapi, 2018). Hoyt's formula can be formulated as:

$$r_{xx'} = 1 - \frac{s^2_r}{s^2_s}$$

s^2_r = residual variance which in the treatment x subject analysis is the mean square of the interaction between the item and the subject, called Mk_{is}

s^2_s = subject variance which is the mean squared between subjects called Mk_s

This study used 3 experts as validators consisting of 2 experts in the field of English and 1 expert in both English and Islamic fields. Data collection used a survey instrument using a Likert scale of 1 to 5. Expert assessment sheets were used to detect how conceptually and operationally relevant the developed assessment instrument is.

RESULTS AND DISCUSSION

This study intends to analyze the outcomes of proving the validity and estimating the reliability of the English text reading comprehension skills test instrument that has been developed by using Islamic English texts. Proving the validity and estimating the reliability of the instrument was accomplished before the instrument was tried out in the field. This study used the Aiken's V formula as a reference in proving the validity of the contents of the instrument. Table 1 shows the results of content validity using the Aiken's V formula.

Table 1. The Results of Content Validity Analysis Using Aiken's V Formula

Item	Aiken Coefficient	Criteria
1	0.58	Valid
2	0.67	Valid
3	0.67	Valid
4	0.83	Valid
5	0.75	Valid
6	0.75	Valid
7	0.75	Valid
8	0.67	Valid
9	0.75	Valid
10	0.92	Valid
11	0.92	Valid
12	0.92	Valid
13	0.92	Valid
14	0.75	Valid
15	0.92	Valid
16	0.67	Valid
17	0.75	Valid
18	0.58	Valid
19	0.67	Valid
20	0.83	Valid
21	0.83	Valid
22	0.83	Valid
23	0.67	Valid
24	0.75	Valid
25	0.67	Valid
26	0.75	Valid
27	0.58	Valid
28	0.75	Valid
29	0.67	Valid
30	0.75	Valid

31	0.92	Valid
32	0.92	Valid
33	0.92	Valid
34	0.75	Valid
35	0.75	Valid
36	0.67	Valid
37	0.75	Valid
38	0.67	Valid
39	0.67	Valid
40	0.75	Valid
41	0.92	Valid
42	0.92	Valid
43	0.92	Valid
44	0.92	Valid
45	0.75	Valid
46	0.75	Valid
47	0.58	Valid
48	0.75	Valid
49	0.92	Valid
50	0.92	Valid
Average	0.77	Valid

Table 1 shows the outcome of Aiken's V analysis. All items are stated valid because the coefficient score of each item is more than 0,3. This is in accordance with the criteria suggested by Azwar (2016) which states that if the validity coefficient is ≥ 0.3 , the item is

valid. The results of this analysis are verified by research held by Arum et al., (2022). This research used Aiken's V formula to analyze the content validity of the developed instrument. The data analyzed showed that all items analyzed were valid. Another research supported this result was conducted by Ad'hiya and Laksono (2018) that also used Aiken's V to analyze 20 items and declared valid with the score of 0.86.

In addition to validity, the reliability of an instrument also needs to be analyzed carefully. After knowing the score of the validity based on specialist decision, the next step is to estimate the reliability of the instrument through the stability agreement among validators (Nugroho et al., 2016). The instrument is said to be reliable if it is not tendentious. The data is correct if the results are still the same after being taken several times (Fitriyanto et al., 2019). The results of the analysis of test instruments performed by experts in this study were estimated for reliability using the two-way ANOVA test procedure through the program of SPSS 27. This procedure aims to estimate the inter-rater reliability of the instrument. Estimation of instrument reliability using SPSS is shown in table 2.

Table 2. The Outcomes of Interrater Reliability Using Two-way ANOVA Test
Test of Between-Subjects Effects
Dependent Variable: SKOR

Source	Type III Sum of Squares	df	Mean Square
Corrected Model	67.873a	149	.456
Intercept	2505.127	1	2505.127
VALIDATOR	10.013	2	5.007
SUBJEK	27.873	49	.569
VALIDATOR*SUBJEK	29.987	98	.306
Error	.000	0	
Total	2573.000	150	
Corrected Total	67.873	149	

R Squared = 1.000 (Adjusted R Squared = .)

Based on the results of reliability estimation by two-way ANOVA test using SPSS 27, it can be seen that the variance between experts is expressed by the Mean Square validator (s^2_v) of 0.306, while the error

variance expressed by the Mean Square validator*subject (s^2_{vs}) gets a result of 5.007. These results are then recalculated using the Hoyt reliability formula with the average of the three experts as follows:

$$r_{xx} = 1 - (s_r^2 / s_s^2) = 1 - 0.306 / 5.007 = 0.94$$

The results of calculations with the Hoyt formula give a reliability coefficient value of 0.94. It shows that the scores given by experts are consistent. This result is also supported by research conducted by Zulpan dan Rusli (2020) who used a two way ANOVA test to analyze the content reliability of an instrument. The study stated that the instrument compiled was reliable. Another

research that also supported this result is a research by Shida et al., (2023) that used inter-rater agreement with the coefficient value of 0.55.

In addition to assessing the test instruments that have been developed, the validators also provide feedbacks and suggestions for improving the quality of the instruments to be tested. Feedback and suggestions from validators can be seen in table 3.

Table 3. Validators' Feedback and Suggestions

Validators	Feedback/suggestions
Validator 1	<p>The order of the difficulty level in the questions can be observed. It can be started from easy items to more difficult items (according to taxonomi level).</p> <p>The order in which the items are presented can be adjusted based on the information presented in the texts.</p> <p>Indicators of determining word equivalents in the texts (without structural clues) can be classified into level C1 (knowledge) or adjustments need to be made in the texts by adding structural clues to match the indicators.</p> <p>For paper-based tests, it is recommended to use line information in the texts and test items to make it easier and faster for test takers to find the information they need or are looking for.</p> <p>Bold letters need to be used for some words or information in the texts and items to make it easier to find and to emphasize.</p> <p>It is necessary to standardize the question display format and the question format settings, such as letters, spaces, tabulation, alignment, etc.</p> <p>The content presented in the texts can be varied. Of the five texts presented, there are three texts with the same history theme.</p>
Validator 2	<p>The texts in the sample questions should have Islamic nuances.</p> <p>It is necessary to check whether the questions have been made sequentially according to the flow of information presented in the texts.</p> <p>It is important to number one of the lines in the texts so that the test takers can answer the questions more effectively.</p>
Validator 3	No feedback/suggestions

The instrument for reading comprehension skill test of Islamic English text was declared valid and reliable based on Aiken's V test and two way ANOVA test.

CONCLUSION

Based on the outcomes and discussion, it can be concluded that the test instrument for reading comprehension skills of Islamic English texts with a validity value using the Aiken's V formula is sufficient, the general

validity of the V result is 0.77, and the estimated reliability among raters uses the two-way ANOVA test with sufficient reliability results. reliable of 0.94. Therefore the reading skills test instrument for understanding Islamic English texts can be continued in the trial step to measure reading comprehension skills for higher education students in Islamic university.

REFERENCES

- Ad'hiya, E., & Laksono, E. W. (2018). Development and validation of an integrated assessment instrument to assess students' analytical thinking skills in chemical literacy. *International Journal of Instruction*, 11(4), 241–256.
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 5(1), 131–142.
- Arum, A. E., Khumaedi, M., & Susilaningih, E. (2022). Validity and Reliability of Development of Self-confidence Assessment Instruments for Students on Chemistry Subject. *Journal of Research and Educational Research Evaluation*, 11(1), 62–69.
- Azwar, S. (2016). *Tes Prestasi*. Pustaka Pelajar.
- Fathurohman, I., & Cahyaningsih, R. D. (2021). Development of Authentic Assessment Instrument for Improving Extensive Capacity of Metacognitive Perspective. *Asian Journal of Assessment in Teaching and Learning*, 11(1), 13–21.
- Fitriyanto, J. N., Widjanarko, D., & Khumaedi, M. (2019). Validity and Reliability Test of Assessment Instrument of the Suitability of Electric Power Steering Media. *Journal of Vocational Career Education*, 4(37), 61–69.
- Guntur, L. M. F., & Pordanjani, S. R. (2019). Exploring the Challenges of Reading Comprehension Teaching for English Proficiency Test Preparation Class in Indonesia. *ELS Journal on Interdisciplinary Studies on Humanities*, 2(3), 321–330.
- Hendryadi, H. (2017). Validitas Isi: Tahap Awal Pengembangan Kuesioner. *Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT*, 2(2), 169–178.
- Kusumastuti, A., Khoiron, A. M., & Achmadi, T. A. (2020). *Metode Penelitian Kuantitatif*. Budi Utama.
- Maghfirotilah, Q., Fajriyah, L., & Hariri, A. (2015). Perspektif Mahasiswa Berlatar Jurusan Keislaman terhadap Pembelajaran Bahasa Inggris. *Jurnal Pendidikan Agama Islam*, 3(2), 257–275.
- Mardapi, D. (2018). *Teknik Penyusunan Instrumen Tes dan Nontes*. Parama Publishing.
- Munir, S., & Hartono, R. (2016). Islamic Magazine Articles To Enhance Students' Reading Skill and Build Their Character Values. *International Journal of Education*, 9(1), 66.
- Nugroho, B. S., Djuniadi, D., & Rusilowati, A. (2016). Pengembangan Penilaian Kinerja Menggambar Teknik Potongan di SMK pada Kurikulum 2013. *Journal of Research and Educational Research Evaluation*, 6(1), 01–07.
- Nugroho, T., Bharati, D. A. L., & Hartono, R. (2018). English Teachers' Perception on Strategies in Teaching Reading Comprehension to Motivate the Students. *English Education Journal*, 9(1), 56–61.
- Shida, N., Abdullah, A. H., Osman, S., & Ismail, N. (2023). Design and development of critical thinking learning strategy in integral calculus. *International Journal of Evaluation and Research in Education*, 12(1), 284–291.
- Suaidi, A., Arifin, M. N., Syahid, A. H., & Sa'diah, S. (2019). Developing Model Test of Reading Comprehension for Islamic Studies. *Proceeding of Annual International Conference on Islamic Education and Language: The Education and 4.0 Industrial Era in Islamic Perspective*, 83–88.
- Sukarni, S., Rukmini, D., Sofwan, A., & Hartono, R. (2017). The Effectiveness of Strategy-based Reading Instruction (SBRI) for Teaching Reading and the Students ' Perception toward the Instruction. *International Journal of English and Education*, 6(3), 213–228.
- Susanto, Y., Alfian, R., Rahim, Z., & Karani. (2018). Uji Validitas dan Reliabilitas Kuesioner EQ-5D Bahasa Indonesia Untuk Mengukur Kualitas Hidup Pasien Hemodialisa Gagal Ginjal Kronik. *Jurnal Ilmiah Manuntung*, 4(1), 41–47.
- Suwito. (2011). *Analisis Isi: Pengantar Metodologi untuk Penelitian Ilmu Komunikasi dan Ilmu-ilmu Sosial Lainnya*. Kencana.
- TOEFL ITP Test Content. (n.d.). Diambil 24 Oktober 2022, dari <https://www.ets.org/toefl/itp/test-content.html#Level1> (IntermediatetoAdvanced)
- Wiratno, T. (2018). *Pengantar Ringkas Linguistik Sistemik Fungsional*. Pustaka Pelajar.
- Zulpan, Z., & Rusli, A. (2020). Validitas Dan Reliabilitas Instrumen Penilaian Membaca Short Functional Text Pada Siswa Smp Kelas Viii. *Jurnal Pendidikan Guru*, 1(1), 86–95.