



Premise: Journal of English Education and Applied Linguistics

e-ISSN: 2442-482x, p-ISSN: 2089-3345

<https://fkip.ummetro.ac.id/journal/index.php/english>

DOI: 10.24127/pj.v14i1.11448

RASCH MODEL CLOZE TEST TO EVALUATE VOCABULARY OF ISLAMIC ELEMENTARY SCHOOL STUDENTS

by

Maulidiyyatul Uswah

Tadris bahasa Inggris, Institut Agama Islam Negeri (IAIN) Kediri, Indonesia

Maulidiyyatul Uswah maulidiyyatulu@gmail.com

Ary Setya Budhi Ningrum

Tadris bahasa Inggris, Institut Agama Islam Negeri (IAIN) Kediri, Indonesia

ary_oyesip@iainkediri.ac.id

Muhammad Dliya Uddin Syafi'i

Syari'ah Islamiyah, Universitas Al Azhar Kairo, Mesir

dliyaudin10@gmail.com

*corresponding author

(Article History: Received:23-11-2024;Reviewed1:17-02-2025;Reviewed2:12-03-2025: Accepted:29-03-2025;Published:30-03-2025).

Abstract:

Numerous studies on cloze tests have been conducted regarding their beneficial use; however, research examining their efficacy for students in primary schools remains scarce. Consequently, existing cloze test designs often lack innovation and practicality for classroom implementation. A cloze test can be an alternative method for assessing elementary school students. Therefore, this study aims to develop a practical, innovative, and user-friendly cloze test tailored explicitly for elementary school students, providing teachers with a valuable tool to enhance vocabulary assessment. Hence, the Research and Development (R&D) under the ADDIE model is employed to develop the final product of English assessment, especially for fourth graders in elementary school. The average score percentage from linguistic and media experts was 93%, indicating that the test product was sufficiently good and ready for usage. The reliability was satisfactory, as indicated by a score of 0.73. The cloze test demonstrated an improvement in students' vocabulary mastery, as evidenced by the increase in scores from pre-test to post-test.

.Keywords: *cloze test, Rasch model, vocabulary knowledge*

Abstrak:

Banyak penelitian tentang tes cloze telah dilakukan terkait manfaat penggunaannya, namun penelitian yang menguji keefektifannya untuk siswa di sekolah dasar masih langka. Akibatnya, desain tes cloze yang ada sering kali kurang inovatif dan kurang praktis untuk diterapkan di kelas. Padahal, tes cloze dapat menjadi salah satu cara alternatif untuk menilai siswa sekolah dasar. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan tes cloze yang efektif, inovatif, dan mudah digunakan yang secara khusus dirancang untuk siswa sekolah dasar, sehingga dapat menjadi alat yang berharga bagi guru untuk meningkatkan penilaian kosakata. Oleh karena itu, Penelitian dan Pengembangan (R&D) di bawah model ADDIE digunakan untuk mendapatkan produk akhir penilaian bahasa Inggris terutama untuk siswa kelas empat sekolah dasar. Persentase skor rata-rata dari ahli bahasa dan media adalah 93%, yang menunjukkan bahwa produk yang diuji cukup baik dan siap untuk digunakan. Reliabilitasnya pun memuaskan, seperti yang ditunjukkan oleh skor 0,73. Tes cloze menunjukkan adanya peningkatan penguasaan kosakata siswa, yang ditunjukkan dengan adanya peningkatan dari skor pre-test ke post-test.

Kata kunci: *tes cloze, model Rasch, pengetahuan kosakata*

How to cite this article:

Uswah, M., Ningrum, A. S. B., & Syafi'i, M. D. U. (2025). Rasch model cloze test to evaluate vocabulary of Islamic elementary school students. *Premise: Journal of English Education and Applied Linguistics*, 14(1), 209–229. <https://doi.org/10.24127/pj.v14i1.11448>

Uswah et al. (2025)

INTRODUCTION

Children's English proficiency in Southeast Asia varies by country, and some countries in the region rank among the highest globally. In general, countries such as Singapore and the Philippines often top global rankings for English proficiency, both among children and adults. Singapore, for example, is often ranked first in Asia and among the highest in the world. The Philippines is also known for having a highly fluent English-speaking population due to its education system and history (Turmudi & Hajan, 2020). Meanwhile, Indonesia is noted for its low proficiency level among groups such as Myanmar (EF English Proficiency Index, 2023).

The low proficiency in English in Indonesia is evident in the limited vocabulary acquisition of their children. Indonesia's students struggle with pronunciation due to several factors. These factors include a lack of understanding of English vocabulary in the material presented and a limited ability to use vocabulary, including synonyms and antonyms, in Writing (Nazhira et al., 2024). Moreover, the students were reluctant to open their dictionaries and showed less interest in learning the English language. The other issues were external factors such as inadequate facilities and motivation (Afidah & Mahfudi, 2022; Suryantini et al., 2024; Budianto et al., 2024).

Besides, one key factor contributing to this challenge is the tendency for children to get the practical test, which applies to developing their vocabulary achievement. Gulshoda (2024) noted that vocabulary plays a crucial role in facilitating effective communication and comprehension. Furthermore, Widodo et al. (2022) claimed that one of the factors that causes students' vocabulary mastery to be not optimal is that the assessment in learning still focuses on the final result displayed through the score of final tests rather than on the student's vocabulary mastery process. In line with Kamaludin et al. (2022) and Zabidin (2021), they stated that vocabulary instructions and strategies such as assessment activities are needed to boost vocabulary attainment among different proficiency levels in English language learning settings.

In the case of vocabulary assessment, various models and strategies are available that can be tailored to suit any level and purpose, as demonstrated by some studies. Studies in the context of vocabulary assessment have found that elementary teachers should employ dynamic, interactive, and more accessible strategies when assessing their students

Uswah et al. (2025)

(Boyinbode, 2018; Kohnke, 2020; Taipang, 2022). However, not all vocabulary assessments or test types are suitable for primary students. In this regard, the cloze test provides a valuable alternative (Sadeghi, 2021). The Cloze test is one of the language assessments that require students to use their critical thinking by filling in the blank words or phrases based on the reading context (Abdullaev et al., 2023). Similarly, studies in the context of the Cloze test supported the benefits of using the Cloze test to evaluate students' vocabulary knowledge being easier to understand, emphasizing practice and innovation to develop students' metacognitive skills, and an effective and efficient way to measure situation-level understanding that draws on language skills, word-level processes, and working memory resources. Moreover, a deeper cloze test can also be used to predict students' performance on important criterion tasks requiring a deeper level of understanding (Silfia, 2024; Huang & Wu, 2024; Bråten et al., 2024; Omuna, 2024; Ani, 2016; Zare & Boori, 2018).

Returning briefly to the issue of the cloze test, it should be consistently measured for reliability and validity. This assessment can be measured by using the Rasch model approach. In this study, the cloze items under examination fit the Rasch model well, confirming the objectivity of these item measurements in evaluating the assumed abilities of the students (Jiayang, 2024). Several extensive cross-sectional studies suggest the implementation of the Rasch model to examine the validity and reliability of the cloze test (Abdullaev et al., 2024; Boori, 2024; Effatpanah et al., 2024). Furthermore, the researchers argued that the cloze test could be analyzed using the Rasch model, and locally dependent items were identified through residual correlations.

Although extensive research has been carried out on the cloze test (Feng, 2024; Wang & Fan, 2024; Hambali et al., 2023; Entika et al., 2024; Fakhruddin et al., 2024), no single study discussed vocabulary knowledge, particularly in primary school students. However, the primary focus of this study is on how the cloze test evolves as an assessment tool and how effectively the cloze technique supports teachers in evaluating students' vocabulary acquisition. Through this test, students present their understanding of the information in the text. It is worthwhile to conduct a study on the vocabulary knowledge of students assessed through cloze tests in an Islamic elementary school in Kediri. The findings will contribute to scholars' understanding of English scholarly publications in the era of the Industrial

Uswah et al. (2025)

Revolution (Turmudi, 2020). Hence, this study functions as a learning process to contribute knowledge to humanity.

Based on the literature above, the research questions of this study are:

1. How is the process of analysis stage in assembling the cloze test for assessing elementary school students' vocabulary mastery?
2. What is the process of designing the cloze test in the assembly stage for assessing elementary school students' vocabulary mastery?
3. How is the process of developing the cloze test for assessing elementary school students' vocabulary mastery?
4. How is the process of implementing the cloze test stage in assembling it for assessing elementary school students' vocabulary mastery?
5. How is the process of evaluation stage in assembling the cloze test for assessing elementary school students' vocabulary mastery?

METHOD

Design

A research and development (R&D) approach is employed in this study. The goal in adapting R&D, in this case, is to create a new cloze test. It means creating an exam that is designed solely to assess students' vocabulary proficiency in English-related courses. Students will be required to fill in reading passages with intentionally placed gaps to demonstrate their understanding of key concepts and themes. The model established by Dick and Carey (1996) is the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model, which is used for the development of this cloze test assessment product.

Participant

This study involved 23 children from the 4th grade A class at an Islamic elementary school. The fourth-grade class of an Islamic primary school was selected because it was deemed suitable for vocabulary enhancement. Fourth-grade students are typically aged 9 to 10 years, during which their language development progresses rapidly. At this stage, children demonstrate an enhanced capacity to comprehend, retain, and utilize new words in appropriate circumstances. Moreover fourth grade is an optimal period to assess the degree of

their vocabulary advancement. By evaluating vocabulary in 4th grade, educators may gauge students' comprehension of fundamental and intermediate vocabulary across the languages being studied, which is essential as a basis for advanced learning.

Instrument

This study employed various instruments to collect data, including questionnaires for students, interviews with English teachers, vocabulary tests for students, and several questionnaires for validation conducted by English teachers, who served as language experts and media experts.

A questionnaire is used to analyze students' prior knowledge, identify their areas of difficulty, and assess their need for vocabulary knowledge, particularly in cloze tests. This instrument is used to fulfill the preliminary study in the Analysis step of the ADDIE model. Five questionnaires should be answered by 23 students who are participants.

The second instrument implemented in this study is an interview. The English teacher is given 10 questions based on her background of education, the course outline, the teaching style, the teaching method, and the learning assessment. It aims to help the researcher arrange the concept, design the format, and design the item number of the cloze test. The interview is still in the preliminary stage of the study.

The second questionnaire is questionnaire validation. The researcher asked the English teacher to serve as an expert validator. Two components need to be measured: the content and the language used in the test. The expert was required to provide validation for each item number of the cloze test. The learning objectives are being considered.

The other instrument is a vocabulary test. This study uses a cloze test to assess the participants. There are 13 item numbers related to their materials, including items and parts of the house, as well as activities within the house and school. The pictures are provided with every item number to make the test easier and more joyful for the students.

Data collecting technique

The data collection techniques employed in this study ensure that each phase of the ADDIE model—Analysis, Design, Development, Implementation, and Evaluation—aligns with the research objectives and enhances instructional effectiveness.

Analysis phase

In the ADDIE model, data collection occurs primarily in the **Analysis** phase to ensure a well-informed foundation for instructional design. Researchers gather qualitative and quantitative data through surveys, interviews, observations, and document analysis. This data helps identify learners' needs, existing knowledge gaps, and technological or pedagogical constraints. In this phase, the procedures are identifying the student's needs, interpreting the curriculum, and assessing the environment as crucial matters to observe. The close-ended questionnaires were conducted to uncover the students' needs and whether the interview guide had been released to the English teacher.

Designing phase

The next step is the design stage. At this stage, various procedures are undertaken, including designing assessment instruments, instructional strategies, and instructional materials. First, we are incorporating the learning objectives into the table specification product, gathering appropriate materials, and formulating suitable indicators. In this case, designing the blueprint of the assessment and cloze test as an instrument related to the provided materials should be established. The researcher adapted the materials from an English textbook entitled 'My Next Words Grade 4 – Student's Book for Elementary School' written by Lilin Rahmawati, Nur Fitria Angrisia, and Lili Nailufary. This book was published by *Pusat Kurikulum dan Perbukuan Badan Penelitian dan Pengembangan dan Perbukuan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi* in 2021—the materials entirely for the specific topics such as daily activity and the room as part of the house.

Developing phase

At this stage, the products are being sent for expert validation. In this stage, a questionnaire served as the primary tool for gathering feedback. This questionnaire assessed seven key areas of the product consisting of clarity and instructions of individual items, overall presentation and organization of the items, how well-suited the items are to their purpose, comprehensiveness of the content covered by the items, whether the questionnaire effectively achieves its intended goals, alignment of the questions with the overall objective, appropriateness of the response scale and evaluation system and the responses from this questionnaire were then used to refine and improve the product.

Implementation phase

After revision, the next step is to conduct a product trial. Implementing the product is required to observe the students' perspective through the evaluation of instruction, duration, and materials, as assessed in this case by the cloze test. The data for the field tryout was collected through a cloze test specifically designed for this study. If student performance data revealed difficulties in achieving the stated learning goals within each test unit, the test was further revised to address these obstacles.

Evaluation phase

The final step is evaluation. Evaluation is a purposeful activity. At this stage, the product is assessed to determine whether it meets the intended objectives of the product development. In this study, the researcher planned to develop a cloze test to evaluate students' vocabulary mastery. In an Islamic elementary school

Data analysis technique

The data analysis technique is applied in the design and development of the product. To analyze data, two categories of valuation were applied. They are linguistic expert and media experts.

In this study, the English teacher was used as the linguistic expert. For media experts, the Rasch model approach is applied. The expert was required to provide validation for each item number of the cloze test. The expert provides an assessment for each item number, ranging from 1 to 4, which is then converted into a percentage. After the value is generated, it will be added to the total value, which will be combined with the values of the other components. The final values indicate whether the product is suitable to try out or needs revision as well.

Furthermore, the media expert applies the Rasch approach, first calculating the item difficulty using Excel, which represents the item discrimination of the instrument. Analyze the test results once participants have finished. The Rasch method was used to assess the instrument's validity and reliability. Rasch measures test reliability at the individual, item, and overall levels more thoroughly. Researchers can be reassured by the validity of the instruments, which ensures that each item is accurately assessed. This study measured

Uswah et al. (2025)

reliability using Cronbach's Alpha, which assesses the internal consistency of a test or questionnaire. Greater Cronbach's Alpha indicates greater reliability. More than 0.70 is acceptable for research instruments. High internal consistency is indicated by a value > 0.80 . Low-reliability scores (below 0.50) indicate the need for redesigning the test items. The Rasch Model evaluates test items using Outfit Mean Square (MNSQ), Outfit Z-standard (ZSTD), and Point Measure Correlation (Pt Measure Corr) values. The acceptable MNSQ range is $0.5 < \text{MNSQ} < 1.5$, with values below 0.5 indicating easy items and above 1.5 indicating complex or inconsistent items with the model. Items having ZSTD values outside of -2.0 to $+2.0$ are either not functioning well or not contributing meaningfully to the test. Additionally, the Point Measure Correlation (Pt Measure Corr) Value should be between 0.4 and 0.85, as values below 0.4 indicate a poor correlation with other items, while values above 0.85 suggest redundancy. The number of tests that were reported as too easy, easy, complex, and very difficult was recorded in this step. The item numbers that were too easy were disregarded, while the challenging item numbers were changed. To determine the increase in students' vocabulary knowledge, data processing was carried out quantitatively using the WinSteps Rasch model computer program, version 5.4.7.0.

RESULT AND DISCUSSION

The result of analyzing the phase

In this stage, it was found that this Islamic elementary school has implemented *Kurikulum Merdeka*. English is not a choice subject. Integrating English into the curriculum is required due to its literacy component, one of the characteristics of *Kurikulum Merdeka*. One of the goals of English teaching is to master a text on a specific topic.

After distributing the questionnaires, the results showed that 86% of the students agreed that they found the English text difficult to understand when taking the vocabulary test. Moreover, most of them, around 83%, agreed that they rarely had a vocabulary test based on the cloze test. However, a test of vocabulary knowledge is needed to determine the students' success criteria.

While the students were filling out the questionnaires, the researcher interviewed the teacher. The English teacher is given 10 questions based on her background of education, the course outline, the teaching style, the teaching method, and the learning assessment. It aims to

Uswah et al. (2025)

help the researcher arrange the concept, design the format, and design the item number of the cloze test. The result is that the teacher uses a conventional method of teaching and learning, as evidenced by the student workbook. She merely explains the material to the students for today and asks them to complete the workbook (LKS). The teacher never gives the students an assessment, such as a cloze test, in an intensive manner.

The result of designing phase

Otherwise, the researcher, as the teacher, arranges to create the cloze test. The vocabulary test adapted the materials from an English textbook entitled 'My Next Words Grade 4 – Student's Book for Elementary School' written by Lilin Rahmawati, Nur Fitriana Angrisia, and Lili Nailufary. This book was published by *Pusat Kurikulum dan Perbukuan Badan Penelitian dan Pengembangan dan Perbukuan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi* in 2021.

The cloze test consists of 13 multiple-choice items. The materials are designed explicitly for topics such as daily activities and the room within the house. The experienced English teacher also validated this product.

The result of the developing phase

The validation of the English teacher was conducted quietly through questionnaires. There are two valuation categories: media experts and linguistics experts. All of the fourteen-item tests are valid. The eight numbers of the item test are understandable enough, and the rest are straightforward. Six-item tests could be used without revision, and the eight numbers should be revised based on the teacher's notes. The topics of the cloze test are also related to the level of students' grades. The word choice is adequate in terms of reading comprehension, encompassing aspects such as nouns, main verbs, adjectives, articles, and pronouns.

The expert provides an assessment for each item number, ranging from 1 to 4, which is then converted into a percentage. After the value is generated, it will be added to the total value, which will be combined with the values of the other components. The final values indicate whether the product is suitable for trial or requires revision. The average score of the media expert was 85.71, while that of the linguistic expert was 100. This average score is divided into two categories due to the two evaluation criteria: the content of the test and the language used in the test.

Uswah et al. (2025)

$$\text{Percentage (100\%)} = \frac{(85,71+100)}{2} = 93\%$$

Figure 1. The percentage of expert's validation

The result of the expert validation showed that the product of the cloze test was good and ready for use in the reading test in the classroom. While giving additional pictures as the illustration of the question is the teacher's suggestion. The use of pictures can enhance students' performance in vocabulary assessments.

The result of the implementation phase

After revising the first draft, the cloze test was administered to the test takers. The test was administered to the students on May 13, 2024. The classroom situation was conducive to the students' response. The tryout was conducted to measure the students' vocabulary knowledge. Twenty-three participants joined to take the cloze test in the classroom.

In this phase, the researcher performed as the teacher. The tryout was conducted by giving instructions to the students. After that, the teacher offered the students the opportunity to ask questions about the difficult terms or vocabulary in the test and finished all the numerical items as well. Then, the students' scores were collected and analyzed. The item difficulty and item discrimination were calculated by using Microsoft Excel. In contrast, the validity and reliability were calculated by using Ministep of Winstep Application version 5.4.7.0.

The result of the evaluation phase

The result of the item difficulty calculation, performed using Excel, showed that several numbers fell into the difficult category. There were 14 numbers of items or questions assisted in the test. The most challenging categories were items 5, 8, and 9. While items 2, 3, 4, 6, 13, and 14 are medium-complex. The remaining item numbers were categorized as easy. For item discrimination, two item numbers were rejected. The rejected item numbers were 1 and 7. In comparison, the item numbers that needed revision were 8, 10, 11, and 12. The good category discrimination was numbers 6 and 9. For item numbers 2, 3, 4, 5, 13, and 14 are excellent categories of discrimination.

This further data showed the results of the reliability analysis of the test based on the Rasch Model. The Cronbach Alpha is 0.73. The reliability of the item test, analyzed using the Rasch model, indicated that the test fell into the good category of reliability. Additionally, the test's validity revealed that 11 of its items had good validity and were not rejected. With a coefficient of less than 0.5 or greater than 1.5, the three items in the test were deemed either too easy or too challenging, and they needed to be replaced.

Table 1: The result of the validity

<i>Entry Number</i>	<i>INFIT</i>		<i>OUTFIT</i>	
	<i>MNSQ</i>	<i>ZSTD</i>	<i>MNSQ</i>	<i>ZSTD</i>
1	1,86	2.87	2.34	2.76
7	1,80	2.77	2.28	2.67
8	1.78	2.71	2.06	2.55
9	1.27	1.15	1.31	1.04
11	1.11	.41	1.06	.51
2	1.06	.32	.85	-.15
4	.96	-.11	.88	-.23
6	.91	-.33	.82	-.33
12	.91	.00	.66	-.01
14	.90	-.39	.79	-.51
3	.83	-.75	.82	-.41
10	.81	.06	.29	-.24
5	.79	-.87	.68	-.97
13	.72	-1.36	.63	-1.39

The validity of the test was demonstrated by 11-item tests, indicating good validity and no rejections. There were three items tested that indicated the Outfit Mean Square (MNSQ) Value was more than 1.5 times the coefficient. It means that the item test was too tricky and needed to be replaced as well. The two remaining numbers indicated less than 0.5. It means that the item test was too easy to answer.

Table 2: The participants' test score

<i>Student's Name</i>	<i>Score</i>	
	<i>Pre-Test</i>	<i>Post Test</i>
VNA	83,80	85,68
NLA	62,20	64,26
FTH	62,26	64,26
ARJN	97,96	99,96
TSY	47,98	49,98
LBB	62,26	64,26
RYHN	40,84	42,84

Uswah et al. (2025)

<i>RHMN</i>	98,00	99,96
<i>ALD</i>	62,26	64,26
<i>DNS</i>	48,80	49,98
<i>AKBR</i>	47,98	49,98
<i>NYSLL</i>	69,4	71,4
<i>ZHFN</i>	90,82	92,82
<i>NZM</i>	69,4	71,4
<i>HFZ</i>	83,70	85,68
<i>ALM</i>	90,82	92,82
<i>VNS</i>	55,12	57,12
<i>MR</i>	76,60	78,54
<i>HLM</i>	97,96	99,96
<i>JY</i>	97,96	99,96
<i>RK</i>	83,68	85,68
<i>ANS</i>	98,00	99,96
<i>AM</i>	76,60	78,54

Based on the participants' scores on the pre-and post-tests, the results are statistically significant. It means that using cloze tests can enhance students' vocabulary knowledge. It showed that students perform better on close tests.

Discussion

The findings of the students' preliminary study are consistent with the research carried out by Marianca et al. (2022), which indicates that due to a lack of vocabulary, many students still struggle to comprehend English texts and respond to reading assignments. Due to their limited vocabulary, they are unable to understand the meaning of words in English texts.

Aligned with the students' questionnaire results, the teacher's interview stated that the teacher rarely conducted cloze tests. Due to the students' reading worksheets, they mostly used the usual or conventional test. The teacher was worried that the students might be confused because they were unfamiliar with the test model. As stated by Shidqi et al. (2024), teachers should continually develop their teaching and assessment methods in innovative ways. The results of the investigation align with the research carried out by Pratiwi (2023). It is stated that English scores in understanding and mastering vocabulary remain low due to the lack of variety in teaching and learning materials.

Furthermore, to assess the student's reading comprehension, a cloze test was conducted. The design of cloze tests is released. The cloze test covered several topics related to daily activities and household equipment based on the students' real-world context.

Uswah et al. (2025)

Focusing on integrating skills into an assessment center to validate those skills. Therefore, the use of contextual and authentic tasks that introduce real-world activities and situations into the language classroom is known as integrated assessment Raymundo (2023). The cloze tests are an engaging and interactive test format for test takers. This test is different from traditional tests, which are often monotonous and boring. Besides that, the teacher rarely used the cloze test and always gave students tests based on their worksheet books. The interactive learning media is required to provide the test due to the easy way to use (Meliana Br. Siahaan et al., 2023).

The key point to remember for implementing the test is the teacher's consideration of using unfamiliar words when administering a cloze test, particularly to youngsters who learn slowly. For example, when children are given an astronomy test, they will have great difficulty with the large amount of vocabulary that they rarely use on a routine basis. Most of the objects related to astronomy are things they have never seen in everyday life (Cahyono et al., 2024; Budianto et al., 2024).

Accordance with the mentioned findings by Abdullaev et al. (2023), Xiping & Jiaying (2024), Zare & Boori (2018), Effatpanah et al. (2024), the use of the Rasch model in calculating quantitative data is an effective way due to this method, students' vocabulary knowledge can be known for each question number. In more detail, the Rasch model can also analyze questions that are too easy and too difficult, allowing the test to be adjusted based on students' abilities.

In other words, the good results of validity and reliability in this test indicate that the product of the developing cloze test can be used to assess students' Vocabulary knowledge. The product of this study offers a valuable alternative to traditional assessment methods, as the old way has limitations and lacks a variety of test models (Gao et al., 2020; Sukarni et al., 2021; Hambali et al., 2023).

Significant progress has been made, as indicated by the participants' pre-and post-test results. It implies that employing cloze exams can improve students' vocabulary. The results showed that students perform better on closed-ended tests and that their performance aligns with their overall reading ability (Febriyanti, 2017; Aprilia, 2021; Evenddy & Utomo, 2021). Moreover, the previously mentioned product of the cloze test is that filling in the blank

Uswah et al. (2025)

vocabulary is more convenient than answering multiple-choice questions. In multiple-choice tests, students can guess the correct answer from several available choices, which can affect the accuracy of the test results.

On the other hand, Cloze tests minimize the possibility of guessing because students must understand the sentence to fill in the correct word without the help of answer choices. In multiple-choice, students who are good at identifying patterns or characteristics of answers may be able to choose the correct answer even though they do not understand the content of the material. Because cloze tests have no options, they measure direct understanding rather than the ability to analyze questions. Hence, the results tend to be more purely indicative of students' understanding (Silfia, 2024).

One of the advantages of the Cloze test product is that it enables students to gain a comprehensive understanding of the provided text. In other words, they must understand the overall context to determine the most appropriate word. This is different from multiple choice, where students can rely on selecting answers from several options provided without understanding the context in depth. Thus, the Cloze test is more effective in measuring the extent to which students truly understand the text and have contextual reading skills (Huang & Wu, 2024)

It seems possible that these cloze test results involve recall skills rather than just recognition, which is often measured by multiple-choice items. Because students must generate or "produce" the appropriate word from memory, rather than simply recognizing the correct answer, they more closely impersonate the process of understanding language in real-world situations, where they must choose the correct words without prompts. Because cloze tests imitate the way people read and understand texts in real life, where people are not given answer choices, they are more authentic in measuring students' abilities. It can provide a better picture of how well students can understand and use vocabulary in real-world contexts (Bråten et al., 2024).

The cloze test product in this study was also designed to help improve grammar comprehension, especially when the test involves filling in words closely related to the grammatical structure of the sentence. When students fill in the missing words, they must understand how sentence structures work, such as the use of subjects, predicates, objects, and

Uswah et al. (2025)

adverbs. It means that they must pay attention to the overall grammar to ensure that the words they choose are appropriate to the sentence's context and grammatical structure. Cloze tests can also improve the use of correct word forms. It often involves using words with the appropriate forms, such as verbs that must follow the subject (for example, "is" or "are" depending on the subject), tense forms (for example, present or past tense), and other types of words. This trains students to use the correct word forms according to grammar (Omuna, 2024).

The final product of this research also helps in recognizing conjunctions and prepositions. When students are asked to fill in the missing parts of sentences that contain prepositions or conjunctions, they are trained to recognize which words are grammatically appropriate in a given context. For example, understanding the use of conjunctions such as "because," "although," or prepositions such as "in," "at," and "on" according to the context of the sentence (Ani, 2016)

On the other hand, even though the cloze test is practical in measuring vocabulary in a particular context, it may not be sufficient to measure various aspects of vocabulary, such as synonyms, antonyms, and understanding of words outside of their immediate context. It can be a weakness if the purpose of the study is to measure broader vocabulary comprehension. If not designed well, the Cloze Test may tend to measure only word-guessing ability rather than actual vocabulary comprehension, which may reduce the validity of the results (Trace, 2023).

Moreover, the students' achievement is integrated into their text comprehension, which is controlled by their language background, memory capability, and basic knowledge of the topic. It means that the cloze test is not a single factor determining reading achievement (Bråten et al., 2024).

Likewise, the deep cloze test has several important advantages, including simplicity and ease of administration, and can be used for research purposes. It was found that although the internal consistency of the deep cloze test is good, its correlations with other measures of text comprehension and final exam scores are mostly low. Deep cloze fills an important gap by facilitating the assessment of reading comprehension outcomes for undergraduate Spanish students (Salmerón et al., 2022).

Uswah et al. (2025)

This distinction is further exemplified in studies using a deep cloze test. This unique method is also employed for this purpose. The cloze test method is used to complete the definition of a vocabulary word. In its use this type of cloze test utilizes the COBUILD Lerner's dictionary application (Demiris & Thompson, 2011). They are written in a simple enough manner for readers in grades 2 and 3 to understand. Because it provides context for the word in the definition and presents the meaning in a complete sentence, the COBUILD-style definition format is preferred over conventional definition formats. Compared to cloze tests that only fill in missing vocabulary by paying attention to the sentence context, the COBUILD style offers higher knowledge and understanding, which should be adjusted to the target being tested as a consideration.

Areas, where significant differences have been found include the cloze test design employed in this study, particularly concerning elementary school children, and the one used in the high school students studied by Feng (2024). Multiple-choice cloze examinations are essentially used equally to assess students' vocabulary. However, it seems that high school students' use of lengthy, compound, and compound-level literature makes the test more challenging than that of elementary school students, who typically use shorter texts. Since high school students are thought to possess more vocabulary and critical thinking skills than elementary school students, these factors are taken into account when designing tests. High school students also reported in the review that the test's lengthy texts were challenging.

The study's results have several flaws. First of all, it uses a random sample without taking into account each student's prior knowledge. Due to potential differences in teaching techniques, instructional time, and English-language textbooks, the findings of this study may not apply to students in other elementary schools across different countries. Moreover, applications other than the Rasch model can be used in future research to assess validity and reliability.

CONCLUSION AND IMPLICATION

Conclusion

There are several ways to find out elementary school students' vocabulary knowledge, one of which is using a cloze test. This assessment instructs students to complete the missing words correctly based on the sentence's context. The advantage of this cloze test is that it can

Uswah et al. (2025)

be used to measure components in English such as vocabulary and grammar. Additionally, a cloze test can be used to measure intensive English language skills, particularly for elementary school students, encompassing both Writing and reading. In contrast, more innovation can also be used to measure speaking and listening skills. Cloze tests can be an alternative to traditional methods in elementary school student assessments. The results of this study also show that the English achievement scores of the students who were tested using the cloze test increased compared to the pre-test scores. Moreover, the design of the cloze test was also validated by assessments by language experts and media experts. While it was tested using the Rasch model application, this allows readers to understand the advantages of the Rasch model, which can identify questions that are too easy or too difficult, enabling revisions to be made.

Limitation

Although cloze tests have been proven to improve students' vocabulary knowledge, these findings may differ if applied to schools with different curricula, teaching methods, and language proficiency levels. Further research is required to confirm the effectiveness across different student populations.

Implication

The positive impact of this finding is that the cloze test can serve as an alternative method for teachers, especially in English assessments, to create high-quality cloze test items that accurately measure student progress. The teachers enable students to use self-designed materials tailored to their levels. Additionally, educational technology, when applied as a Rasch model, is suitable for measuring validity and reliability, thereby enhancing students' performance in assessments. Moreover, advanced educational applications in the English language are widely used to facilitate the learning process and are recommended.

ACKNOWLEDGEMENT

We want to express our heartfelt gratitude to Professor Dr. Rohmani Nur Indah, whose excellent counsel, constructive input, and ongoing support have been instrumental in the creation of this work. We also thank the teachers and students of SD Islam Pojok 2, particularly the fourth-grade students, for their participation and cooperation throughout the

Uswah et al. (2025)

data collection process. Their willingness to participate in this research has resulted in significant insights and contributions.

BIO-PROFILE:

Maulidiyyatul Uswah holds a bachelor's degree in English Education from IAIN Kediri. She is now pursuing her master's degree in English Education at IAIN Kediri. Her expertise is in ELT, Speaking, Writing, and Professional Development. Corresponding email: maulidiyyatulu@gmail.com

Ary Setya Budhi Ningrum is a senior lecturer at the English Language Education Department of The State Islamic Institution (IAIN) Kediri. She is also an editor of the Journal of English Education and Linguistics Studies (JEELS). Her research interests are English language teaching, research methodology, and English for young learners. Email: ary_ovesip@iainkediri.ac.id

Muhammad Dliya Uddin Syafi'i is an undergraduate student of Syari'ah Islamiyyah from Al-Azhar University, Kairo, Mesir. His interest in developing English language learning materials that focus on Islamic topics, such as Islamic legal vocabulary, religious texts, or contemporary issues. Email: dliyaudin10@gmail.com

REFERENCES

- Abdullaev, D., Shukhratovna, D. L., Rasulovna, J. O., Umirzakovich, J. U., & Staroverova, O. V. (2023). Examining local item dependence in a cloze test with the Rasch model. *International Journal of Language Testing*, 14(1), 75–81. <https://doi.org/10.22034/IJLT.2023.409812.1273>
- Ani, M. I. (2016). *Assessment of Readability in Basic Science Textbooks Using the Cloze Procedure*. International Journal of Studies in Education – Vol. 20, Issue 2, 2024, 253-261. <https://ijose.unn.edu.ng/wp-content/uploads/sites/224/2024/07/Paper-20.2.26.pdf>
- Boyinbode, O. (2018). Development of a gamification-based English vocabulary mobile learning system. *International Journal of Computer Science and Mobile Computing*, 7(8), 183–191. www.ijcsmc.com
- Bråten, I., Haverkamp, Y. E., & Anmarkrud, Ø. (2024). Gaining a deeper understanding of the deep cloze reading comprehension test: examining potential contributors and consequences. *Reading and Writing*, 1. <https://doi.org/10.1007/s11145-024-10521-y>
- Budianto, S., Huda, N., Sayidah, N., & Sucipto, S. (2024). Enhancing fourth-grade elementary students' English vocabulary through listening with Received Pronunciation (LWRP). *QALAMUNA: Jurnal Pendidikan, Sosial, Dan Agama*, 16(1), 161–172. <https://doi.org/10.37680/qalamuna.v16i1.4717>
- Demiris, G., & Thompson, H. (2011). Smart homes and ambient assisted living applications: From data to knowledge-empowering or overwhelming older adults? Contribution of the IMIA smart homes and ambient assisted living working group. *Yearbook of Medical Informatics*, 6(1), 51–57. <https://doi.org/10.1055/s-0038-1638738>
- Effatpanah, F., Baghaei, P., & Babaii, E. (2024). A new scoring method for item response theory analysis of C-Tests. *Language Testing Journals*. Sagepub. <https://doi.org/10.1177/02655322241265350>
- Entika, F. P., Lestari, F. P., & Maria Mokhtar, M. (2024). Investigating skills-based language assessment literacy of EFL teachers in Indonesia. *KnE Social Sciences*, 2024, 394–404. <https://doi.org/10.18502/kss.v9i6.15287>
- Kusfitriyatna, F., Evenddy, S. S., & Utomo, D. W. (2021). The effect of using cloze passage technique toward students' reading comprehension in narrative text. *Journal of English Language Learning*, 5(1), 49-62. <https://www.neliti.com/publications/422204/the-effect-of-using-cloze-passage-technique-toward-students-reading-comprehensio#cite>
- Fakhruddin, Z., Arqam, A., & Maming, I. (2024). Analysis of English listening materials and the compatibility to curriculum. *Jelita*, 5(1), 9–18. <https://doi.org/10.56185/jelita.v5i1.591>
- Piyani, P. (2017, November). The correlation between reading comprehension and students' ability in Answering Cloze Test of The Seventh Grade Students at SMP N 1 Kalipuro Banyuwangi in The 2014/2015 Academic Year. In LUNAR (Vol. 1, No. 02, pp. 35-47). <https://doi.org/10.36526/ln.v1i02.463>
- Feng, K. Z. (2024). Exploration of vocational high school students experiencing difficulty in cloze test performances: a mixed-methods study in Taiwan. *Language Testing in Asia*, 14(1). <https://doi.org/10.1186/s40468-024-00274-4>
- Gao, L., Gimpel, K., & Jensson, A. (2020, July). Distractor analysis and selection for multiple-choice

Uswah et al. (2025)

- cloze questions for second-language learners. In Proceedings of the fifteenth workshop on innovative use of NLP for Building Educational Applications (pp. 102-114). <https://aclanthology.org/2020.bea-1.10/>
- Gulshoda, D. (2024). Exploring vocabulary assessment: Diverse evaluation approaches. *Pedagogical Sciences and Teaching Methods*, 3(32), 131-133. <https://interoncof.com/index.php/denmark/article/view/258>
- Hambali, U., Natsir, R. Y., & Nasir, N. (2023). Virtual reality based hypermedia training for teachers of MAN 2 Model Makassar. *Jurnal IPMAS*, 3(3), 132-141. <https://doi.org/10.54065/ipmas.3.3.2023.302>
- Huang, J., & Wu, K. (2024). The impact of metacognition on scores in the cloze tests of the new college entrance examination for English in Jiangsu province: A case study of English teaching in the flow classroom at Nanjing no.5 senior high school. *Journal of Education, Humanities and Social Sciences*, 30(5), 69–76. <https://doi.org/10.54097/5m3jmg90>
- Kamaludin, P. N. H. binti, Mohd Yusof, S. binti, Md Nawi, S. binti, Nordin, N. A. binti, Zabidin, N. binti, & Sain, N. (2022). Group online engagement: An analysis from Tuckman model. *International Journal of Academic Research in Business and Social Sciences*, 12(9). <https://doi.org/10.6007/ijarbss/v12-i9/14625>
- Kohnke, L. (2020). Exploring learner perception, experience, and motivation of using a mobile app in L2 vocabulary acquisition. *International Journal of Computer-Assisted Language Learning and Teaching*, 10(1), 15–26. <https://doi.org/10.4018/IJCALLT.2020010102>
- Marianca, J. S., Liando, N. V. F., & Mamentu, A. C. (2022). Improving students' vocabulary through words mapping strategy at SMP Negeri 6 Tondano. *JoTELL: Journal of Teaching English*, 1(8), 897–911. <https://doi.org/10.36582/jotell.v1i8.4366>
- Siahaan, M.B., Marpaung, T.I., & Hutahaean, D.T. (2023). Students' perception towards orai as learning media of speaking skill at grade eight of SMP Negeri 1 Pematang Siantar. *Jurnal Ilmu Pendidikan Dan Sosial*, 2(4), 414–427. <https://doi.org/10.58540/jipsi.v2i4.449>
- Nazhira, N., Kania, D. A., Canda, I., Panjaitan, N., & Ulayya, S. B. (2024). English vocabulary mastery difficulties for primary school learners. *Madani: Jurnal Ilmiah Multidisiplin*, 2(6), 430–434. <https://doi.org/10.5281/zenodo.11256757>
- Omuna, M. O. (2024). Classroom assessment practices and students' achievement in English grammar in selected secondary schools in Kenya. *European Journal of Education Studies*, 11(2), 182–199. <https://doi.org/10.46827/ejes.v11i2.5203>
- Pratiwi, Y. (2023). Improving young learners' vocabulary mastery through Pop-Up Books. Doctoral dissertation, UIN RADEN INTAN LAMPUNG. <https://repository.radenintan.ac.id/32685/>
- Raymundo, J. L. (2023). Exploring the practices on macro skills integrated assessment in Philippine Higher Education Context: Basis in designing a language training program. *International Journal of Language Education*, 7(3), 429–446. <https://doi.org/10.26858/ijole.v7i3.24252>
- Rizqa, Aprilia (2021) The influence of using jumbled sentences towards students' writing ability in recount text at the second semester of tenth grade of SMAN 1 Tanjung Bintang In The Academic Year Of 2020/2021. Undergraduate Thesis, UIN RADEN INTAN LAMPUNG. <https://repository.radenintan.ac.id/14496/>
- Sadeghi, K. (2021). Judgemental validity of L2 reading tests and cloze procedure: Qualitative and quantitative evidence from test-takers. *Assessing Second Language Reading*.

https://doi.org/https://doi.org/10.1007/978-3-030-84470-7_8

- Salmerón, L., Altamura, L., Carmen Blanco, M., Montagud, S., & Vargas, C. (2022). Validation of a reading comprehension efficiency test for Spanish university students. *Psicologica*, 43(2), 1–22. <https://doi.org/10.20350/digitalCSIC/14809>
- Shidqi, M., Hasbi, M., & Zuhri, K. (2024). Learning management of Islamic education in the international baccalaureate middle years programme curriculum at Al - Firdaus Junior High School in Sukoharjo Academic Year 2023 / 2024. *At Tuots : Jurnal Pendidikan Islam* 6(2), 529–546. <https://doi.org/10.51468/jpi.v6i2.719>
- Silfia, E. (2024). A comparison of using multiple choice and cloze tests on students' achievement in reading comprehension at second semester of English Department of FKIP Batanghari University. *Jurnal Ilmiah Universitas Batanghari Jambi*, 24(2), 1290. <https://doi.org/10.33087/jiubj.v24i2.5253>
- Sukarni, S. (2021). The use of cloze test to test reading comprehension of non-English Department students. *Jo-ELT (Journal of English Language Teaching) Fakultas Pendidikan Bahasa & Seni Prodi Pendidikan Bahasa Inggris IKIP*, 8(1), 74-82. <https://doi.org/10.33394/jo-elt.v8i1.3788>
- Suryantini, M. D., Dewi, N. P. A., & Sintadewi, N. M. D. (2024). Reading challenges in elementary school students' English as a foreign language. *Jurnal Basicedu*, 8(1), 576–581. <https://doi.org/10.31004/basicedu.v8i1.7116>
- Trace, J. (2023). The influence of passage cohesion on cloze test item difficulty. *Language Teaching Research Quarterly*, 37, 161–178. <https://doi.org/10.32038/ltrq.2023.37.08>
- Turmudi, D., & Hajan, B. H. (2020). Education system and English language teaching in the Philippines: Implications for Indonesian EFL learning. *Premise: Journal of English Education and Applied Linguistics*, 9(1), 78. <https://doi.org/10.24127/pj.v9i1.2791>
- Turmudi, D. (2020). English scholarly publishing activities in the Industrial Revolution 4.0 : What , Why , and How ? *ELTEJ*, 3(1), 52–63. <http://journal2.uad.ac.id/index.php/eltej/article/view/1890>
- Widodo, U., Sudarto, S., Walyono, W., Suyatno, T., & Dewi, M. P. (2022). Improving students' vocabulary mastery: strategy, obstacles, and problems. *Edukatif: Jurnal Ilmu Pendidikan*, 4(5), 6814–6822. <https://doi.org/10.31004/edukatif.v4i5.3867>
- Xinping, C., & Jiaying, C. (2024). Application of Rasch model for cloze test quality analysis in language teaching situation (Issue Proms 2023). Atlantis Press International BV. <https://doi.org/10.2991/978-94-6463-494-5>
- Zabidin, N. (2021). Putting laughter in learning: Humour and vocabulary comprehension among ESL students with low proficiency levels. *Proceedings of the 9th international conference on advanced material engineering and technology*. <https://doi.org/https://doi.org/10.1063/5.0181867>
- Zare, S., & Boori, A. A. (2018). Psychometric evaluation of the speeded cloze-elide test as a general test of proficiency in English as a foreign language. *International Journal of Language Testing*, 8(2), 33–43. <https://eric.ed.gov/?id=EJ1299480>