



RiELT JOURNAL

THE EFFECT OF EDBOT IN ASSESSING LITERAL READING COMPREHENSION IN MTsN 4 PADANG

Tridara Rizkia

UIN Imam Bonjol Padang

tridararizkiatanjung@gmail.com

Arwemi

UIN Imam Bonjol Padang

arwemi@uinib.ac.id

Haryudi Nizar

UIN Imam Bonjol Padang

haryudinizar@uinib.ac.id

Abstract

This study investigates the effect of EdBot, an AI-powered digital assessment platform, on seventh-grade students' literal reading comprehension at MTsN 4 Padang. A quasi-experimental post-test only control group design was employed with cluster random sampling, resulting in Class VII.1 (n=31) as the experimental group and Class VII.4 (n=32) as the control group. Both groups received identical guided reading instruction using descriptive texts but differed in assessment method: EdBot for the experimental class and conventional paper-based tests for the control class. The instrument comprised 42 multiple-choice items measuring five literal reading comprehension indicators — specific information, main idea, vocabulary in context, inference, and reference — based on Brown's (2004) framework. Data were analyzed using descriptive statistics, Shapiro-Wilk normality test, Levene's homogeneity test, and Independent Samples T-Test via SPSS version 25. The experimental class achieved a mean score of 80.29 (Good), compared to 70.38 (Fair) in the control class. The T-Test yielded $t = -4.274$ with $\text{sig.} = .000$ ($p < 0.05$), confirming that EdBot has a significant positive effect on students' literal reading comprehension performance. It is concluded that EdBot is an effective digital assessment medium that enhances engagement, provides real-time feedback, and fosters a low-anxiety testing environment at the junior high school level.

Keywords: *EdBot, digital assessment, literal reading comprehension, descriptive text, quasi-experimental*

A. INTRODUCTION

In an era characterized by the rapid flow of digital information, reading skills are no longer merely regarded as academic abilities but as essential survival tools. Students are required to possess strong literal reading comprehension to filter and understand accurate information, thereby avoiding the trap of hoaxes and misinformation. This research is expected to enhance students' literal reading comprehension while simultaneously preparing them to face the challenges of digital literacy in the 21st century.

Reading is a highly fundamental skill in education and one of the main keys to the exploration of knowledge across various disciplines. According to Grellet (1981), reading is a complex process requiring individuals to employ a variety of skills, including letter recognition, vocabulary comprehension, and the ability to draw conclusions and inferences from the text. Reading comprehension encompasses more than just the literal understanding of sentences; it also includes the ability to analyze, evaluate, and integrate information contained in texts (Afflerbach et al., 2008).

One emerging technology with significant potential in the assessment context is EdBot — an Artificial Intelligence-based interactive assessment platform specifically designed to evaluate students' reading comprehension through an innovative and adaptive approach. As an educational chatbot, EdBot not only presents reading texts in various genres but also integrates an interactive questioning system capable of providing immediate and detailed feedback to students. The primary advantage of EdBot lies in its ability to analyze student responses in real-time and provide personalized evaluation based on individual comprehension levels.

Despite the potential of digital assessment tools, teaching reading at the junior high school level faces numerous challenges. Krashen (2004) notes that low reading interest can negatively impact the development of reading skills. Various studies have shown that traditional approaches to teaching reading are sometimes unable to meet the diverse learning needs of students (Pressley, 2000). Field observations indicate that when students are confronted with long and complex texts, many feel overwhelmed, illustrating the need for more innovative approaches in both teaching and assessing reading comprehension.

Based on previous research, the utilization of technology in the learning process has been proven to increase student motivation and engagement (Miller, 2015). Nevertheless, most existing studies have not specifically explored the effectiveness of EdBot in the context of reading comprehension assessment at the junior high school level in Indonesia. Therefore, this study aims to investigate the extent to which EdBot can help improve literal reading comprehension among seventh-grade students at MTsN 4 Padang. The research question addressed is: Does EdBot give a significant effect on students' literal reading comprehension assessment for seventh grade in Junior High School?

B. RESEARCH METHOD

1. Research Design

This research employed a quasi-experimental design with a Post-Test Only Control Group Design. Two class groups were selected through cluster random sampling: one as the experimental class and the other as the control class. As stated by Creswell (2012), the post-test-only control group design compares post-test results for two groups — one receiving the treatment and one that does not — in order to determine whether the intervention produced a significant effect. The experimental class received EdBot-based digital assessment, while the control class used conventional paper-based tests. Both groups received identical guided reading instruction using descriptive texts across four meetings before the post-test

2. Population and Sample

The population of this study comprised all seventh-grade students at MTsN 4 Padang, totaling 363 students distributed across 12 classes (VII.1 through VII.12). Using cluster random sampling via lottery, two classes were selected as the research sample. Class VII.1 (31 students) was assigned as the experimental group and Class VII.4 (32 students) as the control group, giving a total of 63 respondents.

3. Instructional Treatment

Both groups received four meetings of guided reading instruction based on the Guided Reading Technique by Allen (2000), grounded in Vygotsky's (1978) Zone of Proximal Development theory. The instruction covered descriptive texts on the topic 'Introducing People' at the A1 proficiency level. Meeting 1 introduced descriptive text structure; Meeting 2 focused on main idea and specific information; Meeting 3 addressed vocabulary in context and references; and Meeting 4 provided independent practice. Only in the final post-test session (Meeting 5) did the groups differ: the experimental class used EdBot while the control class completed a conventional paper-based test.

4. Research Instrument

The research instrument consisted of 42 multiple-choice items based on six descriptive texts covering familiar social contexts (party, neighborhood, language school, new town). All items were aligned with Brown's (2004) five literal reading comprehension indicators: (1) Specific Information — 31 items; (2) Main Idea — 3 items; (3) Vocabulary in Context — 2 items; (4) Making Inferences — 2 items; and (5) Identifying References — 4 items. Each item offered four options (A, B, C, D) with one correct answer.

5. Scoring Procedures

For the experimental class using EdBot, each correct item was awarded 100 points with a maximum cumulative score of 4,200 points, converted to a 0–100 scale using the formula: $\text{Final Score} = (\text{Total Points} \div 4,200) \times 100$. For the control class, each correct answer received 1

point, converted using: Final Score = (Total Correct ÷ 42) × 2.38. Student performance was classified into five categories based on Brown's (2004) criterion-referenced scale: Excellent (86–100), Good (71–85), Fair (56–70), Poor (41–55), and Very Poor (< 40)

6. Data Analysis

Data were analyzed through three stages: (1) Descriptive statistics, including mean, standard deviation, minimum, maximum, and range; (2) Pre-requisite tests, including the Shapiro-Wilk normality test and Levene's homogeneity of variances test; and (3) Hypothesis testing using the Independent Samples T-Test. All analyses were conducted using SPSS version 25 for Windows, with a significance threshold of $\alpha = 0.05$.

C. FINDING

The result of the analysis presents the descriptive statistics, the normality test, the homogeneity test and hypothesis test. The first is the descriptive test. Table 1 presents the descriptive statistics of post-test scores for both groups. The experimental class ($n = 31$) achieved a mean score of 80.29 (SD = 10.33), with a minimum of 57 and a maximum of 98. The control class ($n = 32$) obtained a mean score of 70.38 (SD = 7.97), with a minimum of 59.5 and a maximum of 90.4. The mean difference between the two groups was approximately 9.91 points, indicating that the experimental class consistently outperformed the control class across all performance measures.

Table 1. Descriptive Statistics of Post-Test Scores

Group	N	Mean	SD	Min	Max	Range
Experimental	31	80.29	10.33	57.0	98.0	41.0
Control	32	70.38	7.97	59.5	90.4	30.9

The experimental class exhibited a wider score range (41.0) and a larger standard deviation (10.33) compared to the control class (range: 30.9, SD: 7.97), suggesting greater variability among students in the experimental group while still achieving higher overall performance. The control class showed a more uniform distribution, with a slight positive skewness (0.708), indicating more students scored below the mean, whereas the experimental class showed a slight negative skewness (−0.198), suggesting more students scored above the mean.

Analysis of the control class data by comprehension indicator revealed that students performed best on inference items (mean: 3.03 out of 4) and most poorly on vocabulary in context and reference items (mean: 1.50 each out of 2). Specific information items yielded a

mean of 21.56 out of 31, and main idea items averaged 2.13 out of 3. These patterns highlight that vocabulary and referential comprehension were areas of particular difficulty for students assessed through conventional methods.

After having the descriptive test results, then the normality test is needed. The Shapiro-Wilk normality test was conducted on post-test data from both groups. For the experimental class ($df = 31$), the Shapiro-Wilk statistic was .976 with a significance value of .696 ($p > 0.05$). For the control class ($df = 32$), the statistic was .934 with a significance value of .052 ($p > 0.05$). Since both significance values exceed the 0.05 threshold, the data from both groups are confirmed to be normally distributed, satisfying the prerequisite for parametric statistical testing.

In finding the variance of the data, Levene's Test for Equality of Variances was conducted using four approaches (based on Mean, Median, Median with adjusted df , and Trimmed Mean). The Levene statistics ranged from 1.664 to 1.828, with significance values of .181 to .202 — all exceeding the 0.05 threshold. Consequently, the null hypothesis of equal variances is retained, confirming that the two groups' score variances are homogeneous. This result validates the use of the Equal Variances Assumed row in the subsequent Independent Samples T-Test.

Last, the result of this research was from conducting the hypothesis testing. The Independent Samples T-Test yielded a t -value of -4.274 ($df = 61$) with a two-tailed significance value of .000 ($p < 0.05$). The mean difference between groups was -9.9122 , with a standard error of 2.3194 and a 95% Confidence Interval ranging from -14.5500 to -5.2744 . Additionally, the One-Way ANOVA produced an F -statistic of 18.264 (Sig. = .000), providing further confirmation of a statistically significant difference between groups. Based on these results, the null hypothesis (H_0) is rejected and the alternative hypothesis (H_1) is accepted: EdBot has a significant positive effect on students' literal reading comprehension performance.

D. DISCUSSION

The findings of this study reveal that seventh-grade students assessed using EdBot achieved a mean score of 80.29 (categorized as Good), significantly higher than the control group's mean of 70.38 (Fair), with a difference of approximately 9.91 points. This result confirms that EdBot, as an AI-powered digital assessment medium, exerts a meaningful and positive influence on students' literal reading comprehension performance at the junior high school level.

The superior performance of the experimental class can be attributed to several key features of the EdBot platform. First, the immediate and personalized feedback mechanism enabled students to actively monitor their comprehension and make real-time adjustments

during the assessment. This aligns with the formative assessment principles advanced by Black and Wiliam (1998), who demonstrated that timely feedback significantly enhances learning outcomes by promoting self-regulated learning and deeper comprehension. When students receive instant corrective feedback, they are more likely to recognize their errors and consolidate correct understanding before the session ends.

Second, the gamification elements and conversational interface embedded in EdBot appear to have reduced assessment anxiety and fostered greater engagement. As noted by Cassady and Johnson (2002), assessment anxiety can compromise test validity. By transforming the evaluation experience into an interactive, low-stakes activity, EdBot created a more authentic representation of students' actual reading abilities. This is consistent with findings by Miller (2015) and Wang and Chen (2022), who demonstrated that digital engagement positively influences student motivation and performance in EFL reading contexts.

Third, the adaptive nature of EdBot's assessment environment, grounded in AI-driven natural language processing (Luckin et al., 2016), allowed for individualized interaction that a conventional paper test cannot provide. This aligns with cognitive theories of reading comprehension — particularly the construction-integration model by Kintsch (1998) — which emphasizes the active construction of meaning from text. By engaging students in a conversational assessment format, EdBot arguably stimulated deeper cognitive processing than passive multiple-choice items on paper.

The analysis of the control class by comprehension indicator revealed notable weaknesses in vocabulary in context and reference identification (both averaging only 1.50 out of 2), suggesting that these dimensions require more explicit instructional attention in conventional classroom settings. The experimental group's wider score range and higher maximum score (98) further indicate that EdBot's interactive features benefited both stronger and weaker students, though variability in digital literacy may account for the broader spread of scores.

These findings corroborate and extend the results of related studies. Nuraini et al. (2025) found that students assessed through Edmodo, a technology-based formative assessment platform, demonstrated significantly better reading comprehension gains than those in conventional settings. Suhartoyo et al. (2025) similarly concluded that online classroom-based assessments are diagnostic, adaptive, and better aligned with instructional objectives. The present study contributes specifically to the Indonesian junior high school context, demonstrating that EdBot's effectiveness extends to A1-level descriptive text comprehension among seventh-grade EFL learners.

However, this study also acknowledges several limitations. EdBot's scoring system

displays results in raw points rather than percentage-based scores, questions are not automatically organized by comprehension indicator, and certain unanswered items are displayed at the session's end in a potentially confusing manner. These limitations necessitate careful teacher preparation, including manual alignment of items with comprehension indicators and thorough student orientation prior to assessment. Future implementations should address these interface issues to maximize EdBot's diagnostic value.

E. CONCLUSION

This study demonstrates that the use of EdBot as a digital assessment platform has a statistically significant and practically meaningful positive effect on seventh-grade students' literal reading comprehension at MTsN 4 Padang. The experimental class, assessed via EdBot, achieved a mean score of 80.29 (Good), compared to 70.38 (Fair) for the control class assessed through conventional paper-based tests. The Independent Samples T-Test ($t = -4.274$, $p = .000$) confirmed this difference at the 0.05 significance level, leading to acceptance of the alternative hypothesis.

The superior outcomes of the EdBot group can be attributed to the platform's core affordances: immediate AI-generated feedback, a gamified and low-anxiety interface, automated scoring, and adaptive question delivery. These features collectively create a more engaging and informative assessment environment than traditional methods, ultimately supporting stronger literal comprehension performance across all five measured indicators — specific information, main idea, vocabulary in context, simple inferences, and reference identification.

Based on these findings, English teachers at the junior high school level are encouraged to integrate EdBot as an alternative assessment tool for reading comprehension, with appropriate preparation for indicator alignment and student orientation. Future research should explore EdBot's effectiveness across different text types, grade levels, and larger samples, and consider pre-test and post-test designs to capture longitudinal learning gains. The study contributes empirical evidence supporting the broader integration of AI-powered assessment platforms in Indonesian EFL education and underscores the importance of technology-enhanced evaluation in developing students' reading competencies for the digital age.

F. REFERENCES

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61(5), 364–373.

- Allen, J. (2000). *Yellow brick roads: Shared and guided paths to independent reading 4–12*. Stenhouse Publishers.
- Baker, E. L. (2008). Formative assessment: A policy perspective. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 21–36). Routledge.
- Barrett, T. C. (1976). Taxonomy of reading comprehension. In R. Smith & T. C. Barrett (Eds.), *Teaching reading in the middle grades*. Addison-Wesley.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270–295.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education.
- Gonzalez, M., & Cunnings, I. (2019). The role of motivation and digital tools in EFL reading comprehension. *Journal of English Language Teaching*, 8(2), 135–145.
- Grellet, F. (1981). *Developing reading skills: A practical guide to reading comprehension exercises*. Cambridge University Press.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Krashen, S. D. (2004). *The power of reading: Insights from the research* (2nd ed.). Libraries Unlimited.