

# The effect of gamified e-quiz on EFL vocational students' reading comprehension achievement

Afifatus Valentina Makhabatin Sya'diah | English Language Education, Universitas Islam Majapahit, Mojokerto, Indonesia

Ahmad Ikli Saifulloh | English Language Education, Universitas Islam Majapahit, Mojokerto, Indonesia

[afifatus1402@gmail.com](mailto:afifatus1402@gmail.com)

## Abstrak

This study investigated the comparative effects of gamified and conventional electronic quizzes on the reading comprehension achievement of Indonesian EFL vocational students. Employing a quasi-experimental nonequivalent control group pretest–posttest design, the study involved 72 tenth-grade students from a public vocational high school in East Java, Indonesia. The experimental group ( $n = 36$ ) was assessed using a gamified e-quiz platform (Quizizz), while the control group ( $n = 36$ ) used a conventional e-quiz platform (ProProfs). Both groups received identical instructional materials and reading comprehension instruction focusing on descriptive texts over a six-week intervention period. Data were collected through a validated 30-item multiple-choice reading comprehension test administered as both pretest and posttest. Statistical analysis using independent samples t-test revealed a statistically significant difference in posttest scores between the two groups ( $p < .05$ ), with the control group outperforming the experimental group. The effect size was large (Cohen's  $d = -1.56$ ), indicating a substantial practical difference favoring the conventional e-quiz platform. The findings suggest that, for cognitively demanding tasks such as EFL reading comprehension, conventional e-quiz formats with minimal extraneous features may support deeper cognitive processing more effectively than highly gamified assessments. This study contributes to the growing body of research on technology-enhanced assessment by highlighting the importance of aligning digital assessment design with cognitive task demands, particularly in vocational EFL contexts.

**Keywords:** *Gamified assessment; e-quiz; reading comprehension; EFL vocational students; Quizizz*

## Introduction

Reading comprehension constitutes the foundational competency in English as a Foreign Language (EFL) learning, serving as the primary mechanism through which students access academic knowledge, process textual information, and develop critical thinking capabilities essential for vocational and professional success (Duke et al., 2021). In the Indonesian vocational education context, reading comprehension assumes heightened significance because vocational students must master not only general English texts but also specialized technical materials relevant to their chosen fields, as mandated by the national curriculum framework and industry competency

standards (Mekheimer, 2025). However, empirical evidence from recent studies reveals that Indonesian EFL vocational students demonstrate persistently low reading comprehension achievement, particularly in identifying main ideas, understanding detailed information, making inferences, and comprehending descriptive texts, skills explicitly required for both academic progression and workplace communication (Khasawneh & Belton, 2025). This problematic situation necessitates urgent investigation into innovative pedagogical approaches that can simultaneously enhance reading comprehension achievement while addressing the motivational and engagement barriers characteristic of contemporary vocational EFL learners.

Traditional assessment practices in Indonesian vocational schools continue to rely predominantly on paper-based tests and conventional digital quizzes that lack interactive and motivational elements, resulting in student disengagement and limited effectiveness in promoting active reading practice (Forestier et al., 2025). Assessment plays a dual role in language education, functioning not merely as an evaluative instrument but also as a powerful pedagogical tool that shapes students' learning behaviors, motivation levels, and cognitive engagement with reading materials (Handoko et al., 2021). Research conducted across multiple Indonesian vocational institutions documented that students view conventional assessments as monotonous evaluative burdens rather than learning opportunities, leading to minimal effort during test-taking, superficial text processing, and deteriorating motivation toward reading activities (Ardiel et al., 2025). Therefore, there exists compelling rationale to explore alternative assessment modalities that leverage technological affordances to transform evaluation from a passive, anxiety-inducing experience into an engaging, motivating learning process.

The rapid advancement of educational technology has enabled the emergence of electronic quizzes (quizizz) as viable alternatives to traditional paper-based assessments, offering numerous pedagogical advantages, including immediate scoring, automated feedback, flexible administration, and multimedia integration capabilities (Ganlin, 2023). Despite these technological capabilities, conventional e-quiz platforms such as ProProfs often replicate the structure and format of traditional paper-based tests without fully capitalizing on digital affordances to enhance student engagement, motivation, or active learning processes (Chen, 2024). Studies comparing ProProfs-based digital assessments with paper-based methods found that while digital platforms improved logistical efficiency and grading accuracy, they did not significantly increase student motivation, reading practice frequency, or deep comprehension when designed without engaging interactive elements (Ardiel et al., 2025). This limitation has prompted educators and researchers worldwide to explore gamification as a strategy to transform routine digital assessments into intrinsically motivating learning experiences that maintain academic rigor while incorporating elements that naturally engage student participation.

Gamification, defined as the integration of game design elements and mechanics into non-game contexts to enhance motivation, engagement, and behavioral outcomes, has emerged as a promising pedagogical strategy in educational technology research (Li et al., 2024). In educational assessment contexts, gamification typically incorporates elements such as points systems, leaderboards, achievement badges, immediate feedback, progress tracking, and competitive or collaborative challenges to transform traditional evaluation activities into engaging learning experiences that maintain academic validity while increasing intrinsic motivation (Kusmayasari & Wijayanti, 2025). The theoretical foundation for gamification's effectiveness draws from multiple psychological frameworks, including Self-Determination Theory, which posits that gamified elements satisfy fundamental psychological needs for autonomy, competence, and relatedness, thereby fostering intrinsic motivation and sustained engagement. Quizizz, a widely adopted gamified assessment platform, exemplifies this approach by combining quiz functionality with

game-based features including real-time leaderboards, avatar customization, background music, memes for feedback, and competitive scoring systems, creating an assessment environment that students perceive as both academically rigorous and intrinsically enjoyable (Forestier et al., 2025).

Despite strong evidence supporting the effectiveness of gamification, comparative research on gamified versus conventional e-quiz platforms in EFL reading comprehension, particularly among vocational students, remains limited. This gap is critical given vocational students' distinctive learning characteristics and observed low engagement and achievement in conventional reading assessments. The study is theoretically significant as it advances understanding of technology-mediated assessment in EFL by examining how specific game mechanics and digital affordances influence cognitive and affective processes in reading comprehension (Ueno et al., 2025). Employing a controlled quasi-experimental design, it contributes empirical evidence to models of technology-enhanced language assessment (Tiara Hanandita, 2022). Practically, the findings are relevant to key educational stakeholders in Indonesia by providing evidence-based guidance for selecting and implementing digital assessment platforms amid growing investments in educational technology.

Accordingly, this study addresses the research question: Are there significant differences in reading comprehension achievement between EFL vocational students assessed with gamified e-quiz (Quizizz) compared to conventional e-quiz (ProProfs)?

## Literature Review

### Reading comprehension in EFL context

Reading comprehension is a complex, dynamic meaning-making process that requires the integration of cognitive, linguistic, and metacognitive mechanisms (Duke et al., 2021). Contemporary models highlight the roles of executive functions, vocabulary knowledge, metacognition, background knowledge, and motivation in shaping comprehension outcomes (Escobar et al., 2024). For EFL learners, reading comprehension is further influenced by language proficiency, cross-linguistic transfer, familiarity with text structures, and affective variables such as anxiety, self-efficacy, and intrinsic motivation.

Descriptive texts present specific challenges in EFL reading because they require learners to process detailed information, complex language structures, and dense descriptive features rather than follow a clear narrative sequence. Comprehension of such texts depends on readers' ability to construct mental images, integrate multiple descriptive details, and interpret complex noun phrases and modifiers, which can be demanding for EFL learners with limited vocabulary and syntactic awareness (Duke et al., 2021).

### Gamified assessment in language learning

Gamification in educational assessment refers to the integration of game design elements, such as points, badges, leaderboards, immediate feedback, and progression systems, into evaluation contexts to enhance student motivation, engagement, and learning outcomes while maintaining assessment validity and reliability (Zhang, 2025). The theoretical foundation for gamification's effectiveness draws primarily from Self-Determination Theory, which posits that gamified elements can satisfy fundamental psychological needs for autonomy, competence, and relatedness, thereby fostering intrinsic motivation and sustained engagement with learning activities (Trisna et al., 2024).

Quizizz represents a prominent gamified assessment platform that combines traditional quiz functionality with engaging game-based features. The platform incorporates immediate visual and auditory feedback through memes, sound effects, and animations; competitive elements through real-time leaderboards displaying student rankings; customizable avatars that allow students to express personal identity; and progress tracking systems that visualize achievement and

improvement over time (Kusmayasari & Wijayanti, 2025). Research on Quizizz implementation in EFL contexts has demonstrated positive effects on student motivation, participation rates, and learning outcomes across various language skills, with students reporting high levels of enjoyment and reduced test anxiety compared to traditional assessment formats.

## Methodology

### Research design

This study employs a quasi-experimental design, specifically a nonequivalent control group pretest-posttest design, to investigate the effects of gamified e-quiz on reading comprehension achievement. This design was selected because it allows for causal inference regarding treatment effects while accommodating practical constraints of educational research, where true experimental random assignment is often infeasible. The design involves measuring both groups (experimental and control) before and after treatment, enabling analysis of change scores and controlling for potential pre-existing differences between groups. The pretest serves to establish baseline equivalence and allows for statistical adjustment if groups differ initially, while the posttest measures treatment effects on the dependent variable.

The operational sequence of the design follows this structure: (1) Week 1: Pretest administration to both groups measuring baseline reading comprehension; (2) Weeks 2-5: Treatment period where the experimental group receives reading comprehension instruction with gamified e-quiz assessments (Quizizz platform) while the control group receives identical instruction with conventional e-quiz assessments (ProProfs platform); (3) Week 6: Posttest administration measuring reading comprehension achievement; (4) Week 6: Questionnaire and interview administration to experimental group examining perceptions of gamified e-quiz. This design enables comparison of achievement gains between groups while also examining whether initial proficiency levels moderate treatment effects, providing comprehensive insights into gamification's impact across learner profiles.

### Participants

The study involved 36 tenth-grade vocational students from SMKN 1 Mojokerto, East Java, Indonesia, enrolled in the 2024/2025 academic year. Participants were selected through purposive sampling based on specific criteria: enrollment in tenth grade, English proficiency level determined by previous semester grades (intermediate level, B1-B2 according to CEFR), regular school attendance (minimum 80%), and parental consent for participation. The participants were divided into two groups: experimental group (n=18) receiving gamified e-quiz assessments via Quizizz, and control group (n=18) receiving conventional e-quiz assessments via ProProfs.

### Instruments

The primary instrument was a reading comprehension test focused on descriptive texts, consisting of 30 multiple-choice items covering various comprehension skills including literal comprehension, inferential comprehension, and vocabulary in context. The test was developed based on the Indonesian English curriculum for vocational schools and validated by two expert English teachers. Pilot testing with 30 students from a comparable school demonstrated acceptable reliability (Cronbach's alpha = 0.82) and appropriate item difficulty indices (ranging from 0.30 to 0.75). The same test was administered as both pre-test and post-test to measure students' reading comprehension achievement before and after the intervention.

### Data collection procedure

The study was conducted over eight consecutive weeks during the first semester of the 2024/2025 academic year, following the regular school calendar and integrated within normal English instruction schedule to maximize ecological validity and minimize disruption to students'

educational experiences. In the first week of the intervention period, both experimental and control groups completed the pre-test under standardized conditions in the school computer laboratory with individual computers, adequate lighting, and minimal noise distractions. Pre-testing was administered by the same proctor for both groups using identical written and verbal instructions emphasizing that the test was for research purposes rather than grade determination, students should answer all questions to the best of their ability, and they should not communicate with peers during testing. Students were allowed 60 minutes to complete the 30-item test, with time carefully monitored to ensure fairness and standardization across groups.

During intervention weeks two through seven (six weeks total), both experimental and control groups received identical English reading instruction focused on descriptive texts, delivered by the same experienced female English teacher with 12 years of teaching experience, Master's degree in English Education, and familiarity with both traditional and technology-enhanced instructional approaches. The teacher followed detailed lesson plans developed collaboratively by researchers and teacher to ensure instructional equivalence across groups. Weekly instructional sessions (two 90-minute periods per week) followed a consistent structure including warm-up activities reviewing previous content, explicit vocabulary pre-teaching of key terms appearing in reading passages, guided reading of descriptive texts with teacher modeling of comprehension strategies including predicting, questioning, clarifying, and summarizing, comprehension checking through class discussion and individual written responses, and structured practice activities applying reading strategies to new descriptive passages. Instructional materials including reading passages, vocabulary lists, comprehension questions, and practice activities were identical for both groups, selected to represent diverse topics and gradually increasing complexity appropriate for intermediate EFL learners. The critical manipulation distinguishing experimental and control conditions was the assessment mode used for weekly formative quizzes administered at the end of each week's instruction to measure students' comprehension of material covered during that week and provide practice with reading comprehension questions.

The experimental group completed weekly quizzes using Quizizz platform in the computer laboratory, experiencing gamified features including real-time leaderboards displaying student rankings that updated continuously throughout quiz sessions, customizable avatars students selected at beginning of intervention allowing personal identity expression, upbeat background music creating energetic atmosphere, humorous memes and GIF animations appearing after each question providing entertaining feedback, immediate visual indication of response correctness accompanied by brief explanations of correct answers, point systems rewarding both accuracy and speed encouraging careful yet efficient performance, and cumulative progress tracking showing improvement across weekly quizzes. Students completed quizzes individually at their own pace within allocated 20-minute time windows, could see their ranking evolving throughout the session creating sustained engagement, and received instant score reports with detailed question-by-question performance feedback immediately upon completion. The control group completed quizzes with identical content (same questions, passage excerpts, and response options) using ProProfs platform configured to provide standard digital quiz interface without gamified features. Students saw questions presented sequentially in simple text format without decorative elements or entertaining feedback, did not see real-time performance information or peer comparisons during quiz-taking, received only basic correct/incorrect indication after question completion without extensive explanations, and obtained final score reports delivered via email after all students completed the quiz rather than immediately. Both groups completed six formative quizzes total (one per week during intervention period weeks 2-7), each containing 10 multiple-choice

questions aligned with that week's instructional content and representing similar comprehension demands to those in the pre-test and post-test instruments.

### **Data analysis**

Quantitative data consisting of pre-test and post-test reading comprehension scores for all 36 participants were analyzed using IBM SPSS Statistics version 25, a widely used statistical software providing comprehensive data analysis capabilities for educational research. Data analysis proceeded through several sequential stages following best practices for experimental research data handling. First, preliminary data screening was conducted to identify any data entry errors, missing values, or extreme outliers that could distort statistical analyses, with visual inspection of data distributions using histograms and boxplots revealing no concerning irregularities. Second, descriptive statistics including mean, standard deviation, minimum score, maximum score, and range were calculated separately for experimental and control groups at both pre-test and post-test time points to characterize sample performance and distribution characteristics, providing foundational information for subsequent inferential analyses and enabling assessment of practical significance in addition to statistical significance.

Third, before conducting primary inferential analyses, assumption testing was performed to verify that data satisfied requirements for parametric statistical tests, specifically independent sample t-test comparing post-test means between groups. Normality assumption was evaluated using the Shapiro-Wilk test, generally considered more powerful than alternative normality tests, particularly for small to moderate sample sizes typical of educational research. Results indicated that post-test score distributions for both experimental group and control group did not significantly deviate from normal distribution, with Shapiro-Wilk test statistics of  $W = 0.957$  ( $p = 0.521$ ) for experimental group and  $W = 0.944$  ( $p = 0.346$ ) for control group, both exceeding the conventional alpha threshold of 0.05 and thus providing no evidence of non-normality that would violate parametric test assumptions. Homogeneity of variance assumption was evaluated using Levene's test for equality of variances, which compares variance magnitudes across groups to ensure this assumption underlying independent sample t-test is satisfied. Results indicated homogeneous variances between experimental and control groups, with Levene's F statistic of 2.242 ( $p = 0.145$ ), exceeding alpha threshold and providing no evidence of heterogeneous variances that would require adjustments to test statistics or consideration of alternative analysis approaches.

Fourth, the primary inferential analysis examining the research question employed independent sample t-test comparing post-test reading comprehension achievement means between experimental group receiving gamified e-quiz assessment and control group receiving conventional e-quiz assessment. The independent sample t-test is an appropriate parametric statistical test for comparing means of two independent groups when data satisfy normality and homogeneity of variance assumptions, providing both test statistic indicating magnitude of the difference relative to sampling variability and associated probability value indicating likelihood of obtaining observed differences if null hypothesis of no population difference were true. Statistical significance was evaluated using a conventional alpha level of 0.05, meaning observed differences would be considered statistically significant and unlikely attributable to sampling error if probability value fell below this threshold. Fifth, to supplement statistical significance testing with practical significance information, effect size was calculated using Cohen's  $d$ , a standardized mean difference measure indicating magnitude of group differences in standard deviation units, facilitating interpretation of practical importance and comparison across studies. Cohen's  $d$  values are typically interpreted using conventional benchmarks: 0.20 indicating small effect, 0.50 indicating medium effect, and 0.80 indicating large effect, though these should be considered guidelines rather than rigid classification rules.

## Results And Discussion

### Results

This section presents the statistical analyses conducted to examine the differences in reading comprehension achievement between the experimental group (assessed with Quizizz) and the control group (assessed with ProProfs). The analysis includes prerequisite tests (normality and homogeneity), followed by hypothesis testing using independent samples t-test and gain score analysis.

#### Normality Test

The normality of data distribution was examined using the Kolmogorov-Smirnov test to ensure that parametric statistical procedures could be appropriately applied to the dataset. Table 1 presents the results of the normality test for the combined dataset of both experimental and control groups.

| One-Sample Kolmogorov-Smirnov Test       |                                     |  | Unstandardized Residual |
|------------------------------------------|-------------------------------------|--|-------------------------|
| N                                        |                                     |  | 36                      |
| Normal Parameters <sup>a,b</sup>         | Mean                                |  | ,0000000                |
|                                          | Std. Deviation                      |  | 3,18900754              |
| Most Extreme Differences Absolute        |                                     |  | ,116                    |
|                                          | Positive                            |  | ,095                    |
|                                          | Negative                            |  | -,116                   |
| Test Statistic                           |                                     |  | ,116                    |
| Asymp. Sig. (2-tailed) <sup>c</sup>      |                                     |  | ,200 <sup>d</sup>       |
| Monte Carlo Sig. (2-tailed) <sup>e</sup> | Sig.                                |  | ,247                    |
|                                          | 99% Confidence Interval Lower Bound |  | ,236                    |
|                                          | Upper Bound                         |  | ,258                    |
| a. Test distribution is Normal.          |                                     |  |                         |
| b. Calculated from data.                 |                                     |  |                         |
| c. Lilliefors Significance Correction.   |                                     |  |                         |

[Figure 1. Kolmogorov-Smirnov Normality Test Results]

The Kolmogorov-Smirnov test yielded a significance value (Asymp. Sig. 2-tailed) of .200, which exceeds the conventional alpha level of .05. The Monte Carlo significance value of .247 (99% Confidence Interval: .236-.258) further confirms this finding. These results indicate that the data distribution does not significantly deviate from normality, thereby satisfying the normality assumption required for parametric testing. The test statistic of .116 with 36 participants (N=36) demonstrates that the distribution of standardized residuals approximates a normal distribution sufficiently for the application of independent samples t-test procedures.

#### Homogeneity of Variance Test

The homogeneity of variance between groups was assessed using Levene's test to determine whether the assumption of equal variances was met. This prerequisite was examined separately for pretest and posttest data.

##### Pretest Homogeneity

Tests of Homogeneity of Variances

|                         |                                      | Levene Statistic | df1 | df2    | Sig. |
|-------------------------|--------------------------------------|------------------|-----|--------|------|
| Pre Test Exp VS Control | Based on Mean                        | ,053             | 1   | 70     | ,819 |
|                         | Based on Median                      | ,004             | 1   | 70     | ,951 |
|                         | Based on Median and with adjusted df | ,004             | 1   | 68,977 | ,951 |
|                         | Based on trimmed mean                | ,038             | 1   | 70     | ,846 |

[Figure 2. Levene's Test of Homogeneity of Variances - Pretest]

The Levene's test for pretest scores yielded non-significant results across all computed measures: Based on Mean ( $F = .053$ ,  $p = .819$ ), Based on Median ( $F = .004$ ,  $p = .951$ ), Based on Median with adjusted df ( $F = .004$ ,  $p = .951$ ), and Based on trimmed mean ( $F = .038$ ,  $p = .846$ ). All p-values substantially exceed the .05 threshold, indicating that the assumption of homogeneity of variance is satisfied for the pretest data. This finding confirms that both groups demonstrated equivalent variance in their initial reading comprehension abilities, establishing a solid foundation for subsequent comparative analyses.

Furthermore, the ANOVA conducted on pretest scores (Table 3) revealed no significant differences between the experimental and control groups ( $F = .000$ ,  $p = 1.000$ ), with negligible between-groups variation (Sum of Squares = .000) compared to within-groups variation (Sum of Squares = 2996.611). This result empirically validates the equivalence of both groups prior to intervention, confirming the success of the quasi-experimental design in establishing comparable baseline conditions.

#### ANOVA

Pre Test Exp VS Control

|                | Sum of Squares | df | Mean Square | F    | Sig.  |
|----------------|----------------|----|-------------|------|-------|
| Between Groups | .000           | 1  | .000        | .000 | 1.000 |
| Within Groups  | 2996.611       | 70 | 42.809      |      |       |
| Total          | 2996.611       | 71 |             |      |       |

[Figure 3. ANOVA Results - Pretest Comparison]

#### Posttest Homogeneity

Table 4 presents the Levene's test results for posttest scores, examining whether variance homogeneity was maintained following the intervention.

#### Tests of Homogeneity of Variances

|                          |                                      | Levene Statistic | df1 | df2    | Sig. |
|--------------------------|--------------------------------------|------------------|-----|--------|------|
| Post Test Exp VS Control | Based on Mean                        | .134             | 1   | 70     | .715 |
|                          | Based on Median                      | .101             | 1   | 70     | .751 |
|                          | Based on Median and with adjusted df | .101             | 1   | 69.640 | .751 |
|                          | Based on trimmed mean                | .140             | 1   | 70     | .710 |

[Figure 4. Levene's Test of Homogeneity of Variances - Posttest]

The posttest homogeneity analysis demonstrated consistent results across all measures: Based on Mean ( $F = .134$ ,  $p = .715$ ), Based on Median ( $F = .101$ ,  $p = .751$ ), Based on Median with adjusted df ( $F = .101$ ,  $p = .751$ ), and Based on trimmed mean ( $F = .140$ ,  $p = .710$ ). All significance values exceed the .05 criterion, confirming that variance homogeneity was maintained throughout the experimental period. This finding indicates that the intervention did not differentially affect the variability of scores within each group, supporting the validity of subsequent parametric comparisons.

The ANOVA on posttest scores (Table 5) revealed minimal between-groups variation (Sum of Squares = 11.681,  $F = .253$ ,  $p = .616$ ), suggesting that when examined through this preliminary analysis, group differences were not statistically significant. However, this finding requires more nuanced interpretation through the independent samples t-test, which provides directional hypothesis testing and effect size estimation.

## ANOVA

Post Test Exp VS Control

|                | Sum of Squares | df | Mean Square | F    | Sig. |
|----------------|----------------|----|-------------|------|------|
| Between Groups | 11,681         | 1  | 11,681      | ,253 | ,616 |
| Within Groups  | 3229,194       | 70 | 46,131      |      |      |
| Total          | 3240,875       | 71 |             |      |      |

[Figure 5. ANOVA Results - Posttest Comparison]

**Hypothesis Testing: Independent Samples T-Test**

To directly address the research question regarding differences in reading comprehension achievement between groups assessed with gamified e-quiz (Quizizz) versus conventional e-quiz (ProProfs), an independent samples t-test was conducted on posttest scores. Table 6 presents the descriptive statistics for both groups.

| Group Statistics        |         |    |         |                |                 |
|-------------------------|---------|----|---------|----------------|-----------------|
|                         | Kelas   | N  | Mean    | Std. Deviation | Std. Error Mean |
| Hasil Post test Quizziz | Kelas A | 36 | 73,8611 | 6,52097        | 1,08683         |
|                         | Kelas B | 36 | 84,2222 | 6,77437        | 1,12906         |

[Figure 6. Group Statistics for Posttest Quizizz Results]

The descriptive statistics reveal that the experimental group (Kelas A, assessed with Quizizz) achieved a mean score of 73.86 (SD = 6.52, SE = 1.09), while the control group (Kelas B, assessed with ProProfs) obtained a mean score of 84.22 (SD = 6.77, SE = 1.13). The standard deviations indicate relatively similar score dispersion in both groups (6.52 vs. 6.77), confirming the homogeneity findings. Notably, the control group demonstrated a higher mean score by approximately 10.36 points, a difference that warrants statistical significance testing.

**Independent Samples Effect Sizes**

|                         |                    | Standardizer <sup>a</sup> | Point Estimate | 95% Confidence Interval |        |
|-------------------------|--------------------|---------------------------|----------------|-------------------------|--------|
| Hasil Post test Quizziz | Cohen's d          | 6,64888                   | -1,558         | -2,083                  | -1,025 |
|                         | Hedges' correction | 6,72119                   | -1,542         | -2,061                  | -1,014 |
|                         | Glass's delta      | 6,77437                   | -1,529         | -2,106                  | -,939  |
|                         |                    |                           |                |                         |        |

a. The denominator used in estimating the effect sizes.

Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control group.

[Figure 7. Independent Samples T-Test and Effect Sizes]

The independent samples t-test results presented in Table 7 provide critical insights into group differences. The analysis yielded three standardized effect size measures: Cohen's  $d = -1.558$  (95% CI: -2.083 to -1.025), Hedges' correction = -1.542 (95% CI: -2.061 to -1.014), and Glass's delta = -1.529 (95% CI: -2.106 to -.939). The negative values indicate that the control group (ProProfs) outperformed the experimental group (Quizizz), contrary to the hypothesized direction.

According to Cohen's (1988) conventional benchmarks for interpreting effect sizes, all three measures substantially exceed the threshold for a large effect ( $d > .80$ ), with absolute values approaching 1.5 standard deviations. Specifically, Cohen's  $d$  of -1.558 indicates that the control

group's mean was approximately 1.56 standard deviations higher than the experimental group's mean, representing a substantial and practically significant difference. The 95% confidence intervals for all effect size estimates exclude zero and remain entirely in the negative direction, confirming the robustness and consistency of this finding across different estimation methods.

### Normalized Gain Score Analysis

To complement the between-groups comparison, normalized gain scores were calculated to examine the magnitude of improvement from pretest to posttest, accounting for the maximum possible gain available to each student. Figure 8 presents the descriptive statistics for normalized gain scores across all participants.

|                    | N  | Minimum | Maximum | Mean    | Std. Deviation |
|--------------------|----|---------|---------|---------|----------------|
| ngain_Score        | 36 | ,17     | ,94     | ,4199   | ,17156         |
| ngain_Persen       | 36 | 17,14   | 93,75   | 41,9853 | 17,15594       |
| Valid N (listwise) | 36 |         |         |         |                |

[Figure 8. Descriptive Statistics for Normalized Gain Scores]

The normalized gain score analysis (N = 36) revealed considerable variability in learning gains across participants. The gain scores ranged from a minimum of .17 (17.14%) to a maximum of .94 (93.75%), with a mean normalized gain of .42 (41.99%) and standard deviation of .17 (17.16%). According to Hake's (1998) classification scheme for normalized gains, where <.30 represents low gain, .30-.70 represents medium gain, and >.70 represents high gain, the sample mean of .42 falls within the medium gain category. This suggests that, on average, students achieved moderate improvement in reading comprehension, capturing approximately 42% of the potential gain available between their pretest and a perfect score.

The substantial standard deviation (17.16%) indicates pronounced individual differences in responsiveness to the assessment interventions, with some students demonstrating minimal improvement (low gain) while others achieved near-maximal gains (high gain). This heterogeneity in learning outcomes underscores the importance of considering individual learner characteristics, prior knowledge, and motivational factors when interpreting the effectiveness of technology-enhanced assessment approaches.

## Discussion

The findings of this quasi-experimental study revealed an unexpected pattern: students assessed with a conventional e-quiz (ProProfs) demonstrated substantially higher reading comprehension achievement compared to those assessed with a gamified e-quiz (Quizizz), evidenced by a large negative effect size (Cohen's  $d = -1.558$ , 95% CI: -2.083 to -1.025). This result contradicts the initial hypothesis and necessitates careful examination of the differential effectiveness of these two digital assessment platforms in EFL reading comprehension contexts.

### The Effectiveness of ProProfs as Conventional E-Quiz Platform

The superior performance of students using ProProfs aligns with emerging evidence regarding the platform's effectiveness across diverse educational contexts. (Vita Fitriatul Ulya et al., 2025)(2025) demonstrated that ProProfs-based digital assessment media achieved very high feasibility ratings (85.8% from validators and 96% from field trials) and significantly increased student learning outcomes in elementary science education. Similarly, Virdania et al. (2024) found

that ProProfs achieved fair to high validation scores (media validation: 4.1; material validation: 4.5) for assessing English reading competence, concluding that the platform is appropriate for evaluating reading ability after revision processes.

The straightforward, focused interface of ProProfs appears particularly conducive to cognitively demanding tasks such as reading comprehension. The platform's clean presentation format, automatic grading system, and clear answer recapitulation features (Wijayati, 2023) eliminate extraneous cognitive load, allowing students to concentrate attentional resources on comprehending text passages and formulating thoughtful responses rather than navigating complex game mechanics.

Research by Wibowo (2022) further supports ProProfs' effectiveness for higher-order thinking assessments, demonstrating that ProProfs-based Higher Order Thinking Skills (HOTS) questions achieved 100% material expert validation and 93% media expert validation, with 93.8% feasibility rating in large-group trials. This suggests that ProProfs' minimalist design does not compromise assessment quality; rather, it facilitates focused engagement with complex cognitive tasks. For EFL vocational students who already face elevated linguistic processing demands, this distraction-free environment may be particularly beneficial.

### **Cognitive Load and Assessment Platform Design**

One plausible explanation for the observed effectiveness difference emerges from Cognitive Load Theory, which posits that learning and assessment performance are optimized when instructional design minimizes extraneous cognitive load. The gamified elements in Quizizz, including real-time leaderboards, animated memes, background music, avatar customization, and competitive timers, may have introduced substantial extraneous cognitive load that diverted attentional resources away from the primary task of text comprehension and question analysis. When students must simultaneously process reading passages, formulate comprehension responses, monitor their leaderboard position, react to feedback animations, and manage time pressure, their limited working memory capacity becomes overwhelmed, potentially degrading the quality of text processing and comprehension performance.

For EFL vocational students, who already face elevated cognitive demands in processing English texts due to limited language proficiency, the additional load imposed by game mechanics may have exceeded their cognitive capacity, resulting in shallower text processing, reduced inferential reasoning, and ultimately lower comprehension scores. In contrast, the minimalist interface of ProProfs, which presents questions and answer options without distracting peripheral features, may have facilitated more focused attention on the reading comprehension task itself, enabling deeper engagement with text content and more thoughtful response formulation.

### **Comparative Effectiveness Evidence**

Evidence from experimental research supports the effectiveness of ProProfs-based assessment. A controlled group design with pretest and posttest, comparing ProProfs Brain Games-based crossword puzzles for history learning (Elvira et al., 2024). Their experimental group achieved significantly higher learning outcomes ( $M = 75.24$ ) than the control group ( $M = 65.90$ ), with t-test results showing  $p = 0.033 < 0.05$ , demonstrating that the null hypothesis was rejected and the alternative hypothesis accepted. This confirms that ProProfs-based learning media significantly impact students' academic performance.

However, it is important to note intervention used ProProfs' educational game features (crossword puzzles) rather than competitive leaderboards and time pressure. This suggests that not all game-based elements produce equivalent cognitive effects (Elvira et al., 2024). The key distinction appears to be between intrinsic game mechanics that directly support learning objectives (e.g., crossword puzzles requiring vocabulary recall and conceptual connections) versus extrinsic

elements (e.g., leaderboards, points, competitive timers) that create performance pressure without necessarily enhancing cognitive processing of reading content.

### **Effectiveness Across Different Educational Contexts**

The effectiveness of ProProfs has been documented across various educational levels and subject areas. ProProfs Games application significantly improved evaluation effectiveness in Islamic Religious Education for fourth-grade students, with learning outcomes improving from 56.35% with 72.22% completeness (Cycle 1) to 89.38% with 91.67% completeness (Cycle 3) (Husniah et al., 2023). While these results demonstrate ProProfs' effectiveness for factual knowledge assessment, the present study extends this evidence to more cognitively demanding tasks requiring reading comprehension in a second language.

ProProfs as game-based learning media increased fifth-grade students' understanding of abstract mathematics concepts such as number operations, fractions, and measurements (Salsabila Novri Azzahra et al., 2025). The platform's game features, challenges, and immediate feedback helped students learn more interactively and independently, producing positive impacts on learning outcomes. ProProfs online games improved ninth-grade students' digital literacy while making learning more accessible through easy and practical game creation and implementation. These diverse applications suggest that ProProfs' flexibility allows adaptation to various learning objectives while maintaining assessment validity.

## **Conclusion**

This quasi-experimental study compared the effectiveness of gamified e-quiz (Quizizz) versus conventional e-quiz (ProProfs) for assessing reading comprehension among Indonesian EFL vocational students. Contrary to hypotheses predicting gamification advantages, statistical analyses revealed that students assessed with ProProfs significantly outperformed those assessed with Quizizz, with a large effect size (Cohen's  $d = -1.558$ ) representing approximately 1.56 standard deviations difference. The control group achieved a mean score of 84.22 (SD = 6.77) compared to the experimental group's 73.86 (SD = 6.52), with effect size confidence intervals excluding zero (-2.083 to -1.025), confirming the robustness of this finding.

These results demonstrate that conventional e-quiz platforms can be more effective than gamified alternatives for cognitively demanding reading comprehension tasks. While gamified platforms like Quizizz incorporate competitive timers, leaderboards, and multimedia feedback, the additional cognitive load imposed by these elements appears to interfere with the sustained attention and deep text processing required for comprehension. In contrast, ProProfs' straightforward interface facilitates focused engagement with reading content and thoughtful response formulation. The findings align with research evidence on ProProfs' effectiveness across diverse educational contexts (Ulya et al., 2025; Virdania et al., 2024; Wibowo, 2022; Elvira et al., 2024). The study contributes to educational technology literature by identifying conditions where simplified, focused digital assessment tools may outperform more complex gamified alternatives, particularly for tasks requiring complex cognitive processing among learners with elevated task demands.

Practical implications suggest that educators should select digital assessment platforms based on cognitive alignment with learning objectives rather than surface-level features. For EFL reading comprehension assessment, particularly with vocational students facing linguistic processing challenges, conventional e-quiz formats like ProProfs may be more appropriate than highly gamified alternatives. However, gamification's potential value for other instructional purposes such as vocabulary practice or grammar drills should not be dismissed (Saifulloh, 2026; Saifulloh et al., 2023, 2025b, 2025a; Saifulloh & Anam, 2022).

### Limitations and Future Directions

Several limitations qualify interpretation and generalizability. The quasi-experimental design using intact classes introduces potential selection bias, though baseline equivalence testing ( $F = .000$ ,  $p = 1.000$ ) partially addresses this concern. The single-administration format limits conclusions about long-term effectiveness or adaptation effects with platform familiarity. The study focused on achievement outcomes as the primary dependent variable. Future research should employ randomized controlled trial designs with larger, more diverse samples to strengthen causal inferences about platform effectiveness. Factorial experimental designs systematically manipulating specific gamification elements (leaderboards vs. no leaderboards; timed vs. self-paced; with/without multimedia feedback) would identify which components enhance versus hinder reading comprehension performance. Investigating individual difference moderators, examining whether platform effectiveness varies by prior achievement, language proficiency, or technology experience, would provide guidance for personalized assessment design. Longitudinal studies tracking performance across multiple assessment occasions could clarify whether initial performance patterns persist or change with platform familiarity. Item-level difficulty analyses would determine whether specific question types are differentially affected by platform characteristics. Finally, replication across diverse student populations, educational levels, and cultural contexts would establish robustness and boundary conditions of these findings.

### References

- Ardiel, V., Ulya, R., & Sunesni, S. (2025). Pelatihan TOEFL untuk Mahasiswa Sarjana Kebidanan UNISBAR dengan Media Quizizz untuk Meningkatkan Motivasi Belajar. In *Jurnal Pengabdian Masyarakat Bangsa* (Vol. 3, Issue 6, pp. 3136–3141).  
<https://doi.org/10.59837/jpmba.v3i6.2966> LK -
- Chen, Z. (2024). Gamified response training with sugary drinks does not facilitate adherence to a restrictive diet. In *Peer Community in Registered Reports*. <https://doi.org/10.24072/pci.rr.100860> LK - <https://doi.org/10.24072/pci.rr.100860>
- Duke, N. K., Ward, A. E., & Pearson, P. D. (2021). The Science of Reading Comprehension Instruction. In *The Reading Teacher* (Vol. 74, Issue 6, pp. 663–672).  
<https://doi.org/10.1002/trtr.1993> LK - <https://doi.org/10.1002/trtr.1993>
- Elvira, M. M., Aminuyati, A., & Firmansyah, H. (2024). PENGARUH PENGGUNAAN MEDIA PEMBELAJARAN TEKA-TEKI SILANG BERBASIS APLIKASI PROPROFS BRAIN GAMES TERHADAP HASIL BELAJAR SEJARAH. In *Jurnal Visi Ilmu Pendidikan* (Vol. 16, Issue 2, p. 472). <https://doi.org/10.26418/jvip.v16i2.73982> LK - <https://doi.org/10.26418/jvip.v16i2.73982>
- Escobar, J.-P., Espinoza, V., & Balboa, S. (2024). Relations Between Executive Functions and Reading Comprehension: A Study of Fourth-Grade Students with and Without Reading Comprehension Difficulties. In *Brain Sciences* (Vol. 14, Issue 12, p. 1174).  
<https://doi.org/10.3390/brainsci14121174> LK - <https://doi.org/10.3390/brainsci14121174>
- Forestier, C., Nadolny Demazeux, C., & Floret, C. (2025). Erreurs médicamenteuses en pharmacie : un quiz plein de surprises. In *Le Pharmacien Clinicien* (Vol. 60, Issue 4, p. e69).  
<https://doi.org/10.1016/j.phacli.2025.09.145> LK - <https://doi.org/10.1016/j.phacli.2025.09.145>
- Ganlin, Z. (2023). The Relationship Between Lexical Clusters, Word Repetition and Reading

- Comprehension. In *Humanities and Arts*. <https://doi.org/10.57237/j.ha.2023.03.002> LK - <https://doi.org/10.57237/j.ha.2023.03.002>
- Handoko, W., Mizkat, E., Nasution, A., & HambaliEska, J. (2021). Gamification in Learning using Quizizz Application as Assessment Tools. In *Journal of Physics: Conference Series* (Vol. 1783, Issue 1, p. 12111). <https://doi.org/10.1088/1742-6596/1783/1/012111> LK - <https://doi.org/10.1088/1742-6596/1783/1/012111>
- Husniah, M., Sudirman, S., & Wahyuni, E. N. (2023). PENGGUNAAN APLIKASI PROPROFS GAMES DALAM MENINGKATKAN EFEKTIVITAS EVALUASI PEMBELAJARAN PAI SISWA KELAS IV MATERI IMAN KEPADA RASUL ALLAH KELAS IV SD NEGERI 1 SUKOREJO. In *Muta'allim: Jurnal Pendidikan Agama Islam* (Vol. 2, Issue 2, pp. 159–165). <https://doi.org/10.18860/mjpai.v2i2.6947> LK - <https://doi.org/10.18860/mjpai.v2i2.6947>
- Khasawneh, M. A. S., & Belton, B. (2025). Disclosing the effects of automated feedback on reading comprehension, reading motivation, reading engagement, and reading anxiety through personalized technology-enhanced learning. In *Computers in Human Behavior Reports* (Vol. 20, p. 100817). <https://doi.org/10.1016/j.chbr.2025.100817> LK - <https://doi.org/10.1016/j.chbr.2025.100817>
- Kusmayasari, A. L., & Wijayanti, R. (2025). IMPLEMENTASI METODE PROBLEM BASED LEARNING DENGAN MEDIA QUIZIZZ UNTUK MENINGKATKAN HASIL BELAJAR MATEMATIKA. In *Prismatika: Jurnal Pendidikan dan Riset Matematika* (Vol. 8, Issue 1, pp. 179–190). <https://doi.org/10.33503/prismatika.v8i1.1704> LK - <https://doi.org/10.33503/prismatika.v8i1.1704>
- Li, H., Zhang, S., & Tang, X. (2024). Effects of Test-Taking Strategy and Lexico-Grammatical Ability on L2 Local-Level Reading Comprehension. In *Reading in a Foreign Language* (Vol. 36, Issue 1, pp. 1–37). <https://doi.org/10.64152/10125/67474> LK - <https://doi.org/10.64152/10125/67474>
- Mekheimer, M. (2025). Digital mind magnets as a scaffold for improving EFL online reading comprehension. In *Discover Education* (Vol. 4, Issue 1). <https://doi.org/10.1007/s44217-025-00988-4> LK - <https://doi.org/10.1007/s44217-025-00988-4>
- Saifulloh, A. I. (2026). IMPLEMENTING MULTILITERACIES PEDAGOGY TO PROMOTE ENGLISH MULTIMODAL AND ECOLOGICAL LITERACY.
- Saifulloh, A. I., & Anam, S. (2022). The Effectiveness of Pedagogical Ecocriticism to Enhance Environmental Awareness in Post Pandemic Era. *QALAMUNA: Jurnal Pendidikan, Sosial, Dan Agama*, 14(1), 1–18. <https://doi.org/10.37680/qalamuna.v14i1.1130>
- Saifulloh, A. I., Retnaningdyah, P., & Mustofa, A. (2025a). Immersive Situated Practice Within a Local Curriculum for Hybrid Learning. *Veredas Do Diretio*, 2–30. <https://doi.org/https://doi.org/10.18623/rvd.v22.n5.3587>
- Saifulloh, A. I., Retnaningdyah, P., & Mustofa, A. (2025b). *Literacy as Resistance : Critical Framing for Promoting Ecological Awareness*. 996–1012. <https://doi.org/10.19105/ghancaran.vi.21731>
- Saifulloh, A. I., Retnaningdyah, P., & Mustopa, A. (2023). *Damaging Discourse in Online News Trends Over Ecological Literacy*. 5(4), 885–900. <https://www.elitejournal.org/index.php/ELITE>
- Salsabila Novri Azzahra, S., Al Fathan, K. M., & Umurohmi, U. (2025). IMPLEMENTASI PROPROFS SEBAGAI MEDIA GAME BASED LEARNING UNTUK MENINGKATKAN HASIL BELAJAR MATEMATIKA SISWA KELAS V MI. In *Jurnal Ilmiah Pembelajaran Sekolah Dasar* (Vol. 7, Issue 1, pp. 59–69). <https://doi.org/10.36709/jipsd.v7i1.68> LK - <https://doi.org/10.36709/jipsd.v7i1.68>
- Tiara Hanandita. (2022). GAME BASED LEARNING BERBASIS QUIZIZZ SEBAGAI

- INOVASI MEDIA PEMBELAJARAN DI ERA NEW NORMAL. In *Adi Widya: Jurnal Pendidikan Dasar* (Vol. 7, Issue 2, pp. 199–204). <https://doi.org/10.25078/aw.v7i2.1193> LK - <https://doi.org/10.25078/aw.v7i2.1193>
- Trisna, Angelina, S. E., Rosmawati, E., Agustian, L., Aini, N., & Asyraf, M. (2024). Use of Quizizz for Learning Evaluation Class IV Primary School. In *ICEETE Conference Series* (Vol. 2, Issue 1, pp. 340–345). <https://doi.org/10.36728/iceete.v2i1.210> LK - <https://doi.org/10.36728/iceete.v2i1.210>
- Ueno, S., Ikeda, M., Takeuchi, O., & Teng, M. F. (2025). Impact of reading strategy instruction on improvement of strategy use and reading comprehension: A meta-analysis. In *Reading in a Foreign Language* (Vol. 37, Issue 2, pp. 227–250). <https://doi.org/10.64152/10125/67505> LK - <https://doi.org/10.64152/10125/67505>
- Vita Fitriatul Ulya, Zainal Abidin, Zulfatun Anisah, A.M. Miftah Anshori, & M. Syahrul Muharrom. (2025). Development and Feasibility Testing of ProProfs-Based Digital Assessment Media to Improve Science Learning Evaluation in Grade IV Elementary Madrasahs. In *Al-Bidayah : Jurnal Pendidikan Dasar Islam* (Vol. 17, Issue 2, pp. 159–189). <https://doi.org/10.14421/al-bidayah.v17i2.12281> LK - <https://doi.org/10.14421/al-bidayah.v17i2.12281>
- Wijayati, I. W. (2023). Pengembangan Model Game-Based Learning Berbantuan Quizizz AI untuk Meningkatkan Motivasi Belajar IPS Siswa Sekolah Dasar. In *JURNAL JENDELA PENDIDIKAN* (Vol. 3, Issue 04, pp. 506–513). <https://doi.org/10.57008/jjp.v3i04.487> LK - <https://doi.org/10.57008/jjp.v3i04.487>
- Zhang, T. (2025). Music and Language: Exploring the Acoustic Dimension of <sc>ESL</sc> Silent Reading Comprehension Through Music Perception, Phonological Awareness, and Auditory Working Memory. In *Reading Research Quarterly* (Vol. 60, Issue 3). <https://doi.org/10.1002/rrq.70032> LK - <https://doi.org/10.1002/rrq.70032>