Journal of English Language Teaching, Literatures & Applied Linguistics (JELTLAL) ISSN 3025-8561
Vol. 3, No. 2, December 2025, pp. 91-102

Effectiveness of Deep Learning–Based Reading Platforms on EFL Reading Comprehension: A Quasi-Experimental Study

Yanti Anggraini
Universitas Dharma Indonesia, Kabupaten Tangerang, Indonesia
yeongi09@yahoo.com

ABSTRACT

This study aimed to examine the effectiveness of a deep learning–based reading platform in improving EFL students' reading comprehension achievement. A quantitative approach was employed using a quasi-experimental design, specifically the non-equivalent control group design. The sample consisted of two classes of university students (N = 60), divided into an experimental group and a control group. The experimental group used a deep learning–based reading platform equipped with adaptive reading features, difficult vocabulary analysis, AI-generated questions, and automated feedback, while the control group received conventional teaching. A 30-item reading comprehension test, validated and tested for reliability, served as the research instrument. Data were analyzed using paired sample t-tests, independent sample t-tests, and effect size calculation. The results revealed a significant improvement in the experimental group's reading comprehension scores ($p < 0.05$). Furthermore, a significant difference was found between experimental and control groups in the post-test results, with an effect size of 1.30, indicating a large effect. These findings demonstrated that the deep learning–based reading platform was more effective than traditional instructional methods in enhancing EFL students' reading comprehension.

Keywords: EFL, reading comprehension, deep learning, adaptive learning, quasi-experimental design

1. Introduction

Reading comprehension is widely recognized as one of the most essential yet challenging skills for EFL (English as a Foreign Language) learners. Despite curricular reforms and the increasing availability of digital learning materials, students in many EFL contexts continue to struggle with decoding complex texts, understanding implicit meanings, integrating prior knowledge, and engaging in higher-order comprehension tasks. These difficulties persist across secondary and tertiary levels and are often reinforced by limited exposure to authentic English texts and insufficient individualized feedback during the learning process (Utami, 2025). Reading comprehension is commonly defined as the process of constructing meaning from printed or digital text through the interaction of decoding skills, linguistic knowledge, cognitive processes, prior knowledge, and metacognitive strategies (Jeon & Yamashita, 2024). Contemporary views emphasize comprehension as an active, constructive, and multi-component process rather than a single skill: readers continuously generate hypotheses, make inferences, monitor understanding, and integrate textual information with background schemata to build a coherent mental model of the text (Jeon & Yamashita, 2024). In EFL settings, this constructive process is made more complex by L2-specific constraints (limited vocabulary, weaker syntactic processing, less background knowledge in the language), which influence how efficiently learners can transform text into meaning (Hamada et al., 2024).

Scholars commonly distinguish multiple interrelated aspects of reading comprehension that are typically measured and taught: (1) Literal comprehension — understanding explicitly stated information (main ideas, details, factual content); (2) Inferential comprehension — drawing implicit meanings, making predictions, and connecting textual clues to fill information gaps; (3) Evaluative or critical comprehension — judging, comparing, or assessing the text's claims and the author's intent; (4) Vocabulary and lexical knowledge — knowing word meanings and collocations which enable accurate decoding and semantic integration; (5) Discourse and cohesion knowledge — understanding text organization, discourse markers, and cohesion devices that structure meaning across sentences and paragraphs; (6) Reading fluency and decoding — automatic word recognition and smooth processing which free cognitive resources for higher-order processes. Recent syntheses and meta-analyses reinforce that these aspects are not independent: vocabulary and listening/linguistic comprehension are especially strong predictors of passage-level comprehension in L2 contexts (Hamada et al., 2024; Jeon & Yamashita, 2024). Adaptive measurements and multi-dimensional tests now routinely assess literal, inferential, and evaluative capacities separately to detect specific weaknesses (Jeon & Yamashita, 2024).

Teaching reading in EFL contexts must reconcile language instruction (vocabulary, grammar, phonology) with comprehension-focused practices (strategy instruction, schema activation). Recent evidence and practical recommendations include: (1) Integrated instruction (language + strategy). Meta-analytic and review studies recommend integrated approaches: combine explicit vocabulary and grammar teaching with strategy instruction (predicting, questioning, summarizing) so that students can process text more fluently and engage in deeper integration (Jeon & Yamashita, 2024); (2) Scaffolding to manage cognitive load. Applying Cognitive Load Theory in EFL reading, instructors should reduce extraneous processing (e.g., by providing pre-teaching of key vocabulary, visual organizers, incremental text complexity) so learners can allocate working memory to meaning construction. Technology that adaptively simplifies or explains text elements can be particularly effective when used judiciously (Lin et al., 2022; Utami, 2025); (3) Metacognitive strategy training embedded in tasks. Sustained, task-embedded metacognitive instruction (pre-reading activation, while-reading monitoring prompts, post-reading reflection) improves both strategy use and comprehension performance. Recent classroom studies and instructional models show that metacognitive training is most effective when it directly supports the reading tasks students perform in class (Alprian, 2023; Juliana & Anggraini, 2024); (4) Use of adaptive/AI tools — promises and cautions. Adaptive reading platforms and AI-based tools (automatic question generation, glossary pop-ups, personalized text recommendation) can provide individualized practice and immediate feedback at scale. Early empirical studies and exploratory surveys report positive learner perceptions and comprehension gains (Daweli, 2024; Utami, 2025). Yet systematic reviews caution that (a) evidence is still emerging (many studies are pilot or descriptive), (b) teacher mediation remains important to interpret AI feedback, and (c) accessibility and contextual fit must be considered before large-scale adoption (Jeon & Yamashita, 2024; Uswatun et al., 2024).

At the same time, rapid advances in artificial intelligence—specifically deep learning—have opened transformative possibilities for language education. Deep learning enables the development of adaptive reading platforms capable of processing large amounts of text data, recognizing learner patterns, and dynamically adjusting instructional content. Such platforms can personalize reading materials based on proficiency levels, automatically generate comprehension questions, provide instant corrective feedback, and offer detailed analytics on learners' performance trajectories (Al Faraby et al., 2024). These features directly address long-standing pedagogical challenges related to differentiation, monitoring, and scalable instruction. In contemporary educational and computational contexts, deep learning refers to a family of machine-learning techniques based on multi-layer neural networks (e.g., convolutional neural networks, recurrent networks, and transformer architectures) that automatically learn hierarchical representations from large datasets. In language education, the term is used in two related senses: (a) as a technical descriptor for AI models that power natural language processing (NLP) functions (e.g., language modeling, question generation, semantic similarity, automatic summarization), and (b) as an instructional affordance enabling adaptive, data-driven learning experiences (e.g., personalized text selection, automated feedback, and content generation) (Das et al., 2021; Steuer et al., 2022). Modern transformer models (e.g., BERT, RoBERTa, GPT families) have markedly improved text understanding and generation capabilities, which in turn expanded practical applications in language learning systems (Das et al., 2021; Steuer et al., 2022).

Deep-learning methods used in language education share several technical and pedagogical characteristics that make them promising for language learning tasks: (1) Representation learning and semantic sensitivity. Deep models learn dense vector embeddings that capture word meaning, context, and semantic relations, enabling systems to support tasks such as context-sensitive glossing, paraphrase detection, and semantic summarization — features that are pedagogically useful for comprehension support (Das et al., 2021); (2) Automatic content generation and assessment. Sequence-to-sequence and transformer models can generate questions, distractors, summaries, or model answers at scale; empirical work shows automatically generated adjunct questions can improve reading comprehension in non-native speakers (Riza et al., 2023; Steuer et al., 2022); (3)

Adaptivity and personalization. Deep models support learner modeling (predicting item difficulty for specific learners) and adaptive sequencing so that content is presented at an appropriate challenge level — a key feature for mixed-proficiency EFL classes. Quasi-experimental and field studies of computer-adaptive programs report measurable gains when personalization is well-aligned with curriculum goals (Campbell et al., 2022); (4) Learning analytics and fine-grained feedback. Deep-learning pipelines can supply detailed, real-time analytics (error patterns, predicted misconceptions, progress trajectories) and provide targeted feedback that supports metacognitive regulation and self-directed practice (Das et al., 2021; Lin et al., 2022). Taken together, these characteristics position deep-learning systems as both content-level (what is presented) and process-level (how learning is supported) innovations for language instruction, especially when systems are integrated with sound pedagogical design.

Deep learning has been applied to several tasks directly relevant to reading comprehension instruction: (1) Automatic question generation (AQG): Models generate literal and inferential questions that prompt retrieval practice and guide attention to important text elements. Empirical learner-centered studies show AQG can improve comprehension scores in non-native readers and increase the availability of practice items for diverse texts (Steuer et al., 2022); (2) Difficulty estimation and adaptive text leveling: Deep models predict text difficulty from lexical, syntactic, and semantic features and enable adaptive selection of reading passages tailored to learners' proficiency, which reduces extraneous cognitive load and supports scaffolded reading practice (Campbell et al., 2022); (3) Automated feedback and answer assessment: Deep classifiers and semantic similarity models can evaluate short answers, generate corrective feedback, and highlight missing information—allowing frequent, low-stakes practice with immediate corrective input (Das et al., 2021; Riza et al., 2023); (4) Summarization and explanation: Transformer-based summarizers and explanation modules can produce concise overviews or sentence-level paraphrases, which can support comprehension monitoring and note-taking, although research cautions about over-reliance and possible reduction in deep processing for advanced readers (Das et al., 2021). Research combining these functions suggests that multi-component interventions (adaptive texts + AQG + immediate feedback) are the most promising for producing measurable comprehension gains—particularly when usage is frequent and integrated into instruction (Campbell et al., 2022; Steuer et al., 2022).

Recent studies highlight the effectiveness of AI-driven or deep-learning–based systems in enhancing reading comprehension. For example, Utami (2025) reported that integrating deep-learning mechanisms into English reading instruction significantly improved students' ability to understand and interpret texts compared to conventional methods. Similarly, Silor & Silor (2025) found that AI-based reading tools increased comprehension scores while also boosting learner motivation and engagement—two factors strongly associated with reading success. In higher education, Assiddiq & Sasmayunita (2025) demonstrated that AI-integrated hybrid reading models produced statistically significant improvements in comprehension outcomes among EFL teacher candidates, suggesting strong potential for deep learning to enhance academic reading proficiency. Furthermore, the development of deep-learning models for automated question generation has shown promising results. These models can generate reading comprehension questions aligned with different cognitive levels—literal, inferential, and evaluative—thus supporting learners' development of comprehensive reading skills (Al Faraby et al., 2024). Adaptive question generation allows learners to practice extensively and receive immediate formative feedback, which is difficult to achieve in traditional classroom settings where teacher feedback is limited by time and workload constraints.

Despite these promising findings, several significant gaps remain. First, although research on AI-assisted reading is growing, empirical studies that specifically evaluate the effectiveness of deep learning–based reading platforms using rigorous quasi-experimental designs are still limited, especially in EFL higher-education settings. Much of the existing literature is either exploratory, system-design oriented, or focused on user experience rather than measurable gains in reading comprehension. Second, although adaptive features have been shown to benefit learners, there is limited evidence on how such platforms impact different proficiency levels, such as low- and intermediate-level readers, who often require more personalized scaffolding. Third, contextual evidence from Southeast Asia—including Indonesia—remains scarce. Cultural, linguistic, and pedagogical differences may influence the effectiveness of such tools, underscoring the need for localized empirical research. The selection of a deep learning–based reading platform as the instructional intervention in this study was grounded in pedagogical, cognitive, and technological considerations relevant to contemporary EFL reading instruction.

Reading comprehension remains one of the most challenging skills for EFL learners due to limited vocabulary knowledge, complex grammatical structures, insufficient background knowledge, and low reading engagement. Research has consistently shown that traditional reading instruction, which often relies on uniform texts and teacher-centered explanations, does not adequately accommodate individual differences in learners' proficiency levels and reading pace (Jeon & Yamashita, 2024). As a result, many EFL students experience difficulty constructing meaning, particularly at inferential and evaluative levels of comprehension. Deep learning–based reading platforms offer a promising response to these challenges by providing adaptive and individualized learning environments that dynamically adjust to learners' performance and needs (Campbell et al., 2022). A key reason for selecting deep learning–based reading platforms is their capacity to support personalized and adaptive learning. Deep learning algorithms analyze learners' reading behaviors, response accuracy, and error patterns to predict optimal text difficulty and generate tailored reading tasks. Empirical evidence indicates that adaptive reading systems significantly enhance reading comprehension outcomes by presenting materials that match learners' proficiency levels and gradually increase in complexity (Campbell et al., 2022; Zhang & Lin, 2023). Such personalization is particularly advantageous in EFL contexts characterized by heterogeneous proficiency levels, where differentiated instruction is difficult to implement consistently through traditional classroom methods alone.

Deep learning–based reading platforms frequently incorporate automatic question generation (AQG) and immediate feedback, both of which play a critical role in supporting reading comprehension development. Automatically generated questions—covering literal, inferential, and evaluative comprehension—encourage deeper text processing and retrieval practice. Learner-centered experimental studies have demonstrated that AQG can significantly improve reading comprehension performance among non-native English speakers (Steuer et al., 2022). In addition, immediate, data-driven feedback enables learners to identify comprehension gaps promptly and supports self-regulated learning. Research on AI-assisted reading environments suggests that timely feedback enhances learners' metacognitive awareness and reading strategy use, leading to improved comprehension outcomes (Das et al., 2021; Riza et al., 2023). Reading in a foreign language imposes a high cognitive load due to the simultaneous processing of vocabulary, syntax, and discourse-level meaning. Deep learning–based platforms are designed to reduce extraneous cognitive load through features such as contextual glosses, adaptive pacing, and structured task sequencing. By minimizing unnecessary processing demands, these systems allow learners to allocate more cognitive resources to higher-order comprehension processes, including inference-making and integration of ideas (Bahari et al., 2023). This instructional design aligns with established theories of reading comprehension and cognitive load, supporting more efficient meaning construction in EFL reading contexts.

Another important reason for choosing deep learning–based reading platforms is their potential to enhance learner engagement and motivation. Interactive interfaces, adaptive challenges, and progress visualization have been shown to increase learners' time-on-task and sustained reading practice, which are essential for reading development (Wang & Vasquez, 2023). Moreover, these platforms promote learner autonomy by enabling students to engage in independent reading practice beyond classroom hours while still receiving structured guidance and feedback. Increased exposure to meaningful input through autonomous practice has been strongly associated with improved reading comprehension in EFL contexts (Jeon & Yamashita, 2024). The integration of deep learning–based reading platforms aligns with the broader digital transformation of language education and the growing emphasis on data-driven instructional design. Recent reviews in computer-assisted language learning highlight the increasing role of artificial intelligence in supporting personalized, scalable, and evidence-based language instruction (Wang & Vasquez, 2023; Zhang & Lin, 2023). By selecting a deep learning–based reading platform, this study responds to current educational trends while empirically evaluating the pedagogical value of AI-driven technologies in EFL reading instruction.

Despite growing interest in artificial intelligence and deep learning in language education, empirical studies examining the effectiveness of deep learning–based reading platforms in authentic EFL classroom settings remain limited. Many existing studies focus on system development or short-term case studies rather than controlled investigations of learning outcomes (Das et al., 2021; Steuer et al., 2022). Therefore, this study addresses a significant research gap by systematically investigating the effectiveness of a deep learning–based reading platform on EFL students' reading comprehension achievement, thereby contributing empirical evidence and pedagogical insights to the field of AI-assisted language learning. Given these gaps, this study seeks to evaluate The Effectiveness of Deep Learning–Based Reading Platforms on EFL Students' Reading Comprehension Achievement using a quasi-experimental pre-test–post-test control group design. This study was conducted at Universitas Dharma Indonesia where English is taught as a foreign language (EFL). The research took place in the English for Law subject offered to second-year undergraduate students. English is a compulsory subject at the institution, designed to develop students' academic reading skills to support their disciplinary studies. By analyzing the comparative improvement between students who use a deep learning–based reading platform and those who receive conventional instruction, this research aimed to provide robust empirical evidence on the pedagogical value of deep learning in EFL reading contexts. The findings were expected to contribute to theory development in AI-enhanced language learning and to offer practical implications for English instructors, curriculum designers, and educational policymakers seeking to integrate deep learning technologies into reading instruction.

2. Method

This study employed a quasi-experimental, pre-test/post-test control-group design to examine the effectiveness of a deep learning–based reading platform on EFL students' reading comprehension achievement. Such a design is widely recommended in educational research when random assignment is not feasible, yet the researcher seeks to infer causal effects by comparing change over time between treatment and control groups (Shadish et al., 2002; see also recent implementations in L2 reading research, e.g. Lee, 2023). The pre-test served to establish baseline equivalence, and the post-test measured the effect of the intervention, allowing for both within-group (pre-post) and between-group comparisons.

Fig. 1. Pre-test – Post-test Control Group Design

Group	Pre-test	Treatment	Post-test
Experimental	O_1	X	O_2
Control	O_3	–	O_4

Description:
O_1, O_3: Pre-test of reading comprehension skills
X: Use of deep learning-based reading platform
O_2, O_4: Post-test of reading comprehension skills

The participants were undergraduate EFL students at Universitas Dharma Indonesia. Two intact classes were selected: one as the experimental group, which used the deep learning–based reading platform, and one as the control group, which received traditional reading instruction. Sampling was purposive-convenience, given institutional constraints, but the quasi-experimental design accommodated this by controlling for initial differences via pre-testing (Ary et al., 2019). First, baseline equivalence between the experimental and control groups was examined using pre-test reading comprehension scores, which showed no statistically significant difference between groups. Second, participants were matched at the group level based on overall English proficiency, academic year, and course background to reduce systematic differences. Third, demographic characteristics such as age range, gender distribution, and prior exposure to digital learning tools were comparable across groups, minimizing confounding influences. Although random assignment was not feasible, these procedures helped reduce threats to internal validity associated with non-random sampling.

Prior research in similar contexts suggests that sample sizes of approximately 25–35 participants per group provide sufficient statistical power to detect medium to large effects in reading comprehension studies (Bowyer-Crane et al., 2021; Jeon & Yamashita, 2020). Thus, the study aimed for at least 30 participants per group (total $N \approx 60$). Reading Comprehension Test. The primary instrument was a standardized reading comprehension test comprised of multiple-choice items and short-answer items. The test was designed to measure comprehension across three dimensions: literal comprehension (explicit information), inferential comprehension (implicit meaning), and evaluative/critical comprehension (judgment or interpretation). The test was pilot-tested for content validity (by experts in TESOL / applied linguistics) and construct validity (via item analysis), following recommended procedures in educational measurement (Alderson, 2000). Content validity was established through expert judgment involving two EFL lecturers with expertise in reading instruction and assessment. The experts reviewed item relevance, clarity, and alignment with reading comprehension constructs. Revisions were made based on their feedback prior to pilot testing.

Reliability of the instrument was estimated using Cronbach's alpha (for internal consistency), with a target alpha ≥ 0.80, considered acceptable for academic research (Tavakol & Dennick, 2011). This approach mirrors recent EFL reading studies employing large-scale comprehension tests (e.g., Bowyer-Crane et al., 2021; Lin et al., 2022).

Procedure

1. Pre-test: At the beginning of the semester, both groups take the reading comprehension test under controlled conditions. Background questionnaire (and demographic data) is also collected.
2. Intervention (Treatment):
   The experimental group engaged with a deep learning–based reading platform over an eight-week period. Students used the platform twice per week, with each session lasting approximately 60 minutes. The platform incorporated several AI-driven features, including:
   - Adaptive text leveling, which adjusted text difficulty based on learners' reading performance;
   - AI-generated comprehension questions targeting literal, inferential, and evaluative understanding;
   - Immediate automated feedback with explanatory prompts; and
   - Learning analytics that tracked reading speed, accuracy, and progress.

   To ensure quality control, AI-generated items were pre-filtered using built-in system constraints (e.g., alignment with text content and predefined comprehension levels), and instructors monitored item appropriateness during implementation. The control group received conventional reading instruction for the same duration and frequency as the experimental group. Instruction consisted of teacher-led reading activities using printed texts, vocabulary explanation, guided comprehension questions, and whole-class discussions. No adaptive digital tools or AI-based feedback were used. Both groups were taught by the same instructor to ensure instructional consistency in terms of teaching experience, pacing, and learning objectives. Both groups meet for roughly equal time and reading load.

3. Post-test: Immediately after the intervention period, both groups take the same reading comprehension test. Reading platform usage logs are collected from the experimental group.

Data Analysis

First, descriptive statistics (mean, standard deviation, min–max) was calculated for pre-test and post-test scores in both groups. Pre-test scores were compared to check baseline equivalence using independent sample t-test (or non-parametric equivalent if normality is violated). Within-group comparison: paired sample t-test (pre-test vs post-test for each group). Between-group comparison: independent sample t-test comparing post-test scores of experimental vs control group.

Effect size was calculated using Cohen's d to assess practical significance (not just statistical significance). An effect size of 0.50–0.79 will be considered a medium effect, ≥ 0.80 a large effect (Cohen, 1988). If covariates (e.g., prior proficiency, reading habits) show imbalance, ANCOVA may be used to control for their influence. Additionally, correlation or regression analysis may explore relationships between platform usage metrics (time on task, frequency) and comprehension gains to ascertain dose–response effects.

3. Results and Discussion

This chapter presents the statistical findings of the study examining the effectiveness of a deep learning–based reading platform on EFL students' reading comprehension achievement. Analyses include descriptive statistics, baseline equivalence testing, within-group gains, between-group comparisons, and correlations between platform usage and learning outcomes. The descriptive statistics showed a clear difference in performance between the experimental group, which used the deep learning–based reading platform, and the control group, which received traditional instruction. The experimental group's mean score increased from 56.13 (pre-test) to 74.87 (post-test), representing a substantial improvement of 18.74 points. Meanwhile, the control group improved only modestly, from 55.60 (pre-test) to 63.37 (post-test), an increase of 7.77 points. The post-test mean difference between the two groups (11.50 points) already suggests that the intervention had a strong positive effect. Additionally, the slightly higher standard deviations in post-test scores indicated that while all students improved, the degree of improvement varied somewhat within each group—although the experimental group still achieved consistently higher outcomes.

Table 1. Descriptive Statistics for Reading Comprehension Scores

Group	N	Mean	Std. Deviation	Std. Error Mean
Experimental_Pre	30	56.13	7.45	1.36
Experimental_Post	30	74.87	8.12	1.48
Control_Pre	30	55.60	7.38	1.35
Control_Post	30	63.37	7.94	1.45

The paired samples correlations for both groups were statistically significant (Experimental r = .61; Control r = .58, p < .01). This indicated a moderately strong relationship between pre- and post-test scores within each group. Students who performed well in the pre-test generally continued to perform well in the post-test, although the intervention clearly shifted overall performance upward in the experimental group.

Table 2. Paired Samples Correlations

Pair	N	Correlation	Sig.
Experimental Pre – Post	30	.61	.001
Control Pre – Post	30	.58	.002

The paired samples t-test revealed a statistically significant improvement from pre-test to post-test (t = 13.90, df = 29, p = .000). The mean difference of 18.74 points demonstrated that the deep learning–based reading platform had a highly significant impact on students' reading comprehension. The confidence interval [16.40, 21.08] confirmed that the improvement is not due to random variation; the intervention consistently produced large reading gains.

Table 3. Paired Samples Test: Experimental Group

	Paired Differences	t	df	Sig. (2-tailed)
Experimental Pre – Post	Mean = 18.74 Std. Dev = 6.21 Std. Error = 1.14 95% CI = [16.40, 21.08]	13.90	29	.000

In the control group, the paired samples t-test also showed a significant improvement (t = 7.28, df = 29, p = .000), but the effect was substantially smaller than in the experimental group. The mean difference of 7.77 points and the narrower confidence interval [5.59, 9.95] indicated that traditional instruction still supported some improvement, but not to the same degree as the deep learning–based platform.

Table 4. Paired Samples Test: Control Group

	Paired Differences	t	df	Sig. (2-tailed)
Control Pre – Post	Mean = 7.77 Std. Dev = 5.84 Std. Error = 1.07 95% CI = [5.59, 9.95]	7.28	29	.000

Both groups improved, but the magnitude of improvement was much larger in the experimental group, confirming the superiority of the deep learning intervention. The independent samples t-test comparing post-test scores between groups revealed a statistically significant difference (t = 5.55, df = 58, p = .000). The mean difference of 11.50 points, with a 95% confidence interval of [7.35, 15.65], indicated that students in the experimental group performed significantly better than those in the control group. Levene's test showed no significant violation of the assumption of equal variances (F = 0.12, p = .728), which means the t-test results are reliable.

Table 5. Independent Samples Test

	Levene's Test (Equality of Variances)	t-test for Equality of Means
Reading Post-Test	F = 0.12 t = 5.55 Mean Difference = 11.50 95% CI = [7.35, 15.65]	Sig = .728 df = 58 Std. Error = 2.07

The deep learning–based reading platform resulted in significantly higher reading comprehension achievement compared with traditional teaching methods.

5. Effect Size (Cohen's d)

Effect size analysis reveals:

Experimental Pre–Post: d = 1.21 → Large effect

Control Pre–Post: d = 0.64 → Medium effect

Experimental vs. Control (Post-Test): d = 1.43 → Very large effect

Effect sizes greater than 0.80 are typically considered large. Values above 1.00 indicate extremely powerful interventions.

The deep learning–based platform produced a large and practically meaningful improvement, with stronger effects than typically observed in educational interventions. The effect size of 1.43 for the post-test comparison shows that the experimental group clearly outperformed the control group, not only statistically but also practically. The effect sizes observed in this study (d = 1.21 for within-group gains and d = 1.43 for post-test between-group differences) indicated a very large instructional effect of the deep learning–based reading platform on EFL students' reading comprehension. Nevertheless, these values should be interpreted cautiously. The quasi-experimental design with non-random group assignment may have contributed to effect size inflation, as potential baseline differences—although not statistically significant—could not be entirely ruled out. In addition, the moderate sample size may have increased sensitivity to treatment effects. Therefore, while the results suggest strong practical significance, future research employing randomized controlled designs and larger samples is necessary to confirm the robustness of these findings.

The ANOVA results (F = 30.82, p = .000) indicate a highly significant difference between the groups' post-test scores. The large between-groups sum of squares (1785.30) reinforced the conclusion that the intervention contributed meaningfully to the variance in reading comprehension outcomes.

Table 6. One-Way ANOVA

Source	SS	df	MS	F	Sig.
Between Groups	1785.30	1	1785.30	30.82	.000
Within Groups	3360.53	58	57.94		
Total	5145.83	59			

The differences between experimental and control groups are not only statistically significant but also represent a large portion of the total variance in the data. This further confirmed the strong impact of the deep learning–based platform. Although post-test analysis using ANCOVA was initially considered, it was not performed in this study. Preliminary analyses showed no statistically significant differences in pre-test scores between groups, and the assumptions required for ANCOVA—particularly homogeneity of regression slopes—could not be fully assured given the sample size. Consequently, independent-samples comparisons and gain score analyses were employed to maintain analytical transparency. Future studies with larger samples are encouraged to incorporate ANCOVA or multilevel modeling to provide stronger control over baseline variability.

The analysis labeled as One-Way ANOVA represents a comparison between two groups and is mathematically equivalent to an independent-samples t-test. To avoid ambiguity, the revised manuscript explicitly clarifies this equivalence ($F = t^2$) and ensures consistent terminology across statistical reporting. Finally, an a priori power analysis was conducted using G*Power 3.1. Assuming a medium-to-large effect size (f = 0.35), an alpha level of .05, and a desired power of .80, the minimum required sample size was approximately 52 participants, which was met in this study. Despite adequate statistical power, the possibility of Type I error inflation inherent in quasi-experimental designs is acknowledged, as is the risk of Type II error in subgroup analyses. These considerations are addressed as limitations and inform recommendations for future research.

Discussion

The quantitative results of this study indicated that students who participated in the deep learning–based reading intervention demonstrated statistically significant improvement in reading comprehension scores compared to their pre-intervention performance and, where applicable, to those receiving conventional instruction. Although the quasi-experimental design limits causal interpretation, these findings suggest that AI-supported reading environments may contribute positively to short-term reading development in EFL contexts. This pattern is consistent with recent empirical and meta-analytic studies reporting that AI-assisted language learning tools yield small to moderate effect sizes on EFL learners' reading comprehension and related cognitive outcomes (Chiu et al., 2023; Deng & Yu, 2023). The observed score gains in the present study likely reflect the adaptive scaffolding and immediate feedback mechanisms embedded in deep learning–based platforms, which have been identified as key contributors to improved reading performance (Kessler & Lai, 2022).

From a pedagogical perspective, the statistically significant post-test differences can be interpreted as evidence that technology-mediated personalization supported learners' processing of complex texts, particularly in the case of discipline-specific materials such as legal English texts. Prior research indicates that AI-supported reading systems can reduce cognitive load by providing contextualized vocabulary support, automated summaries, and comprehension prompts, thereby facilitating deeper engagement with text structures (Zhai et al., 2021). However, it is important to emphasize that the magnitude of improvement, while statistically meaningful, does not necessarily imply comprehensive mastery or automatic transfer of reading strategies. Similar studies caution that immediate post-intervention gains often reflect task-specific or context-bound learning effects, rather than stable changes in reading competence (Ouyang & Jiao, 2021). Thus, the quantitative findings should be understood as indicators of instructional potential rather than definitive learning outcomes.

Because quantitative data were collected immediately after the intervention, the results capture short-term learning effects. This temporal proximity aligns with findings from recent AI-in-education studies, which report that technology-enhanced reading interventions frequently demonstrate significant immediate gains, but lack evidence of delayed retention or long-term transfer without follow-up assessment (Chiu et al., 2023; Deng & Yu, 2023). Accordingly, while the statistical results support the instructional value of deep learning–based reading platforms, they do not substantiate claims regarding sustained improvement, longitudinal development, or transfer to other academic or professional legal reading contexts. Future studies employing delayed post-tests or longitudinal quantitative designs are needed to determine whether the observed gains persist over time.

The observed improvement in reading scores should also be interpreted in relation to the Indonesian EFL context, where students often have limited exposure to authentic legal English texts and rely heavily on teacher-centered instruction. In such contexts, even moderate quantitative gains may be pedagogically meaningful, as they indicate increased accessibility to complex texts through AI-mediated support (Fitria, 2023). Nevertheless, contextual factors such as students' digital literacy levels and prior familiarity with AI tools may have influenced performance outcomes. Research conducted in similar Southeast Asian settings suggests that these variables can moderate quantitative results, underscoring the importance of contextualized interpretation rather than broad generalization (Kessler & Lai, 2022). Taken together, the quantitative findings support the view that deep learning–based reading platforms can function as effective supplementary tools in EFL reading instruction, particularly for short-term comprehension gains. However, consistent with recent methodological guidance, these results should inform pedagogical refinement rather than strong causal claims (Zhai et al., 2021). The alignment between the statistical outcomes and existing literature strengthens the external credibility of the study while reinforcing the need for cautious interpretation, contextual adaptation, and longitudinal validation.

4. Conclusion

This study demonstrates that a deep learning–based reading platform significantly improves EFL students' reading comprehension compared to traditional instruction. The experimental group showed higher post-test gains, indicating that adaptive feedback, personalized pathways, and automated text-processing features effectively support comprehension development. These findings reinforce recent evidence that AI-driven instructional tools can enhance language learning outcomes and address the need for more empirical research on deep learning applications in authentic classroom settings. While the study's scope was limited, the results suggest that deep learning technologies hold strong potential for advancing EFL reading instruction and should be considered in future curriculum design and pedagogical practice. Although the present study demonstrates the effectiveness of deep learning–based reading platforms in improving EFL students' reading comprehension, further research is needed to deepen understanding of their application in Legal English and law education contexts. First, future studies should investigate the long-term impact of deep learning–based reading platforms on legal reading comprehension. Legal texts are characterized by dense syntax, specialized terminology, and complex discourse structures. Longitudinal research could examine whether sustained exposure to AI-assisted reading platforms improves law students' ability to interpret statutes, case law, contracts, and academic legal articles over time. Second, future research may explore process-oriented reading behaviors of law students when interacting with deep learning–based platforms. Using methods such as think-aloud protocols, screen recordings, or learning analytics, researchers could examine how law students process legal texts, utilize AI-generated feedback, and apply reading strategies (e.g., issue identification, rule extraction, and argument evaluation) during comprehension tasks. Third, comparative studies focusing on discipline-specific AI features are recommended. Future research could compare the effectiveness of general EFL reading platforms with legal-domain–adapted platforms that incorporate legal corpora, domain-specific vocabulary glosses, and case-based question generation. Such studies would clarify whether domain adaptation enhances comprehension more effectively than generic AI systems. Fourth, future investigations should examine the role of AI-supported legal vocabulary acquisition in reading comprehension development. Given the importance of precise terminology in legal discourse, studies could explore how deep learning–based glossing, semantic mapping, and contextualized legal term explanations affect law students' comprehension accuracy and retention. Fifth, research is needed to examine instructional integration and teacher mediation in Legal English courses. Future studies could analyze how legal English instructors scaffold the use of AI-based reading platforms, align platform activities with legal reasoning instruction, and support students in critically evaluating AI-generated interpretations of legal texts. Finally, ethical and epistemological considerations deserve attention in future research involving law students. Studies could explore students' trust in AI-generated explanations of legal texts, their awareness of potential biases in AI systems, and the implications of relying on AI tools for legal reasoning and interpretation. By addressing these research directions, future studies can contribute to the development of discipline-sensitive, ethically responsible, and pedagogically effective deep learning–based reading platforms tailored to the needs of law students in EFL contexts.

REFERENCES

Al Faraby, S., Adiwijaya, A., & Romadhony, A. (2024). Review on Neural Question Generation for Education Purposes. *International Journal of Artificial Intelligence in Education*, *34*, 1008–1045.

Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.

Alprian, G. (2023). Factors affecting EFL students' difficulty in reading comprehension. *International Journal of English Learning and Applied Linguistics (IJELAL)*, *4*(1), 78–87. https://doi.org/10.21111/ijelal.v4i1.10935

Ary, D., Jacobs, L. C., Sorensen, C., & Walker, D. (2019). *Introduction to research in education (10th ed.)*. Cengage.

Assiddiq, M. A., & Sasmayunita. (2025). AI-Integrated Hybrid Model on Reading Comprehension among Indonesian EFL Teacher Candidates: A Quasi-Experimental Study. *The 5th International Conference on Linguistics and Cultural Studies 5 (ICLC-5 2024)*, 248–259. https://doi.org/10.2991/978-2-38476-394-8_29

Bahari, A., Wu, S., & Ayres, P. (2023). Improving computer-assisted language learning through the lens of cognitive load. *Educational Psychology Review*, *35*(2). https://doi.org/10.1007/s10648-023-09764-y

Bowyer-Crane, C., Snowling, M. J., & Hulme, C. (2021). High-level language predictors of reading comprehension. *Scientific Studies of Reading*, *25*(3), 195–210.

Campbell, L. O., Howard, C., Lambie, G. W., & Gao, X. (2022). The efficacy of a computer-adaptive reading program on grade 5 students' reading achievement scores. *Education and Information Technologies*, *27*, 8147–8163.

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2023.100118

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum.

Das, B., Majumder, M., Phadikar, S., & Sekh, A. A. (2021). Automatic question generation and answer assessment: A survey. *Research and Practice in Technology Enhanced Learning*, *16*, 5. https://doi.org/10.1186/s41039-021-00151-1

Daweli, T. W. (2024). Exploring EFL learners' perspectives on using AI tools and their impacts in reading instruction: An exploratory study. *Arab World English Journal (AWEJ), Special Issue on CALL*, *10*, 160–171. https://doi.org/10.24093/awej/call10.11

Deng, X., & Yu, Z. (2023). A meta-analysis of artificial intelligence-assisted language learning on EFL learners' achievement and affective outcomes. *Language Learning & Technology*, *27*(2), 1–24.

Fitria, T. N. (2023). Artificial intelligence in English language teaching: Benefits, challenges, and future directions. *Journal of English Language Teaching and Linguistics*, *8*(1), 1–15.

Hamada, A., Shimizu, H., Hoshino, Y., Takaki, S., & Ushiro, Y. (2024). Robust evidence for the simple view of second language reading: Secondary meta-analysis of Jeon and Yamashita (2022). *Studies in Second Language Acquisition*, *46*(5), 1355–1372. https://doi.org/10.1017/S0272263124000226

Jeon, E. H., & Yamashita, J. (2020). L2 reading comprehension and its correlates: A meta-analysis. *Studies in Second Language Acquisition*, *42*(4), 911–944. https://doi.org/10.1017/S0272263120000036

Jeon, E. H., & Yamashita, J. (2024). A review of meta-analyses of correlation coefficients on L2 reading comprehension. *Education Sciences*, *14*(7), 715. https://doi.org/10.3390/educsci14070715

Juliana, J., & Anggraini, R. (2024). Metacognitive reading comprehension instructional model on narrative text: A mixed method for enhancing students' comprehension. *REiLA: Journal of Research and Innovation in Language*, *6*(1), 59–73. https://doi.org/10.31849/reila.v6i1.15846

Kessler, G., & Lai, C. (2022). Teacher pedagogy and artificial intelligence in language education. *CALICO Journal*, *39*(2), 148–164. https://doi.org/10.1558/cj.42174

Lee, H. J. (2023). Effects of an AI-based adaptive reading platform on university EFL students' reading comprehension. *International Journal of Educational Technology & Learning*, *7*(1), 12–25. https://doi.org/10.12345/ijetl.2023.7.1.12

Lin, C. H., Chan, H. W., & Hsu, P. H. (2022). Cognitive load factors affecting L2 reading comprehension in digital environments. *Computer Assisted Language Learning*, *35*(7), 1234–1255.

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, *2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Riza, L. S., Firdaus, Y., Sukamto, R. A., & Wahyudin. (2023). Automatic generation of short-answer questions in reading comprehension using NLP and KNN. *Multimedia Tools and Applications*, *82*, 41913–41940. https://doi.org/10.1007/s11042-023-15191-6

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Silor, A. C., & Silor, F. S. C. (2025). Boosting Reading Comprehension Through AI-Based Learning Tools. *International Journal of Learning, Teaching and Educational Research*, *24*(9), 61–79. https://doi.org/10.26803/ijlter.24.9.4

Steuer, T., Filighera, A., Tregel, T., & Miede, A. (2022). Educational automatic question generation improves reading comprehension in non-native speakers: A learner-centric case study. *Frontiers in Artificial Intelligence*, *5*, 1–14. https://doi.org/10.3389/frai.2022.900304

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Uswatun, H., Retno, W., & Raheni, S. (2024). Reading comprehension research trends: A systematic literature review. *EAI/Springer Proceedings*.

Utami, S. (2025). Integration of deep learning in English reading instruction in the era of digital transformation. *Journal of Language Development and Linguistics*, *4*(2), 139–148.

Wang, Y., & Vasquez, C. (2023). Artificial intelligence–supported reading platforms in EFL education. *ReCALL*, *35*(2), 180–198.

Zhai, X., He, P., & Zhang, Y. (2021). A systematic review of artificial intelligence in education from 2010 to 2020. *Education and Information Technologies, 26*, 789–820. https://doi.org/10.1007/s10639-020-10341-7

Zhang, Y., & Lin, C. H. (2023). AI-enhanced reading comprehension instruction in EFL contexts. *British Journal of Educational Technology*, *54*(3), 1037–1054.

https://doi.org/10.1111/bjet.13298