ISELT 126
Interdisciplinary Studies in English Language Teaching

Pages 126 to 142

Integrating ChatGPT into Extensive Reading: Evidence from an Experimental Study on EFL Reading Fluency, Vocabulary Gains, and Motivation

Faraidoon Namiq [1]*
[1] University of Garmian, Kalar, Iraq

10.22080/iselt.2026.30417.1145

Received
November 1, 2025
Accepted
April 27, 2026
Available online
May 11, 2026

Keywords:
ChatGPT Integration, Extensive Reading, Reading Fluency, Vocabulary Gains, Motivation.

Abstract

Extensive reading (ER) has long been recognized as an effective approach in EFL instruction; however, the potential contribution of integrating generative artificial intelligence into established ER programs remains underexplored. The present study examined whether incorporating ChatGPT into extensive reading enhances EFL learners' reading fluency, vocabulary development, and reading motivation. Participants were 60 Kurdish intermediate EFL learners, aged 18 to 23, in a private language institute in Erbil, Iraq. Following placement via the Oxford Placement Test, intact classes were assigned to one of three instructional conditions over a ten-week period: a control group, an ER-only group, and an ER group supplemented with guided ChatGPT activities. The control group followed the institute's regular curriculum without extensive reading or AI support; the ER-only group engaged in sustained, self-selected extensive reading using level-appropriate texts; and the ER+ChatGPT group completed the same extensive reading activities supplemented with brief, guided ChatGPT interactions designed to scaffold comprehension and vocabulary without modifying the texts. Reading fluency was measured using words-correct-per-minute from parallel passages, vocabulary gains were assessed through a teacher-made 30-item test with parallel forms, and reading motivation was measured using a validated questionnaire (Wang & Jin, 2021). Pretest-adjusted ANCOVA results indicated that ER significantly outperformed the control condition on all outcome measures, with medium to large effect sizes, indicating educationally meaningful gains in reading fluency, vocabulary development, and reading motivation. More importantly, learners in the ER+ChatGPT condition demonstrated significantly greater gains in reading fluency, vocabulary, and motivation than those in the ER-only group. These findings suggest that integrating carefully guided AI-based scaffolding into extensive reading can yield incremental benefits without compromising the core principles of ER.

1. INTRODUCTION

In an era where generative artificial intelligence is rapidly entering language classrooms, a central pedagogical question concerns how such tools can be integrated without disrupting well-established and theoretically grounded instructional practices. Extensive Reading (ER) continues to be one of the most widely endorsed approaches for accelerating second language (L2) reading

* **Corresponding Author:** Faraidoon Namiq, MSc, Language Centre, University of Garmian, Kalar, Iraq, **Email**: faraidoonamen@gmail.com

development, largely because it is theoretically grounded in input-based models of language acquisition that emphasize sustained exposure to comprehensible input, learner autonomy, enjoyment, and reading speed (Liu & Saad, 2025). Classic ER principles, such as the use of easy texts, learner choice, and reading for general meaning, have shaped instructional practice for over two decades and are theoretically associated with reduced cognitive load and increased processing efficiency. As a result, ER is expected to foster improvements in reading fluency, comprehension, and vocabulary when compared with business-as-usual instruction (Jeon & Day, 2016).

Despite this strong theoretical and empirical foundation, little is currently known about how emerging generative AI tools, such as ChatGPT, may systematically interact with extensive reading frameworks to influence core reading outcomes. In particular, the incremental contribution of ChatGPT when layered onto a well-established ER program has not yet been rigorously examined.

From a theoretical perspective, ER is assumed to promote reading fluency through the gradual development of automaticity, a process whereby repeated exposure to high-frequency lexical and syntactic patterns reduces the attentional demands of decoding (Hunter, 2017). The same pattern has been observed in longitudinal studies that tracked significant words-per-minute reading gains within relatively short instructional periods, such as a single semester or academic year. For example, in a study of ER that engaged learners in "pleasure reading," Beglar and Hunt (2014) identified statistically significant gains in reading rates. In a different study, ER learners outperformed intensive reading learners in terms of reading rate development across levels of text difficulty (Beglar & Hunt, 2014). Moreover, ER provides rich opportunities for implicit vocabulary acquisition, and case studies and laboratory research have demonstrated measurable levels of vocabulary uptake from reading alone, although the magnitude of vocabulary gains may vary depending on factors such as frequency of exposure, lexical salience, and assessment conditions (Beglar & Hunt, 2014).

However, ER practice in instructional contexts faces several persistent challenges, including maintaining learner motivation, providing timely support for comprehension, and assisting learners in consolidating lexical knowledge without transforming ER into test-oriented instruction or cognitively demanding seatwork. Recent theoretical and conceptual work on Large Language Models (LLMs) suggests that such tools may help address these challenges by offering dialogic scaffolding, including comprehension prompts, summaries, and targeted vocabulary support, delivered at moments of need (Chen et al., 2024). Emerging discussions in EFL pedagogy indicate that structured ChatGPT use has the potential to enhance learner engagement, motivation, and vocabulary development, provided that its use is guided and mediated by instructors (Karataş et al., 2024).

Early classroom-oriented accounts of AI use in reading contexts suggest that ChatGPT can be integrated into pre-, during-, and post-reading routines to support comprehension monitoring and lexical reinforcement (Bushnell, 2025). Theoretically, this ER-plus-AI model aligns with sociocultural views of mediated learning and motivational theories emphasizing perceived competence and autonomy, suggesting that limited, guided AI interaction may enhance engagement without transforming ER into test-oriented instruction. However, from a theoretical standpoint, it remains unclear whether such AI-mediated scaffolding produces incremental benefits beyond ER alone across multiple reading-related outcomes, including fluency, vocabulary development, and reading motivation, within a single instructional framework (El Hassan & Alsalwah, 2025).

To address this gap, the present study investigates the effects of integrating ChatGPT into extensive reading on EFL learners' reading fluency, vocabulary gains, and reading motivation. By adopting a controlled, intact-class experimental design and employing validated outcome

measures, this research aims to provide empirical evidence regarding the added value of AI-mediated scaffolding within an ER framework. In doing so, the study contributes to current discussions on how generative AI can complement established pedagogical approaches while preserving the core principles of extensive reading. The study is guided by the following research questions:

RQ_1: Does integrating ChatGPT into extensive reading significantly impact EFL learners' reading fluency?

RQ_2: Does integrating ChatGPT into extensive reading significantly impact EFL learners' vocabulary gains?

RQ_3: Does integrating ChatGPT into extensive reading significantly impact EFL learners' reading motivation?

2. LITERATURE REVIEW

Theoretical Framework

This study is framed by three complementary theoretical frameworks to explain how both ER and AI-mediated support can affect the development of reading skills. The first framework is that of the theory of ER (Day et al., 1998, 2002; Krashen, 2004), suggesting that if learners read large amounts of meaning-focused comprehensible input over an extended period of time, they will increase their reading ability. This type of reading promotes the development of automaticity by allowing the reader to develop their fluency through the continued experience of effortless reading; it also promotes the formation of connections between words and their meanings by allowing the reader to establish connections through reading decodable texts. Finally, it allows for optimal conditions to promote incidental vocabulary learning. The practice of ER also creates an environment in which learners can become engaged emotionally, owing to the opportunity to select their own books, read materials that are at an appropriate level of difficulty for them, and create a no-stress reading environment, which in turn, sends positive messages of reading success and enhances reading motivation and confidence in the ability to read a foreign language. These theoretical perspectives justify the expectation that participants' reading fluency, vocabulary development, and reading motivation will be positively affected by ER.

Sociocultural theory (Vygotsky, 1978) is the second theoretical perspective of this research and defines learning as a social process. Learning is achieved through social interaction and support provided by others in the learner's Zone of Proximal Development (ZPD). Learners frequently confuse texts because when they read a text independently, they may not be able to understand what they read when facing unfamiliar words or complex structure. Teachers in a classroom setting help facilitate the learner's understanding through scaffolding by asking questions, clarifying information, or directing the learner's attention to essential features of the text. The role of ChatGPT in the study is to serve as a mediational tool that provides learners with dialogic scaffolding in the form of adaptive explanatory, paraphrases, prediction prompts, and comprehension checks. By having this type of interaction, the AI system can provide learners with timely assistance in closing gaps in comprehension and supporting continued engagement with the text, which corresponds with Vygotsky's view of guided assistance.

The third motivation theory strand is based on Deci and Ryan's (2000) Self-Determination Theory (SDT), which underscores autonomy, competence, and relatedness for intrinsic motivation. For learners, motivation increases if they are capable of understanding what they read, can make decisions about their learning, interact with others in a meaningful way, and generally feel competent in understanding the underlying content of their studies. ChatGPT can enhance learners'

motivation by providing learners with immediate, personalized feedback that enhances their sense of competently understanding texts, allowing learners to ask questions, guide the interaction, and access feedback from ChatGPT through interactive and responsive interactions.

In the present study, SDT complements the ER framework by explaining why AI-supported reading may produce stronger motivational outcomes, where ChatGPT can increase perceived competence through immediate clarification and feedback, preserve autonomy through learner-initiated questioning, and provide a sense of responsive support that can approximate relatedness. Together with sociocultural scaffolding, these motivational conditions are expected to increase reading persistence and engagement, which in turn supports the ER mechanisms of volume and repeated exposure underlying fluency and vocabulary development.

Empirical Studies

Studies on ChatGPT in EFL reading contexts are emerging, but the empirical base is not uniformly distributed across outcomes and designs. Kim et al. (2025) conducted a classroom-based study with undergraduates that incorporated structured, text-based dialogues with ChatGPT in reading lessons and provided evidence of significant growth in five dimensions of "ChatGPT literacy", including technical proficiency, critical evaluation, communication proficiency, creative application, and ethical competence. The participants indicated they generally trusted the AI tool but had some concerns over accuracy, describing adaptive, metacognitive strategies. While these findings illuminate how learners engage with AI tools and develop strategic awareness, the study primarily foregrounds dispositions toward the tool rather than objective gains in reading performance (e.g., fluency) or language development (e.g., vocabulary), thereby limiting its contribution to outcome-based reading research.

With respect to achievement outcomes, some quasi-experimental studies show that ChatGPT enhances vocabulary learning when it is integrated theory-informed and in an interactional way. Abdelhalim and Alsehibany (2025) adopted a sociocultural theoretical stance to compare vocabulary practice mediated by ChatGPT to learning through lecture. The experimental group experienced real-time feedback and AI-generated, adaptive vocabulary exercises, experiencing robust improvements on measures of productive vocabulary and overall measures compared to the control group with moderate to large effects, and across participants the analysis showed substantial improvements. Interactions between the learners and the chatbots were also qualitatively analyzed. They documented increased engagement alongside "situated" use of vocabulary and technology, aided by scaffolding. These results suggest that dialogic AI-mediated interaction can support lexical development; however, the focus on vocabulary alone leaves open questions about whether such benefits extend to broader reading processes central to extensive reading.

Aldowsari and Aljebreen (2024) also reported that learners in high school using a ChatGPT-based application had benefits posttest for vocabulary learning and measured positive attitudes towards vocabulary learning compared to a control condition. Although these findings reinforce the potential of AI-supported scaffolding for lexical learning and motivation, they similarly examine vocabulary outcomes in isolation and do not address reading fluency development or sustained engagement over time.

These studies broadly indicate that AI-led dialogic, scaffolded exchanges can lead to improvements in vocabulary learning and learner motivation. They are also both studies that while advancing the research in terms of using AI with lexis and vocabulary, still have only really examined dimensions of lexis and vocabulary learning, while dimensions of reading fluency

development and sustained reading motivation benefits for example in relation to deepening engagement, merit further consideration.

Evolving evidence on comprehension is emerging but remains methodologically limited. Yousefi and Askari (2024) found very large post-intervention gains in reading comprehension for university EFL learners who had been exposed to ChatGPT as compared to a control group over a ten-week period. The study specifically considered comprehension accuracy rather than fluency (rate/automaticity) and did not isolate which AI-mediated reading processes (e.g., previewing, questioning, lexical scaffolding) were responsible for gains. Furthermore, without an ER component, it is difficult to speculate if AI efficacy would scale within ER ecosystems that rely on volume, autonomy, and pleasure to build speed and vocabulary incidentally.

Lo et al. (2024) reviewed 70 empirical ESL/EFL studies during the 18 months following ChatGPT introduction that suggested research has focused too much on writing and short-term perception studies but not enough on reading, motivation, and performance outcomes using more rigorous designs. They suggest that longer-term quasi-experimental/experimental studies that use objective measures (e.g., standardized or validated tests) would lend credibility to our learning claims. The authors characterize the contemporary single-skill or perception-heavy investigations as a necessary basis for continuing our investigations, while noting a clear necessity for causal, incremental testing of ChatGPT in ecologically legitimate reading investigations (Kim et al., 2025).

The literature presents gaps that inform the current research. First, while there is encouraging evidence for vocabulary and attitude (Abdelhalim & Alsehibany, 2025; Aldowsari & Aljebreen, 2024), there has been limited cause to demonstrate if the insertion of ChatGPT into ER improves reading fluency, which is a primary outcome of ER beyond simply improving ER. Second, most previous research has not examined multiple outcomes; and as such, there is no way of knowing about trade-offs or tentatively complimentary outcomes. Third, few studies have implemented or utilized AI alongside a typical cycle of ER, with guided prompts that maintain the principles of ER with light-touch scaffolding. Finally, there is a contextual gap with private institutes in Kurdistan for a specific demographic.

3. METHODOLOGY AND DESIGN

Design

This study employed a pretest–posttest quasi-experimental design with intact classes randomly assigned to one of three instructional conditions: (a) Experimental Group 1 (Extensive Reading, ER), a standardized, ER program which uses level-appropriate graded readers, but does not include AI; (b) Experimental Group 2 (ER+ChatGPT), which uses the ER program and includes guided interactions with ChatGPT; and (c) Control, which continues the standard curriculum without any ER or AI. The independent variable is the instructional condition with 3 levels; (ER, ER + ChatGPT, Control). The dependent variables are (i) reading fluency, (ii) vocabulary gains, and (iii) reading motivation. Baseline measures are taken prior to the intervention and used as covariates, and posttest outcomes are compared across groups. The main effect of ER and the incremental effect of adding ChatGPT were also estimated.

Participants

The initial sample included 80 Kurdish English learners from a private language institute in Erbil, Iraq. Participants were selected based on the following criteria: (a) current enrollment in the

institute's general English program, (b) age between 18 and 23 years, (c) and regular attendance availability for the duration of the intervention. Following a placement test, 60 intermediate-level students (both male and female) were selected to participate in the study after meeting all criteria. Using intact classes to avoid mixing groups, these 60 participants were assigned to Experimental Group 1 for Extensive Reading (ER) only, Experimental Group 2 for integrated ER+ChatGPT, and Control (C); each group consisting of 20 students. All participants provided informed consent, and prior to the intervention, there was evidence of baseline equivalence across groups.

Instruments

Oxford Placement Test

The Oxford Placement Test (OPT; Allan, 2004) was conducted to assess the participants' initial proficiency level and to screen their selection for placement at the intermediate level. The OPT consists of 200 total items in two sections: 100 vocabulary questions, testing receptive lexical knowledge, and 100 listening questions. testing understanding of listening. All tests were taken under standardized conditions and were treated as total scores to check for eligibility. To enhance group comparability, only learners classified as intermediate based on OPT scores were retained using the test's standardized scoring guidelines, after which intact classes were randomly assigned to the study conditions. The OPT results also provided proficiency stratification in this study, serving as a covariate check to verify beginnings baseline equivalence in proficiency across the three groups.

Reading Materials

The primary reading resource for the program was *Top Notch 2B* (Saslow & Ascher, 2005), a textbook systematically aligned with intermediate-level proficiency according to its curricular sequencing, lexical load, and grammatical progression, with chapters selected to align with the 10-week intervention. The reading passages (approximately 250-450 words) in each unit present broadly familiar academic and experiential topics. They include pre- and post-reading tasks, glossaries, and skills work designed to foster fluency development while recycling targeted vocabulary. *Top Notch 2B* was carefully selected to comply with Krashen's (2004) principles concerning comprehensible input (i+1); it has an overall lexical complexity level of CEFR A2-B1, which is above the independently readable text complexity level of many intermediate learners as indicated by the OPT. As such, this was due to a combination of the graded structure of the vocabulary, the controlled syntactic complexity, and the familiarity of the topics related to learners' knowledge of English to allow them to comprehend the text, while at the same time developing learners' linguistic abilities according to the principles of extensive reading.

During the intervention, Experimental Group 1 independently engaged in sustained reading of *Top Notch 2B* passages and supplementary extension texts based on the same unit themes. Experimental Group 2 completed the same readings but received additional brief, guided ChatGPT activities, such as comprehension checks, vocabulary clarification, and paraphrasing prompts, which were researcher-designed to align with *Top Notch 2B* and calibrated for intermediate-level learners, serving to scaffold comprehension rather than modify the texts. This AI-assisted support served as a bridge that complemented the users' understanding and decreased their possibility of experiencing a temporary lapse in understanding, which aligns with Krashen's concept of comprehensible input as well as Vygotskian views on assisted learning. The Control group also used *Top Notch 2B* following the institute's regular syllabus but did not participate in ER routines or any AI-mediated activities.

Reading Fluency Pre and Posttest

Reading fluency was measured through two parallel sets of leveled passages (approximately 250–300 words per passage) aligned to the lexical and syntactic profile of *Top Notch 2B* and equated for length, average sentence length, and BNC/COCA frequency bands. After a one-minute practice passage, participants engaged in one minute of oral reading on the test passage at each of the time points (pre/post) with trained raters noting total words read, errors (mispronunciations, substitutions, omissions; self-corrections in less than three seconds were not considered errors), and words correct per minute (WCPM) as the primary dependent measure. To minimize potential for a form effect on reading fluency disparity between time points, pre- and post-passages were counterbalanced across intact classes, and conditions were standardized to the greatest extent possible. Secondary indices included accuracy rate (% words correct) and prosodic notes (e.g., phrasing/pausing) for qualitative triangulation. Inter-rater reliability was examined using double scoring on 25% of the recordings, and large outliers resulting from disfluency unrelated to reading (e.g., coughing/sneezing) were flagged for sensitivity analyses; the primary analyses examined all available data using an ITT approach.

Vocabulary Gain Test

Vocabulary learning was evaluated using a researcher-designed assessment tool based on the institution's official textbook, *Top Notch 2B* (Saslow & Ascher, 2005). The assessment consisted of 30 multiple-choice questions intended to sample target vocabulary (headwords, collocations, and phrase frames) from the unit word lists and reading selections. A test specification table (blueprint) was first prepared to ensure proportional representation of unit vocabulary and item types. Items were then drafted following the blueprint and reviewed for clarity, plausibility of distractors, and alignment with instructional objectives. Content validity was supported through systematic alignment with the learning objectives and vocabulary targets of the 10-week intervention and through expert review. Specifically, two experienced EFL instructors familiar with the *Top Notch* series evaluated each item for relevance, representativeness, and clarity; items were revised or replaced based on this feedback.

To assess changes in vocabulary learning, two parallel forms (i.e., Form A and Form B) with aligned test blueprints (content coverage, difficulty, and item format) were created; Form A was administered as the pretest and Form B was administered as the posttest. Parallel-form equivalence was examined through pilot testing and item analysis (item facility and discrimination), and forms were adjusted to ensure comparable difficulty profiles.

The test was designed to minimize cueing validity with items been reviewed by two experienced EFL instructors. The scoring system was dichotomous (1 = correct, 0 = incorrect; total score range of 0-30). Assessment contexts were standardized (fixed time limit, no dictionaries). Internal consistency reliability was examined using KR-20 coefficients, which yielded acceptable values for both forms (Form A = .82; Form B = .85), indicating satisfactory reliability. Basic item analysis (item facility and discrimination) and internal consistency checks were conducted before administration to verify test form equivalency.

Motivation Questionnaire

The 50-item motivation questionnaire developed by Wang and Jin (2021) was employed to measure learners' reading motivation. This questionnaire evaluated the following ten dimensions: self-efficacy, challenge, work avoidance, curiosity, involvement, importance, grades, competition, social, and compliance. Participants were provided with each item and asked to respond using a

four-point Likert-type scale (1 = very different from me; 2 = somewhat different from me; 3 = somewhat like me; and 4 = a lot like me). Each subscale and overall motivation index were computed and reported as subscale scores. Preliminary findings reported by Wang and Jin (2021) indicated that subscales demonstrated acceptable to good internal consistency ranging from .71 to .86 for Cronbach's α.

Data Collection Procedure

The purpose of the present study was to examine the effects of integrating ChatGPT into an ER program on EFL learners' reading fluency, vocabulary gains, and reading motivation. After receiving institutional approval and informed consent, all eligible learners were administered the OPT in controlled environments. The listening and vocabulary sections of the OPT were administered individually in paper-based format in a classroom setting, in which all participants were provided with an identical sequence of instructions and fixed time limits. Raw scores were used to screen learners for at least intermediate proficiency and to confirm baseline comparability across groups. Of the total pool (N = 80), 60 eligible students were retained and assigned by intact classes to one of three instructional conditions (ER, ER+ChatGPT, and Control).

On a different day, to relieve any fatigue, participants completed all baseline outcome measures, conducted in the following order: 1) the reading fluency test, which includes 1-minute oral reading following a practice passage; 2) vocabulary gain test, including Form A aligned to *Top Notch 2B* targets; and 3) motivation questionnaire with standard instructions. Across groups, testing rooms, seating arrangements, proctoring procedures, and time limits were held constant. In cases of student absenteeism, missed assessments were completed within 72 hours under identical conditions.

The intervention phase lasted 10 sessions, each 90 minutes in duration. In the ER group, students participated in sustained, reading solely from *Top Notch 2B* passages or approved extensions. Each week, students would also complete weekly reading logs consisting of the minutes read, the pages read, difficulty level, and a summary of approximately 1-2 sentences. In the ER+ChatGPT group, students followed the same ER routine, with the addition of a structured ChatGPT component lasting approximately 20–30 minutes per week, conducted in a supervised computer lab. ChatGPT interactions followed a fixed instructional sequence across all sessions, consisting of (a) pre-reading prediction prompts (e.g., activating background knowledge and anticipating content), (b) during-reading comprehension checks (e.g., clarification requests, paraphrasing of selected sentences, and confirmation of inferred meaning), and (c) post-reading vocabulary scaffolding (e.g., explanation of target words, example sentence generation, and synonym checks). All prompts were researcher-designed, standardized across sessions, and aligned with the weekly reading selections to ensure consistency across learners.

To ensure fidelity to the intended intervention, implementation checks were conducted throughout the study for both the ER and ER+ChatGPT conditions. In the ER+ChatGPT group, all ChatGPT interaction transcripts were exported at the end of each session and reviewed for fidelity purposes only. The transcripts were not coded for analytic purposes, as they were not treated as qualitative data; rather, they were examined to verify that interactions adhered to the predefined instructional prompts and procedures. Specifically, fidelity checks focused on confirming that ChatGPT use was limited to brief comprehension scaffolding (e.g., clarification, paraphrasing, vocabulary explanation, and prediction prompts), that no additional content was introduced, and that the original reading texts were not modified.

With respect to the ER intervention, fidelity was ensured by adhering to a consistent set of ER routines across sessions, including sustained silent reading of level-appropriate texts, student-

selected reading within assigned units, and the absence of explicit testing or form-focused instruction during reading time. The ER condition did not include AI support, comprehension questioning, or teacher-led analysis beyond routine monitoring, thereby preserving the core principles of extensive reading. In the control group, students followed the institute's standard syllabus, which used *Top Notch 2B* but did not incorporate ER routines or AI-based activities.

At the final day of the ten-week, all the participants re-used the measures of outcome under the same conditions: (1) a reading fluency test on a parallel passage, which was counterbalanced by class to avoid any form effects; (2) a vocabulary gain test, Form B, which was parallel to Form A in terms of blueprints and difficulty; and (3) the motivation questionnaire was given in full again in the same manner as the first test. The order of testing, room conditions, timing of testing, and proctoring matched the pretest. Students were instructed to not discuss the content of the test with their peers between various sessions; proctors collected all materials immediately following the completion of each section.

Raters completed calibration training using anchor recordings prior to scoring. A difference greater than 5 words-correct-per-minute (WCPM) or disagreement on error classification triggered adjudication by a third rater. Vocabulary tests were machine-scored, and item-level characteristics (item facility and discrimination), missing-response patterns, and timing distributions were examined to confirm parallel-form equivalence. Motivation subscale scores were computed according to the instrument manual, and internal consistency indices (α/ω) were examined for both pretest and posttest administrations. All datasets were anonymized using unique identifiers, double-entered to reduce transcription error, and stored on an encrypted drive. Primary analyses followed an intention-to-treat approach, and any protocol deviations or make-up testing were documented and examined through sensitivity analyses.

Data Analysis Procedure

The data were analyzed using an intention-to-treat approach with $\alpha = .05$. Given the pretest–posttest design and the primary interest in comparing post-intervention group differences while statistically controlling for baseline variation, ANCOVA was selected as the principal analytic model. For each outcome, reading fluency (WCPM), vocabulary score (0–30), and reading motivation (total score and subscales), ANCOVA models were fitted with instructional Group as a fixed factor and the corresponding pretest score entered as a covariate. This approach increases statistical power and reduces error variance associated with pre-existing individual differences.

In addition, confirmatory linear mixed-effects models (LMMs) were estimated to evaluate the robustness of the ANCOVA findings and to account for the nested structure of the data. These models included random intercepts for class and participant, with robust standard errors reported where appropriate. The convergence between ANCOVA and LMM results supported the stability of the findings; therefore, ANCOVA results are reported as primary, with LMM results used as sensitivity checks.

Two a priori contrasts were specified to test the study's primary hypotheses: (a) ER > Control (the ER effect) and (b) ER+ChatGPT > ER (the incremental AI effect). Linearity between the covariate and dependent variables, homogeneity of regression slopes across groups, and normality of residuals were examined through scatterplots, interaction tests between group and covariate, Q–Q plots, and Shapiro–Wilk tests. Where minor violations were detected, Box–Cox transformations or rank-based robust ANCOVAs were employed to confirm the consistency of results.

Effect sizes were reported as adjusted mean differences with 95% confidence intervals and partial η^2 for ANCOVA models, and as standardized β coefficients for LMMs. Holm–Bonferroni

adjustments were applied to control the familywise error rate across the six primary tests. Exploratory mediation analyses were conducted to examine whether reading volume (minutes/pages logged) partially accounted for group differences in outcomes, recognizing that these analyses were supplementary and interpretive in nature.

Inter-rater reliability for the reading fluency measure (based on a double-scored subset) and internal consistency for the motivation subscales (α/ω) were calculated at both pretest and posttest. Missing data were handled via multiple imputation (m = 20) under the missing-at-random assumption, with sensitivity analyses conducted and reported in the supplementary materials.

4. RESULTS

All analyses utilized an intention-to-treat procedure with α = .05 and Holm–Bonferroni correction applied to the three primary outcomes (fluency, vocabulary, motivation). For every dependent variable, a one-way ANCOVA was performed, assigning Group (Control, ER, ER+ChatGPT) as the fixed factor and the relevant pretest score as a covariate. Initial assessments confirmed the performance of ANCOVA: pretest means were similar among groups; the linear relationship between the covariate and posttest was satisfactory; the homogeneity of regression slopes was achieved; and the residuals exhibited no significant deviations from normality or homoscedasticity. In instances of minor assumption tensions, robust standard errors yielded identical inferences and are not reported separately. The covariate (pretest) was a significant positive predictor in all models, demonstrating suitable adjustment for baseline disparities. Adjusted posttest means (SE), omnibus F-tests, planned contrasts (ER > Control; ER+ChatGPT > ER), and effect sizes (partial η^2) with 95% confidence intervals are present. Descriptive pre/post means and standard deviations are included in the RQ-specific tables, while the supplementary materials contain full item-level diagnostics (vocabulary test) and inter-rater reliability (fluency).

Research Question 1: Reading Fluency

The Words Correct Per Minute (WCPM) index was used to measure reading fluency. It was based on one-minute oral readings of parallel passages given before and after the test. A one-way ANCOVA was performed to compare posttest WCPM scores between the three groups: Control, Extensive Reading (ER), and Extensive Reading with ChatGPT (ER+ChatGPT). Pretest fluency scores were used as a covariate. Initial assessments validated the assumptions of linearity, homogeneity of regression slopes, and normal distribution of residuals. The covariate (pretest fluency) was statistically significant ($\beta \approx .62$, $p < .001$), suggesting that participants with higher initial fluency tended to maintain higher posttest scores, thus supporting the appropriateness of ANCOVA.

Table 1 shows both the raw descriptive statistics (the means and standard deviations for the pretest and posttest) and the adjusted posttest means that come from the ANCOVA model. As can be seen, all three groups started at about the same level (about 82 WCPM). The Control group only got a little better (from 82.1 to 85.3 WCPM), but the ER group performed greatly better (from 81.7 to 96.4 WCPM). The ER+ChatGPT group had the biggest gain, going from 82.5 to 104.8 WCPM. After taking into account the differences before the test, the estimated marginal means were 86.0 (SE = 1.9) for the Control group, 98.5 (SE = 1.9) for the ER group, and 105.9 (SE = 1.9) for the ER+ChatGPT group. If all groups had started at the same level on the pretest, these adjusted means show what the group averages would become.

Table 1: Descriptive Statistics for Reading Fluency

Group	n	Pretest Mean (SD)	Posttest Mean (SD)	Adjusted Posttest Mean (SE)*
Control	20	82.1 (11.9)	85.3 (12.5)	86.0 (1.9)
ER	20	81.7 (12.3)	96.4 (13.0)	98.5 (1.9)
ER+ChatGPT	20	82.5 (12.0)	104.8 (13.2)	105.9 (1.9)

***Adjusted for pretest WCPM.**

Table 2 shows the ANCOVA inferential statistics, which include the omnibus test, the covariate effect, and the planned pairwise contrasts. The omnibus test for group effect was statistically significant, $F(2, 56) = 32.41$, $p < .001$, with a partial $\eta^2 = .54$, indicating a substantial effect size. This result substantiates that significant disparities in fluency outcomes existed among the three instructional conditions after accounting for pretest performance. Planned contrasts elucidated these differences: the ER group significantly surpassed the Control group with an adjusted mean difference of 12.5 WCPM (95% CI [8.2, 16.8], $p < .001$), whereas the ER+ChatGPT group outperformed the ER group by an additional 7.4 WCPM (95% CI [3.3, 11.5], $p < .001$).

Table 2: Reading Fluency (WCPM): ANCOVA and Planned Contrasts Results

Effect / Contrast	df	F / Δ (95% CI)	p	Effect size
Omnibus Group (ANCOVA)	2, 56	F = 32.41	< .001	partial $\eta^2 = .54$
Covariate: Pretest WCPM	1, 56	F = 54.72	< .001	partial $\eta^2 = .49$
Planned: ER − Control	—	Δ = +12.5 WCPM [8.2, 16.8]	< .001	$g \approx 0.95$
Planned: ER+ChatGPT − ER	—	Δ = +7.4 WCPM [3.3, 11.5]	< .001	$g \approx 0.56$

The ANCOVA results indicated that both ER and ER+ChatGPT significantly outperformed the control condition in reading fluency. Interpretation of these results was based on statistical significance, adjusted mean differences, and effect size estimates. The ER+ChatGPT group demonstrated a clear incremental advantage over the ER-only group, with an adjusted mean difference of approximately seven words correct per minute (WCPM). This difference was considered pedagogically meaningful, as gains of this magnitude have been associated with improved reading automaticity and efficiency in fluency-oriented research. Accordingly, the decision that ER+ChatGPT produced greater fluency gains was grounded in both statistical evidence and practical significance rather than statistical significance alone. These results indicate that the integration of AI-driven scaffolding (such as comprehension prompts, vocabulary clarification, or prediction activities) within extensive reading routines facilitates the maintenance of pace, decreases hesitations, and enhances reading performance.

Research Question 2: Vocabulary Gains

A teacher-made 30-question multiple-choice test was used to measure vocabulary acquisition based on *Top Notch 2B*. Two parallel versions of the test (Form A for the pretest and Form B for the posttest) made sure that the content was the same and that the test-retest bias was lower. The test questions assessed the ability to recognize definitions, contextual meanings, and collocations.

The KR-20 coefficients for the pretest and posttest were .82 and .85, which were good enough for reliability estimates. A one-way ANCOVA was conducted on posttest scores (out of 30), utilizing Group (Control, ER, ER+ChatGPT) as the independent variable and pretest score as the covariate. Table 3 presents the descriptive statistics and adjusted posttest means for each group. The covariate significantly predicted posttest scores ($\beta \approx .58$, $p < .001$), affirming that learners' initial vocabulary levels affected post-intervention performance.

Table 3 shows the descriptive and adjusted posttest means for the performance of participants did on the vocabulary test. The pretest means show that the groups started at about the same level (≈15/30). The Control group had only a small improvement at the posttest (M = 16.1, SD = 3.5), but the ER group had a moderate improvement (M = 21.2, SD = 3.6). The ER+ChatGPT group had the best scores (M = 23.7, SD = 3.4). When the differences between the pretests were taken into account, the adjusted means were 16.3 (SE = 0.4) for the Control group, 21.5 (SE = 0.4) for the ER group, and 24.0 (SE = 0.4) for the ER+ChatGPT group. This shows that there were clear stepwise improvements that matched the level of instruction.

Table 3: Descriptive Statistics for Vocabulary

Group	n	Pretest Mean (SD)	Posttest Mean (SD)	Adjusted Posttest Mean (SE)
Control	20	14.6 (3.2)	16.1 (3.5)	16.3 (0.4)
ER	20	14.8 (3.1)	21.2 (3.6)	21.5 (0.4)
ER+ChatGPT	20	14.9 (3.3)	23.7 (3.4)	24.0 (0.4)

Table 4 summarizes the inferential ANCOVA statistics. The omnibus group effect was statistically significant, $F(2, 56) = 46.18$, $p < .001$, with partial $\eta^2 = .62$, indicating a large effect size. The covariate was also significant, $F(1, 56) = 45.03$, $p < .001$, confirming its predictive value. Planned contrasts revealed that the ER group significantly outperformed the Control group (adjusted mean difference = +5.2 points, 95% CI [3.7, 6.7], $p < .001$; $g \approx 0.98$), and the ER+ChatGPT group outperformed the ER group (adjusted mean difference = +2.5 points, 95% CI [1.2, 3.8], $p < .001$; $g \approx 0.55$).

Table 4: Vocabulary: ANCOVA and Planned Contrasts Results

Effect / Contrast	df	F / Δ (95% CI)	p	Effect size
Omnibus Group (ANCOVA)	2, 56	F = 46.18	< .001	partial $\eta^2 = .62$
Covariate: Pretest Vocabulary	1, 56	F = 45.03	< .001	partial $\eta^2 = .45$
Planned: ER − Control	—	Δ = +5.2 points [3.7, 6.7]	< .001	$g \approx 0.98$
Planned: ER+ChatGPT − ER	—	Δ = +2.5 points [1.2, 3.8]	< .001	$g \approx 0.55$

The significantly omnibus and contrast results demonstrate that both interventions improved learners' vocabulary performance relative to traditional instruction, with the incorporation of ChatGPT leading to an additional, statistically significant benefit. Item-level analysis indicated that the most significant enhancement for the ER+ChatGPT group was observed in collocation and context-cloze items, suggesting that concise, focused interactions with ChatGPT likely enhanced lexical application and contextual comprehension during reading. In general, adding ChatGPT to extensive reading not only helped students learn more vocabulary, but it also helped them remember it better by making it more interactive and structured.

Research Question 3: Reading Motivation

The 50-item Motivation Questionnaire (Wang & Jin, 2021) was used to measure reading motivation. It used a 1–4 Likert scale (1 = very different from me; 4 = a lot like me). The total motivation score, which is the mean of all the items, was checked; the internal consistency was high (α = .88 before the test and α = .90 after the test). A one-way ANCOVA analyzed posttest motivation among the Control, ER, and ER+ChatGPT groups, based on questionnaire data collected before and after the intervention, with pretest motivation scores serving as the covariate. Assumption checks validated ANCOVA (linear covariate–DV relationship, homogeneous slopes, and residuals that are approximately normal and homoscedastic). The covariate significantly predicted posttest scores ($\beta \approx .60$, $p < .001$), indicating proper baseline adjustment.

Table 5 reports descriptive and adjusted means. All groups began at similar motivation levels (≈2.62–2.64). By posttest, the Control group showed a minimal increase, the ER group showed a moderate increase, and the ER+ChatGPT group showed the largest gain. Adjusted posttest means (estimated marginal means controlling for baseline) were 2.73 (SE = 0.04) for Control, 3.05 (SE = 0.04) for ER, and 3.24 (SE = 0.04) for ER+ChatGPT.

Table 5: Descriptive Statistics for Reading Motivation

Group	n	Pretest Mean (SD)	Posttest Mean (SD)	Adjusted Posttest Mean (SE)*
Control	20	2.63 (0.29)	2.70 (0.31)	2.73 (0.04)
ER	20	2.62 (0.30)	3.02 (0.33)	3.05 (0.04)
ER+ChatGPT	20	2.64 (0.31)	3.22 (0.34)	3.24 (0.04)

Table 6 summarizes the inferential results. The omnibus ANCOVA was significant, $F(2, 56) = 28.47$, $p < .001$, and partial $\eta^2 = .50$ (large). Planned contrasts, preregistered to match the study aims, showed that ER exceeded Control by +0.32 points on the 1–4 scale (95% CI [+0.22, +0.42], $p < .001$; Hedges' $g \approx 0.84$), and ER+ChatGPT exceeded ER by +0.19 points (95% CI [+0.10, +0.28], $p < .001$; $g \approx 0.52$).

Table 6: Reading Motivation: ANCOVA and Planned Contrasts Results

Effect / Contrast	df	F / Δ (95% CI)	p	Effect size
Omnibus Group (ANCOVA)	2, 56	F = 28.47	< .001	partial $\eta^2 = .50$
Covariate: Pretest Motivation	1, 56	F = 49.66	< .001	partial $\eta^2 = .47$
Planned: ER − Control	—	Δ = +0.32 [0.22, 0.42]	< .001	$g \approx 0.84$
Planned: ER+ChatGPT − ER	—	Δ = +0.19 [0.10, 0.28]	< .001	$g \approx 0.52$

After controlling for baseline motivation, both ER and ER+ChatGPT significantly outperformed the control condition. Most significantly, ER+ChatGPT led to an additional, medium-sized enhancement over ER. A +0.19 adjusted increment on a 1–4 scale indicates a significant change in learners' self-reported levels of engagement, enjoyment, and task value. These results, when combined with those on fluency and vocabulary, suggest that light, guided ChatGPT routines built into ER likely increased self-efficacy and perceived usefulness, leading to longer periods of reading.

In summary, the ANCOVA results for all three research questions demonstrated a consistent and coherent pattern: both ER and ER+ChatGPT yielded significant improvements in

reading fluency, vocabulary acquisition, and reading motivation compared to traditional instruction, with the ER+ChatGPT group consistently outperforming the ER group. These findings collectively illustrate that the incorporation of ChatGPT into extensive reading produces incremental and pedagogically significant enhancements that surpass those attained through extensive reading alone. The improved performance of the ER+ChatGPT group highlights the educational benefits of incorporating AI-driven dialogue and feedback into existing language learning frameworks. The data indicate that utilizing ChatGPT as a guided support tool—enhancing comprehension checks, lexical clarification, and tailored feedback— strengthens both cognitive and affective aspects of the reading process. The results confirm that integrating human-designed extensive reading with intelligent digital scaffolding constitutes an effective and advantageous way to enhancing EFL learners' literacy development and engagement.

5. DISCUSSION AND CONCLUSION

The purpose of the present study was to examine whether integrating ChatGPT into an ER framework enhances EFL learners' reading fluency, vocabulary gains, and reading motivation compared with ER alone and traditional instruction. To address this aim, a quasi-experimental design with intact classes was employed, using pretest–posttest measures and pretest-adjusted ANCOVA analyses to compare outcomes across instructional conditions. The results demonstrated that extensive reading led to significant improvements across all three outcome measures and, more importantly, that the ER+ChatGPT condition produced significantly greater gains than ER alone. These findings suggest that AI-mediated scaffolding can meaningfully enhance learner engagement, comprehension support, and lexical consolidation when embedded within an ER framework.

The pattern of ER versus Control aligns with the literature indicating ER's benefits for reading development. Meta-analytic syntheses (e.g., Jeon & Day, 2016; Karataş et al., 2024), consistently report small-to-medium gains for fluency and measurable gains for vocabulary when learners read large amounts of level-appropriate text. In addition, the documented fluency gains extend evidence from the classroom (e.g. Beglar & Hunt, 2014) that consistent exposure and volume of reading promote automaticity, illustrated through higher WCPM. The parallel gains in motivation appear to be aligned with the core principles of ER, choice, enjoyment, and low-stakes accountability, and have been cited regularly as underlying reasons for engaging in ER.

Aside from ER, the additional benefit of ER+ChatGPT over ER in fluency, vocabulary, and motivation gives us causal evidence that engagement with dialogic AI can be a pedagogical approach that supports language educational development. For vocabulary, the findings of the present research align with quasi-experimental studies illustrating that ChatGPT-mediated, scaffolded interaction and dialogic practices with ChatGPT lead to meaningful outcomes for lexical growth and learner engagement (Abdelhalim & Alsehibany, 2025; Aldowsari & Aljebreen, 2024).

One plausible mechanism for the effects can be the brief, structured prompts (prediction, Socratic checks, targeted glossing, collocations, etc.) that promote an acceleration of noticing and perpetuate recycling without ER requiring test-like work and possibly chronic learning. For reading performance, our findings enhance other work that reports benefits for AI-supported reading comprehension outcomes (Yousefi & Askari, 2024), but more benefits are offered in the current research by showing effects for fluency, another meaning dimension (rate/accuracy) that previous AI research rarely measured. The fluency gain could reflect how prompt-supported staged engagement increases a purposeful reading and enhances pacing (e.g., using summarized rehearsal), decreasing breakdowns when comprehension breaks down.

From the emotional/strategic side, motivation effects align with students' reports of high perceived usefulness, with careful consideration of AI (Kim et al. 2025). Students in the ER+ChatGPT condition received just-in-time support (e.g., clarification questions, tips about pronunciation, goal-oriented summaries) that likely increased self-efficacy and task-value as theoretical motivation, along with qualitative themes of engagement and agency documented in previous studies focused on blending/using AI. Notably, the present design goes beyond just perception only work by demonstrating objective outcome gains in controlled-comparisons.

Ultimately, our research addresses each of the gaps noted in Lo et al.'s (2024) comprehensive scan. It examines reading, utilizes a robust multi-arm design, and gathers validated, objective data on three outcomes at once (fluency, vocabulary, motivation). By embedding ChatGPT inside an authentic ER program and not treating it as a separate tool, we are investigating its incremental value, rather than the additive value over an established baseline. Together, the pattern suggests a practical point for practitioners: retain the central tenets of ER, and then layer on low-level, guided ChatGPT routines to help with comprehension and lexical consolidation; this was a dose which appeared to provide additional value without subverting the purpose of ER. Future studies could explore mechanism (e.g., mediation via reading volume or strategy use), dosage (frequency/length of AI interactions), or durability (Posttest after a delay and transfer to independent reading).

This research offers causal evidence that both ER and ER supplemented with guided ChatGPT routines improve reading fluency, vocabulary gains, and reading motivation for intermediate EFL learners, with the ER+ChatGPT condition yielding additional benefits over ER alone. Regarding context, the participants were Kurdish learners aged 18–23 from a private institute in Erbil, Iraq. They completed *Top Notch 2B* and ER activities on a weekly basis. Accordingly, the findings are most directly generalizable to similar private or institute-based EFL programs that can ensure sustained reading time alongside light-touch, supervised AI support. More cautious generalization may extend to other monolingual EFL contexts with comparable curricular structures, instructional expertise, and access to digital resources.

From a pedagogical perspective, the findings support a practical “ER-first, AI-light” instructional model that preserves the core principles of extensive reading while strategically incorporating short, guided ChatGPT interactions for previewing, comprehension monitoring, and vocabulary recycling. Programs may benefit from (a) embedding brief weekly AI-supported routines (approximately 20–30 minutes) linked to reading selections, (b) developing educator AI literacy and instructional guardrails, and (c) aligning assessment practices to capture reading rate, incremental vocabulary growth, and learner engagement. At the institutional level, policies should promote equitable access to devices, communicate expectations regarding academic integrity, and provide introductory AI-orientation workshops so that the benefits of AI-supported learning are not restricted to students with greater personal technological resources. For instructors, the use of structured prompt guides and fidelity checklists may facilitate feasible integration without necessitating major curricular redesign.

Limitations raise important priorities for future research. The initial sample was drawn from a single private language institute and consisted of 60 intermediate Kurdish EFL learners (ages 18–23), which necessarily constrains the generalizability of the findings. Consequently, the results should be interpreted cautiously when extended to other educational settings, age groups, or proficiency levels. Although intact classes were randomly assigned to conditions, the absence of individual-level random assignment represents a limitation of the study. The use of ANCOVA helped control for pre-existing group differences by adjusting for baseline scores, but this statistical approach cannot fully eliminate the possibility of selection bias. Therefore, the findings

should be interpreted with appropriate caution, and future research employing randomized controlled designs is needed to further validate these results.

In addition, the duration of the intervention and the absence of delayed posttest data limit claims regarding the long-term durability of observed gains; future studies should therefore examine retention and transfer to independent reading over extended periods. Moreover, although motivation and vocabulary learning were assessed using reliable instruments aligned with the instructional content, future research could incorporate external standardized measures and behavioral or usage-based analytics to better model learning mechanisms and instructional dosage. Although the vocabulary test demonstrated acceptable internal consistency and content alignment with the intervention, external validation evidence and test–retest reliability were not collected, which limits broader interpretability beyond the targeted textbook vocabulary domain. Potential novelty effects and access inequities related to device availability and connectivity also warrant further investigation through longitudinal or cost–benefit analyses.

References

Abdelhalim, S. M., & Alsehibany, R. (2025). Integrating ChatGPT for vocabulary learning and retention: A classroom-based study of Saudi EFL learners. *Language Learning & Technology*, *29*(1), 1–24. https://doi.org/10.64152/10125/73635

Aldowsari, B. I., & Aljebreen, S. G. (2024). The impact of using a ChatGPT-based application to enhance Saudi students' EFL vocabulary learning. *International Journal of Language and Literary Studies*, *6*(4), 380-397. https://doi.org/10.36892/ijlls.v6i4.1955

Allan, D. (2004). *Oxford placement test 1: Test pack*. Oxford University Press.

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language*, *26*(1), 29–48. https://doi.org/10.64152/10125/66684

Bushnell, A. (2025). *Using AI to power primary literacy: A classroom guide for teachers*. Taylor & Francis.

Chen, J., Liu, Z., Huang, X., Wu, C., Liu, Q., Jiang, G., ... & Chen, E. (2024). When large language models meet personalization: Perspectives of challenges and opportunities. *World Wide Web*, *27*(4), 42. https://doi.org/10.1007/s11280-024-01276-1

Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, *14*(2), 136–141. https://doi.org/10.64152/10125%2F66761

Day, R. R., Bamford, J., Renandya, W. A., Jacobs, G. M., & Yu, V. W. S. (1998). Extensive reading in the second language classroom. *RELC Journal*, *29*(2), 187-191. https://doi.org/10.1177/003368829802900211

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

El Hassan, F. A. M., & Alsalwah, A. F. (2025). Exploring the impact of ChatGPT on EFL reading practices: Opportunities and challenges. *International Journal of English Language Teaching*, *13*(1), 85-93. https://doi.org/10.37745/ijelt.13/vol13n18593

Hunter, A. M. (2017). *Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency* [Doctoral dissertation, St Mary's University, Twickenham].

Jeon, E.-Y., & Day, R. R. (2016). The effectiveness of extensive reading on reading proficiency: A meta-analysis. *Reading in a Foreign Language*, *28*(2), 246–265. https://doi.org/10.64152/10125/66901

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, *29*(15), 19343-19366. https://doi.org/10.1007/s10639-024-12574-6

Kim, N. Y., Cha, Y., & Kim, H. S. (2025). Exploring ChatGPT literacy in EFL reading education. *Multimedia-Assisted Language Learning*, *28*(1), 23-41. https://doi.org/10.15702/mall.2025.28.1.39

Krashen, S. D. (2004). *The power of reading: Insights from the research*. Bloomsbury Publishing USA.

Liu, S., & Saad, M. R. B. M. (2025). Role of extensive reading in vocabulary development, reading comprehension, and reading speed: A systematic literature review. *Eurasian Journal of Applied Linguistics*, *11*(1), 87-99. https://doi.org/10.32601/ejal.11108

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*(1), 50-60. https://doi.org/10.1186/s40561-024-00342-5

Saslow, J. M. & Ascher, A. (2005). *Top Notch 2B*. Pearson.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, X., & Jin, Y. (2021). A validation of the Chinese motivation for reading questionnaire. *Journal of Literacy Research*, *53*(3), 336-360.

Yousefi, M., & Askari, M. I. (2024). Exploring the effectiveness of artificial intelligence (AI) on reading comprehension among Iranian EFL learners. *Journal of New Trends in English Language Learning*, *3*, 1-17. https://doi.org/10.71528/2024.202406221123351