Language Teaching Research Quarterly **EUROKD**

2026 (54), 385–409
https://doi.org/10.32038/ltrq.2026.54.16

Integrating Reading Bots with Guided Reading Strategies to Enhance EFL Nursing Students' Comprehension of Medical Texts

Nada S. Al Ruwaili
North Private College of Nursing, Arar, Saudi Arabia
Abdul Aziz Mohamed Mohamed Ali El Deen*
Department of English Language and Literature, College of Languages and Translation, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia
Shahad Z. Aldandan
North Private College of Nursing, Arar, Saudi Arabia
Noura A. Alhudaib
North Private College of Nursing, Arar, Saudi Arabia
Mohammed A. Alrashed
Department of English Language and Literature, College of Languages and Translation, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia
Amr M. Mohamed
North Private College of Nursing, Arar, Saudi Arabia

Correspondence
Email: AmAlieddin@imamu.edu.sa

Abstract

This mixed-methods study examined the role of chatbot-supported guided reading in EFL reading comprehension among undergraduate nursing students. Using a quasi-experimental design, 84 participants in experimental and control groups completed pre- and post-tests assessing overall comprehension and sub-skills. The experimental group used chatbot-guided reading, while the control group received traditional instruction. Semi-structured interviews with 27 participants explored learners' perceptions. Quantitative results showed significant improvement in overall comprehension and lower-level skills (e.g., skimming), but not critical reading. Qualitative findings indicated that chatbot-guided reading enhanced motivation and reduced anxiety through instant feedback, simplified explanations, and a non-judgmental environment. However, concerns emerged regarding response reliability and limited cognitive engagement. Overall, chatbot-guided reading supports basic skill development and affective engagement but is insufficient for fostering higher-order critical reading skills. The research suggests that a mixed teaching method that combines reading tasks based

ARTICLE HISTORY
Received: 29 December 2025
Revised: 17 April 2026
Accepted: 15 May 2026

KEYWORDS
Reading Bots, Guided Reading, EFL Nursing Students, Medical Comprehension, AI-Assisted Learning

on AI support with training on critical thinking conducted by an instructor is recommended.

How to cite this article (APA 7th Edition):
Al Ruwaili, N. S., Ali El Deen, A. M. M., Aldandan, S. Z., Alhudaib, N. A., Alrashed, M. A., & Mohamed, A. M. (2026). Integrating reading bots with guided reading strategies to enhance EFL nursing students' comprehension of medical texts. *Language Teaching Research Quarterly*, *54*, 385–409. https://doi.org/10.32038/ltrq.2026.54.16

Introduction

Nursing students often struggle to understand complex medical literature, which is a key to academic success and further practice (Nghia & Yen, 2018). Recent research continues to confirm these challenges while emphasizing the importance of structured and technology-supported interventions to improve reading outcomes in specialized domains (AlTwijri & Abdelhalim, 2026; Huang et al., 2026). Guided Reading is a long-standing approach to education that aims to support students in comprehending and interpreting texts (Duke et al., 2021). According to student perceptions, guided reading questions have positive effects on students' motivation, reading comprehension, effort level, and prior knowledge of the material (Brown et al., 2016). In addition, Liu et al. (2022) found that both triggered and sustained situational interest were associated with perceived social connection with the chatbot and with students' engagement in the reading activity. More recent studies also highlight that AI-supported learning environments can enhance learner engagement and motivation when integrated with pedagogical strategies such as guided reading (Griche & Bennis, 2026).

The use of such a strategy in nursing education may enhance reading comprehension, yield better academic results, and increase students' confidence in their ability to work with medical literature.

Medical texts are difficult to read due to specialized language and the high level of information, which may not be readily comprehended by students of non-English languages. This ignorance can lead to poor school performance as well as in the clinic. Despite the importance of reading comprehension, not many studies have investigated the application of guided reading strategies in medical education.

To evaluate the effectiveness of the Guided Reading Strategy in enhancing nursing students' understanding of medical texts, explicitly focusing on vocabulary acquisition, critical thinking, and the ability to interpret complex content.
In this study, we aim to answer the following questions:
RQ_1: Is there a statistically significant improvement in nursing students' reading comprehension scores (critical reading, skimming and scanning, vocabulary acquisition, and inference and deduction) after integrating reading AI chatbots with guided reading strategies?

RQ2: Do parental educational attainment, GPA, and gender significantly moderate the improvement in reading comprehension skills following the intervention with reading AI chatbots and guided reading strategies?
RQ3: What is the effect of using AI chatbots in guided reading on motivation, anxiety, and cognitive load in nursing students' perspectives?

Theoretical Framework

The current research is based on a combined theoretical approach comprising sociocultural learning theory, cognitive load theory, and effective motivational perspectives. The given framework is not like a list of applicable theories but the very conceptual framework according to which the intervention is planned to be designed, implemented, and analyzed, chatbot-assisted guided reading. These theoretical lenses shed light on how the chatbot was expected to improve EFL nursing students' reading comprehension and critical reading skills (Pan et al., 2024; Shafiee Rad, 2025).

From a sociocultural perspective, learning is a mediated social process where tools and signs support higher mental functions (Vygotsky, 1978). This is directly applied in guided reading, as a teacher or chatbot provides scaffolded interaction within the learner's Zone of Proximal Development (ZPD) to develop reading strategies (Kitsili & Murray, 2024; Le & Nguyen, 2024). This study frames the AI chatbot as a digital scaffold or mediator. It was developed not as an information warehouse but as an interactive companion to involve students in dialogic reading comprehension processes, expanding guided reading practice beyond the dimensions of time and space within the classroom. The most important research question in this perspective is whether reading development can be scaffolded by AI-mediated social interaction, thereby facilitating it in a similar way or complementing the effects of human guidance (Pan et al., 2024; Yetişensoy & Karaduman, 2024).

The micro-level cognitive mechanism that supplements the sociocultural macro-level is cognitive load theory (Paas, 1992; Sweller, 1988). It hypothesizes that learning is maximized when instruction design addresses the intrinsic load (the complexity of the material), minimizes extraneous load (processing information that is irrelevant), and considers Germane load (cognitive resources utilized in schema construction) (Mayer, 2024). This is because, in this study, the guided reading questions aim to control intrinsic load by dissecting texts. It was theorized that the chatbot would decrease extraneous load, namely by offering on-demand (immediate) clarifications on vocabulary and key ideas (Mohamed et al., 2026a; Wardhono et al., 2024), thereby enabling learners to redirect cognitive resources (germane load) toward a deeper understanding and critical analysis. There is, however, critical tension: these responses from chatbots can increase extraneous cognitive load, especially in complex, critical reading tasks that demand subtle judgment; therefore, they may be counterproductive to the very processes they are supposed to assist.

The third pillar encompasses affective-motivational theories, particularly self-determination and foreign language anxiety. The chatbot was used to create a low-stakes, 24/7 environment to reduce the cognitive burden of reading complex L2 nursing texts (Krashen, 1982). It sought to increase motivation and decrease anxiety by enhancing learners' autonomy and providing tailored feedback, leveraging perceived competence and a favorable affective climate regarding cognitive involvement (Teng et al., 2024).

This model considers chatbot-guided reading as an improvement of comprehension via cognitive and affective functions (Pan et al., 2024; Shafiee Rad, 2025). The chatbot acts as a sociocultural mediator (Molenaar, 2022), and its effects are determined by cognitive load and positive effects. Mixed-methods design is a design that combines outcome measures with learners' experiences (Teng et al., 2024). The framework also indicates the varied impacts of basic and critical reading due to their demands.

Guided Reading and Diverse Learning Needs

Guided reading is a long-standing technique used to assist students with various linguistic and cognitive requirements. It facilitates understanding and supports independent reading. Nevertheless, its application in most learning settings has been problematic due to resource scarcity. Kitsili and Murray (2024) found that group-guided reading is underutilized in South Africa due to high student-to-teacher ratios and inadequate teaching resources. These limitations underscore the importance of scalable, adaptable models that can complement teachers' instruction on a periodic basis.

Along with the established guided reading, learner-driven, and multimodal strategies that can underpin active, playful, and embodied learning, these strategies can also be used to promote literacy and language success across all disciplines. These findings suggest that the process of reading development is strongly related to the work of extended cognition, as well as affective mechanisms, including motivation, interest, and cognitive effort. Recent studies further confirm that AI-supported reading environments can enhance comprehension, motivation, and cognitive engagement among language learners (AlTwijri & Abdelhalim, 2026; Huang et al., 2026), while broader reviews emphasize the growing role of generative AI in reshaping foreign language instruction (Griche & Bennis, 2026). Similarly, CAI was established to have a positive impact on reading in students with autism spectrum disorders and intellectual disabilities because this method provides close pacing and no-judgment feedback (Kurzeja et al., 2024). Collectively, these studies highlight the need to have flexible reading interventions that consider the needs of multi-skilled learners and the context in which they are learning.

Critical reading is particularly relevant when evaluating and countering information. It has already been established that active learning strategies may enhance secondary students' critical reading skills (Phimphimon et al., 2024). Competency in critical reading is essential for academic success and evidence-based practice in professional learning

and education, especially in nursing. According to previous studies, reading tests can be used to assess nursing students and predict academic and professional success (Mitchell, 2024).

However, one of the most common challenges for nursing students is developing core reading subskills, such as identifying primary ideas, skimming, and working with field-specific vocabulary (Wardhono et al., 2024). Despite interventions such as reading automated feedback, which have enhanced reading performance and classroom participation, the overall critical thinking skills of nursing students remain moderate and not uniformly developed (Basco-Prado et al., 2024). Previous systematic reviews also indicate that medical and nursing learners with academic difficulties benefit from structured educational interventions and targeted academic support (Montreuil et al., 2025). They are frequently complicated by limited exposure to strategic reading instruction and insufficient instructional support (Le et al., 2024), indicating that creative, scaffolded reading methods should be introduced into nursing education.

AI Chatbots in Reading Instruction

Recent developments in educational technology have sparked increased interest in AI-driven chatbots to assist language learning, including reading instruction. In this new research direction, generative AI applications like ChatGPT have been theorized as learning scaffolds that facilitate cognitive and affective facets of learning when used in tandem with the teaching goals (Ali El Deen et al., 2026; Barrot, 2024; Dong, 2024). Empirical studies indicate that chatbots may help relieve foreign-language reading fear and improve student performance and confidence (Zheng, 2024), while recent research also highlights their role in shaping learners' motivation and emotional engagement in AI-mediated learning environments (Goktas et al., 2024; Mohamed et al., 2026b).

Individualized, self-directed reading that offers instant feedback, strategy recommendations, and clarifications related to demands is also a beneficial application of chatbots (Jmaiel et al., 2025; Pan et al., 2024). Liu et al. (2024) also found that learning-by-teaching approaches supported by chatbots are more likely to enhance learners' interest and engagement in the reading process. In terms of sociocognitive factors, chatbots can help learners converse with texts and thereby acquire and understand vocabulary, especially among more sophisticated learners (Cheng et al., 2024; Mohamed et al., 2026a).

Even with its benefits, literature also has significant shortcomings. There is still concern about the quality and consistency of chatbot responses, the effectiveness of automated questioning strategies, and the levels of cognitive engagement of learners in AI-mediated interactions (Alfaleh et al., 2025; Cheng et al., 2024; Liu et al., 2024). Additionally, although receptive skills, especially reading, are understudied, productive skills, including writing and speaking, among others, have been the targets of much existing

generative AI research (Barrot, 2024; Baskara et al., 2024). The imbalance highlights the importance of empirical research examining the effects of chatbot-assisted instruction on various aspects of reading comprehension.

AI Applications in Nursing Education

Recent research further indicates that AI-driven instructional environments can enhance learner engagement, motivation, and adaptive learning when integrated within structured pedagogical designs (Ali El Deen et al., 2026; Mohamed et al., 2026b), while also emphasizing the importance of teacher mediation and pedagogical alignment in technology-enhanced language education (Xerri et al., 2025).

AI chatbots in the nursing curriculum have also shown potential to increase students' access to current clinical knowledge and aid decision-making (Makhlouf et al., 2024). Researchers criticize the fact that the application of AI can harm academic integrity, foster cognitive dependence, and lead to inaccuracies when used blindly. Sociocultural and constructivist models based on ZPD by Vygotsky, in turn, tend to focus on guided interaction and critical assessment and advocate hybrid solutions that combine AI tools with teacher-centered teaching (Dong, 2024; Simms, 2024).

The Study

Although several previous studies indicate that AI chatbots have pedagogical potential for nursing training and language learning (Ali et al., 2024; Makhlouf et al., 2024), several significant gaps remain. First, empirical studies on the use of chatbots as guided reading aids have not been conducted to date, particularly within department-specific fields such as nursing education. Second, the literature is one-sided, as it primarily focuses on engagement and basic understanding outcomes and pays insufficient attention to more advanced reading skills, including critical reading and critical evaluation. Third, no previous research has evaluated the affective and cognitive impacts of chatbot implementation, as well as motivation, anxiety, and cognitive load, simultaneously in a single study. In order to fill these gaps, the current study investigates how chatbot-mediated guided reading affects the reading comprehension and affective reactions of EFL students in nursing, thus becoming a part of a bigger picture of AI-based reading instruction within the context of the professional learning process.

Methodology

Participants, Context, and Procedure

The study was conducted at North Private College of Nursing during the second semester of the 2025 academic year (January–May) in an EFL nursing course that included discipline-specific reading material. The sample consisted of 84 undergraduate nursing students, selected through stratified random sampling to achieve proportional representation across academic years. The sample was small, comprising first- and second-year students (88.1 and 9.5 percent, respectively), and a few third and fourth-

year students (2.4). The majority of respondents were young adults, with 64.3% aged 18 to 24 and 25% aged 18 and below, which aligns with the institution's demographic profile.

Subsequently, following a baseline test, the participants were randomly divided into an experimental group (n=46), in which the applicant was guided through chatbot-supported reading, and a control group (n=38), in which the applicant was guided through instructor-supported reading. Reading proficiency was evenly distributed across the groups. Time and content factors were eliminated by having the same materials, objectives, and assessments, and combining the three-month intervention with regular coursework. Informed consent was given by all the participants, and their right to withdraw without academic penalty was guaranteed. The institution gave ethical approval.

Instruments

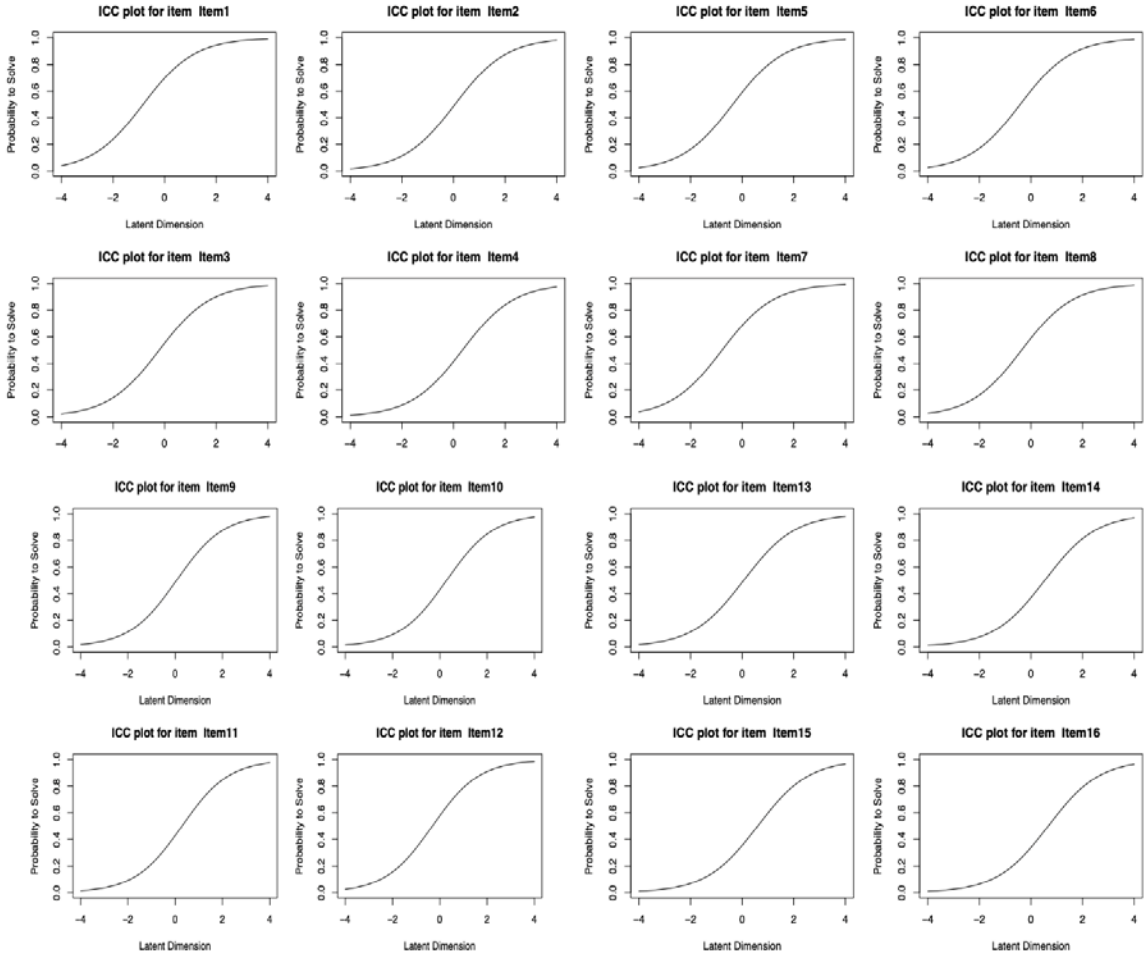
The study included a pre-post test comprising 20 items with four primary reading skills (critical reading, skimming/scanning, vocabulary acquisition, and inference) using a medical text about cardiac anatomy. The measuring instrument was a balanced design (5 items per domain of the skills, 1 mark per item) to assess baseline proficiency and intervention gains among EFL nursing students. The complexity and types of questions in the passage (evidence-based MCQs to measure critical reading, factual recall to measure skimming) were also purposely designed to assess surface-level understanding and critical reading abilities. The subgroup analysis of the influence of technological familiarity was possible because demographic information (age, year level, and device/AI tool usage) was collected. There was consistency in the scoring between pre- and post-administration, and a special concern about the quality of distracters in MCQs (why not the wrong option for possible cardiac arrest due to valve failure in Question 4). This model enabled a direct comparison of chatbot-assisted and traditional guided reading results in controlling text difficulty and cognitive load.

Test Validation

After collecting the pre-test responses, we conducted an initial psychometric validation of the test items using the Rasch model. We first tested the item difficulty using the item characteristic curve (ICC), item difficulty parameters, and the person-item map. Figure 9 presents a sample of ICC plots, which assess the probability that a participant will answer an item correctly based on their ability level.

Figure 1

ICC Plot for the First 16 Items



The ICC plots are sigmoidal, indicating that different items do not make sharp jumps among students with different ability levels and instead gradually differentiate among them. The bottom part of the curve (the flat part near 0 probability) shows that less-skilled participants are less likely to provide the correct answer (according to the literature). The transition zone is indicated by the middle of the curve, where the probability of a correct response is directly proportional to ability. The high-ability participants are quite likely to get it right (the high-value portion of the curve), suggesting that the probability is very large (Lee & Bolt, 2018; Rosenbaum, 1987). Based on these measures, the exam seems to be item-balanced. In order to ensure this, we analyzed the item difficulty estimates (see Table 1). The difficulty parameters range from 0.835 to -0.664, indicating a quite wide distribution of item challenges along the ability scale. This dispersion of about 1.5 logits indicates that the test can successfully differentiate between people with low, average, and slightly above-average ability levels. Notably, the average item difficulty is close to zero, similar to the Rasch scaling, which sets items at the average item difficulty. In summary, Table 1 supports the ICC-based interpretation: the test provides a generally balanced range of item difficulties, with a few items targeting high and low ends of the ability continuum.

Table 1

Item Difficulty and Ease of Estimation

	Estimate	Std. Error	95% CI	
			Lower	Upper
Item1	0.835	0.224	0.396	1.275
Item2	-0.031	0.230	-0.509	0.391
Item3	0.211	0.224	-0.229	0.651
Item4	-0.346	0.239	-0.814	0.121
Item5	0.368	0.223	-0.069	0.805
Item6	0.420	0.223	-0.016	0.856
Item7	0.783	0.224	0.344	1.221
Item8	0.368	0.223	-0.069	0.805
Item9	-0.059	0.230	-0.509	0.391
Item10	-0.287	0.236	-0.750	0.176
Item11	-0.287	0.236	-0.750	0.176
Item12	0.316	0.223	-0.121	0.754
Item13	-0.315	0.230	-0.509	0.391
Item14	-0.533	0.246	-1.016	-0.050
Item15	-0.598	0.250	-1.087	-0.109
Item16	-0.664	0.253	-1.160	-0.169
Item17	-0.004	0.228	-0.451	0.444
Item18	-0.287	0.236	-0.750	0.176
Item19	0.051	0.227	-0.394	0.496
Item20	-0.171	0.233	-0.627	0.285

Given the limited distribution of difficulty or ability as indicated by the person-item map, we further our analysis to understand the level of item discrimination and the overall model fit (see Table 2). The p-values associated with the chi-square statistics indicate that all items exhibit an acceptable fit to the Rasch model, with none showing a statistically significant misfit at the 0.05 level. Regarding the fit indices, both Outfit and Infit Mean Squared Error (MSE) values are largely within the commonly accepted range of 0.7 to 1.3, supporting the appropriateness of the items. In terms of discrimination, most items show a moderate to strong ability to differentiate between respondents of different ability levels. Overall, these results support the test's internal consistency and discriminative capacity. Therefore, we retained all 20 items for the study.

Interview

The survey instrument was designed based on users' experiences with AI chatbots when reading medical texts. It also quantifies the frequency and use of chatbots to simplify text, support vocabulary, and summarize, as well as the challenges, motivations, and anxiety experienced in learning a foreign language. The validated scale used to determine perceived mental effort was that developed by Paas (1992). The utilization of an item mental effort scale is in line with the cognitive load research tradition. The scale was first developed by Paas (1992) as a subjective rating of invested mental effort across the world, and it has been widely applied in experimental instructional design studies because it is sensitive to instructional manipulations and is not overly intrusive during learning tasks. According to research in the context of cognitive load theory, single-item mental effort testing is expected to yield appropriate validity rates and is typically used

when cognitive load is measured immediately after the learning process (Paas, 1992; Sweller, 1988). Since the current study is a quasi-experimental classroom study, the application of the Paas scale caused minimal disruption to measurement, and participants perceived minimal cognitive effort during reading assisted by AI.

Table 2

Itemfit Statistics

	Chisq	Df	p-value	Outfit MSQ	Infit MSQ	Discrim
Item1	88.300	83	0.325	0.851	0.921	0.461
Item2	87.527	83	0.346	0.742	0.910	0.480
Item3	61.733	83	0.961	0.735	0.787	0.625
Item4	66.632	83	0.905	0.793	0.882	0.483
Item5	67.920	83	0.884	0.809	0.838	0.558
Item6	85.899	83	0.392	0.723	0.846	0.434
Item7	63.259	83	0.947	0.753	0.797	0.575
Item8	77.527	83	0.649	0.923	0.973	0.335
Item9	92.122	83	0.231	0.897	0.989	0.582
Item10	95.646	83	0.162	0.739	0.845	0.586
Item11	82.491	83	0.495	0.982	0.929	0.570
Item12	67.548	83	0.891	0.804	0.846	0.531
Item13	82.689	83	0.489	0.984	0.918	0.473
Item14	87.531	83	0.172	0.821	0.811	0.414
Item15	98.127	83	0.123	0.868	0.885	0.452
Item16	91.127	83	0.254	0.785	0.884	0.589
Item17	88.462	83	0.320	0.753	0.925	0.379
Item18	86.149	83	0.385	0.826	0.914	0.309
Item19	84.992	83	0.419	0.812	0.821	0.478
Item20	83.984	83	0.159	0.938	0.948	0.496

The questionnaire was validated through expert review by medical educators and linguists, cognitive interviews, and a pilot survey. Feedback-informed refinements to wording and response scales. Q6 used Paas's (1992) mental effort scale; motivation and anxiety items included open-ended prompts for qualitative data. This process strengthened the instrument's validity prior to implementation.

Data Analysis

Participant demographics (age, education, parental education) were summarized using descriptive statistics (frequencies and percentages) to provide context for interpreting the findings. A mixed-design ANOVA was used to address the research questions, incorporating within-subjects (pre–post scores) and between-subjects (chatbot vs. no chatbot) factors. This allowed analysis of changes over time, group differences, and interaction effects, enabling a robust assessment of the impact of chatbot-mediated guided reading on students' critical reading and comprehension.

Ethical Approval

This study was approved by the Research Ethics Committee of North Private College, Saudi Arabia, in accordance with the Declaration of Helsinki. Participation was voluntary,

with informed consent obtained electronically. All data were anonymized to ensure confidentiality, and participants could withdraw at any time without consequence.

Results

Demographic Characteristics

Table 3 shows that most participants were aged 18–24 (64.3%) or under 18 (25.0%), with fewer aged 25–30 (8.3%) and above (2.4%). First-year students comprised the majority (88.1%), followed by second-year students (9.5%), third-year students (1.2%), and fourth-year students (1.2%). Regarding parental education, 58.3% of fathers had a basic education, 28.6% had a college education, and 13.1% had a graduate degree. Maternal education was slightly higher: 46.4% held college degrees, 44.0% basic education, and 9.5% graduate degrees, providing context for participants' backgrounds.

Table 3

Demographic Profile of Participants

Profile	Frequency	Percent
Age		
Under 18	21	25.0
18-24	54	64.3
25-30	7	8.3
Over 30	2	2.4
Study level		
First year	74	88.1
Second year	8	9.5
Third year	1	1.2
Fourth year	1	1.2
Father's education		
Basic Education	49	58.3
College	24	28.6
Graduate Studies	11	13.1
Mother's education		
Basic Education	37	44.0
College	39	46.4
Graduate Studies	8	9.5

A mixed-design ANOVA was used, with time (pre/post-test) as the within-subjects factor and intervention group as the between-subjects factor. Preliminary tests confirmed normality and no significant outliers, supporting the appropriateness of this analysis.

To address the first research question, a mixed-design ANOVA was conducted to examine the effects of Time (pre-test vs. post-test) and Condition (guided vs. no guided reading) on students' critical reading skills. From the analysis presented, the within-subjects effect of time of intervention (see Table 4) was statistically significant, F (1, 82) = 196.39, p < .001, with a large effect size (partial η^2 = .705). This indicates that participants' critical reading scores significantly improved from pre-test to post-test, regardless of group condition. However, the Time × Condition interaction was not statistically significant, F (1, 82) = 0.032, p = .859, partial η^2 < .001, suggesting that the degree of improvement in

critical reading skills over time did not differ between students who received guided reading and those who did not.

The between-subjects effect (represented by the type of intervention) was also not statistically significant, F (1, 82) = 1.93, p = .168, partial η^2 = .023, indicating that there was no overall difference in critical reading scores between the guided and no guided reading groups when averaging across time points. However, descriptive statistics (Table 5) show that the guided reading group had slightly higher mean scores both at pre-test (M = 2.30, SD = 1.38) and post-test (M = 4.76, SD = 0.60) compared to the no guided reading group (Pre: M = 2.08, SD = 1.60; Post: M = 4.47, SD = 0.76). Despite this trend, the lack of significant interaction or condition effects indicates that these differences were not statistically meaningful in the current sample.

Table 4

Effect of Guided Reading Strategy on Critical Reading across Pre- and Post-Test

Source		Type III Sum of Squares	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	244.873	196.385	.000	.705
	Greenhouse-Geisser	244.873	196.385	.000	.705
	Huynh-Feldt	244.873	196.385	.000	.705
	Lower-bound	244.873	196.385	.000	.705
Time Condition	Sphericity Assumed	.040	.032	.859	.000
	Greenhouse-Geisser	.040	.032	.859	.000
	Huynh-Feldt	.040	.032	.859	.000
	Lower-bound	.040	.032	.859	.000
Condition		2.734	1.931	.168	.023

Table 5

Critical Reading Pre- and Post-Test Scores by Condition

	Condition	Mean	Std. Deviation	N
CR_Pre	No guided reading	2.079	1.6004	38
	Guided reading	2.304	1.3803	46
CR_Post	No guided reading	4.474	.7618	38
	Guided reading	4.761	.6031	46

To answer the second research question, a mixed-design ANOVA was conducted to evaluate the effects of Time (pre-test vs. post-test) and Condition (reading bot vs. no reading bot) on students' comprehension of medical texts, as measured by scores on 'skimming and scanning', 'vocabulary acquisition', and 'inference and deduction'. The within-subjects effect of Time (see Table 6) was statistically significant, F (1, 82) = 298.53, p < .001, with a very large effect size (partial η^2 = .785). This indicates that overall, students' comprehension of medical texts significantly improved from pre-test to post-test, regardless of whether they used a reading bot or not. The interaction of Time and Condition was not statistically significant, F (1, 82) = 2.82, p = .097, partial η^2 = .033. Although this interaction approached significance, it does not provide strong statistical

evidence that the amount of improvement over time differed significantly between the two groups.

However, the between-subjects effect was significant, $F(1, 82) = 6.44$, $p = .047$, partial $\eta^2 = .050$, indicating that students using the reading bot achieved higher overall comprehension scores. Both groups had similar pre-test scores (No Bot: M = 5.71, SD = 2.75; Reading Bot: M = 5.76, SD = 2.45), but by post-test the reading bot group outperformed the no-bot group (M = 15.04, SD = 1.97 vs. M = 12.53, SD = 3.81), showing greater gains following the intervention. In summary, while both groups improved significantly over time, using reading AI chatbots led to higher overall comprehension scores, suggesting that reading bots may enhance students' understanding of medical texts.

Table 6

Mixed ANOVA: The Effects of Time and Instructional Condition on Comprehension

Source		Type III Sum of Squares	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	2371.886	298.525	.000	.785
	Greenhouse-Geisser	2371.886	298.525	.000	.785
	Huynh-Feldt	2371.886	298.525	.000	.785
	Lower-bound	2371.886	298.525	.000	.785
Time Condition	Sphericity Assumed	22.386	2.818	.097	.033
	Greenhouse-Geisser	22.386	2.818	.097	.033
	Huynh-Feldt	22.386	2.818	.097	.033
	Lower-bound	22.386	2.818	.097	.033
Condition		310.054	4.436	.047	.050

Table 7

Reading Comprehension Pre- and Post-Test Scores by Condition

	Condition	Mean	Std. Deviation	N
Comprehension_Pre	No guided reading	5.711	2.7500	38
	Guided reading	5.761	2.4512	46
Comprehension_Post	No guided reading	12.526	3.8113	38
	Guided reading	15.043	1.9659	46

Since the difference between participants exposed to the reading AI chatbot and those not exposed was significant, an additional analysis was made to determine which aspect of reading comprehension was statistically significant. The statistically significant (significant) element is scores in "skimming and scanning" ($F[1, 82] = 12.810$, $p < .001$). But the difference was moderate as the partial eta-squared value was 0.35. A comparison of the mean scores reveals that, although there is no significant difference in pre-test scores among the groups, the post test scores vary with the participants exposed to the guided reading strategy form of using reading bot (M = 4.85, SD = 0.41) performing better at the skills of skimming and scanning compared to the rest of the participants who were not exposed to any reading strategy (M = 2.89, SD = 0.97).

Table 8

Mixed ANOVA: Effects of Time and Condition on Skimming and Scanning Scores

Source		Type III Sum of Squares	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	236.968	204.947	.000	.714
	Greenhouse-Geisser	236.968	204.947	.000	.714
	Huynh-Feldt	236.968	204.947	.000	.714
	Lower-bound	236.968	204.947	.000	.714
Time Condition	Sphericity Assumed	.206	.178	.674	.002
	Greenhouse-Geisser	.206	.178	.674	.002
	Huynh-Feldt	.206	.178	.674	.002
	Lower-bound	.206	.178	.674	.002
Condition		128.096	12.810	.000	.355

Table 9

Skimming and Scanning Pre- and Post-Test Scores by Condition

	Condition	Mean	Std. Deviation	N
SkimPre	No guided reading	2.379	1.2815	38
	Guided reading	2.391	1.4526	46
SkimPost	No guided reading	2.895	.9737	38
	Guided reading	4.848	.4199	46

Table 10

Mixed ANOVA: Effects of Time and Condition on Vocabulary Acquisition Scores

Source		Type III Sum of Squares	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	315.268	289.736	.000	.779
	Greenhouse-Geisser	315.268	289.736	.000	.779
	Huynh-Feldt	315.268	289.736	.000	.779
	Lower-bound	315.268	289.736	.000	.779
Time Condition	Sphericity Assumed	2.125	1.953	.166	.023
	Greenhouse-Geisser	2.125	1.953	.166	.023
	Huynh-Feldt	2.125	1.953	.166	.023
	Huynh-Feldt	89.226			
	Lower-bound	89.226			
Condition		.150	0.126	.724	.002

Table 11

Mixed ANOVA: Effects of Time and Condition on Inference and Deduction Scores

Source		Type III Sum of Squares	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	364.779	330.951	.000	.801
	Greenhouse-Geisser	364.779	330.951	.000	.801
	Huynh-Feldt	364.779	330.951	.000	.801
	Lower-bound	364.779	330.951	.000	.801
Time Condition	Sphericity Assumed	.065	.059	.809	.001
	Greenhouse-Geisser	.065	.059	.809	.001
	Huynh-Feldt	.065	.059	.809	.001
	Lower-bound	.065	.059	.809	.001
Condition		.168	.146	.703	.002

Qualitative Results

This section presents a thematic analysis of participants' interview responses on their perceptions of guided reading with AI chatbots. This analysis seeks to identify recurring patterns and attitudes in learners' use of AI chatbots during reading. Additionally, it examines how these tools may influence comprehension, motivation, anxiety, and cognitive load. Participants' responses were inductively coded and categorized into key themes. In addition, illustrative quotes from learners' responses were selected to support the findings. The letter "P" that follows these quotes refers to the word participant.

A total of 27 participants responded to the interview questions via a Google Form link. All participants confirmed utilizing the AI chatbots such as ChatGPT, Gemini, and DeepSeek. Regarding the reasons for using these AI chatbots in reading activities, participants answered a checkbox question. 20 participants reported using AI chatbots to define new vocabulary. 12 participants confirmed using them to simplify the text for better comprehension, and the same number approved using them to summarize the text. 7 participants stated that they use AI chatbots to understand hidden messages or inferences (e.g., the author's intention), and one participant added that he uses AI chatbots to find more references on the topics he reads.

As regards the difficulties participants faced while using AI chatbots, out of 27 participants, 16 participants constituting 59.3 % confirmed that they did not face any difficulties whereas 2 participants representing 7.4 % provided a contrasting perspective and 9 (33.3%) selected "maybe". These reasons included: unclear answers and inaccuracy of information or sources (see Extracts Group 1), which confirms the findings of Dahlkemper et al. (2023).

Extracts Group 1

- "Not clear answers." (P 1)
- "Not clear answers." (P 4)
- "Sometimes not accurate." (P 12)
- "Sometimes the answers aren't specified enough to the major point I'm asking for." (P 20)

Since motivation is considered a central concept in language learning, it is essential to investigate it within the context of the study and in relation to the research aims. 15 out of 27 participants (55.6%) confirmed that the AI chatbot motivated them to read more medical texts, whereas only 4 stated the opposite. These AI chatbots motivated them because they were easy to understand, provided references and sources, and offered summaries of the medical topics (see Extracts Group 2).

Extracts Group 2

- "Sometimes AI chatbots put references, therefore I usually read them." (P 2)
- "It breaks down the topic to make it easier to understand." (P 14)

- “Because it helps me to be more confident and keep going, as AI chats have quick clarifications, can simplify complex words and summarize the article.” (P 22)
- “Made it easy to understand some diagnosis by giving me the most important points about it”. (P 24)

Anxiety is another concept that warrants exploration when introducing any recent strategies in language learning. 14 participants representing 51.9 % selected “yes” for the question of “Did these AI chatbots reduce your reading anxiety?” whereas five stated the opposite. The reasons included enhancing self-confidence, a non-judgmental environment, availability at all times, and AI chatbots' simple explanations (see Extracts Group 3).

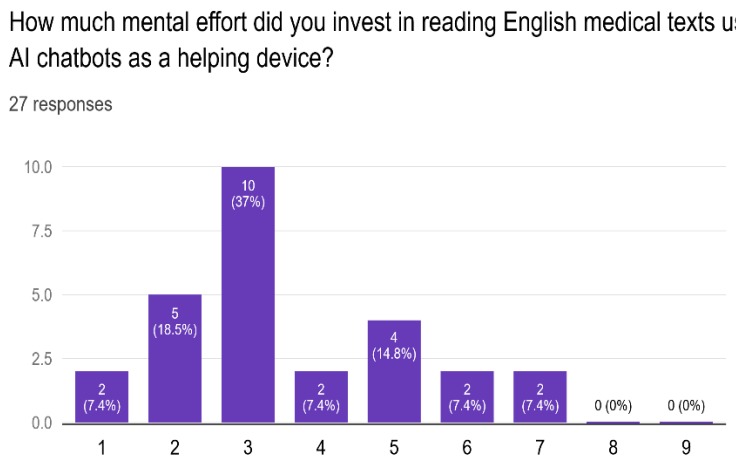
Extracts Group 3

- “Made me feel comfort and sure that if I faced any difficulties, I would find an appropriate answer easily.” (P 7)
- “You Can Ask “Stupid” Questions Without Judgment. Many students are afraid to ask basic questions in class. With me, you can ask anything, as many times as you want.” (P 16)
- “They explained hard words and ideas in a simple way, so I understood better.” (P 22)

The cognitive load question has also provided an answer to whether integrating AI chatbots would reduce the cognitive effort learners exert when reading medical texts. This is because of the complex nature of these texts, which include medical terminology, complex grammar, and long descriptions. This was measured through a one-question scale developed by Paas (1992). As illustrated in Figure 2, 10 out of 27 participants, constituting 37% selected a mental effort of three, followed by 5 (18.5%) selecting a mental effort of 2, and 4 (14.8%) selecting a mental effort of 5. In addition, 2 participants selected a cognitive effort of 1, and the same number selected mental effort of 4, 6, and 7.

Figure 2

Distribution of Cognitive Load Scale Frequencies



In conclusion, the results imply that AI chatbots would be a helpful addition to the traditional guided reading process, as they would lead to greater learning incentives, lower anxiety levels (Yuan, 2025), and reduced cognitive load among nursing students. However, successful integration should be carefully considered in terms of the instructional design applied to ensure that students employ AI chatbots to assist, not to substitute for, active work on medical texts.

Discussion

This study investigated the efficacy of chatbot-assisted guided reading in enhancing the reading comprehension and critical reading skills of EFL nursing students. The results showed significant improvement in overall reading performance in both the experimental and control groups over time, consistent with prior research highlighting the effectiveness of structured reading interventions (Brown et al., 2016; Duke et al., 2021). However, no statistically significant improvement was found in critical reading skills attributable to the addition of the chatbot compared to traditional instructor-guided reading. This finding reflects broader concerns in the literature regarding the limitations of current AI tools in supporting higher-order cognitive processes such as critical thinking and evaluative reasoning (Ali et al., 2024; Davis, 2024)**.** These findings must be further elaborated analytically on the processes and causes behind them, with explanations in reference to the conceptual framework the study uses, a combination of cognitive load theory and sociocultural scaffolding methods.

The mentioned time effect is a good argument in favor of one of the key principles of the organized guided reading as an effective pedagogical intervention, as the concept of the zone of proximal development developed by Vygotsky would suggest. Internal skill acquisition is supported by systematic support (human or digital). This supports sociocultural perspectives that emphasize mediated learning and scaffolded interaction as central to language development (Le & Nguyen, 2024; Vygotsky, 1978)**.**

A significant finding, though, is that the interaction effect is not significant. This indicates that, in its current form, the chatbot was not an effective scaffold for higher-order thinking. Such a limitation is consistent with studies that highlight that AI-driven tools often lack the adaptive, dialogic scaffolding necessary for deep cognitive engagement (Molenaar, 2022; Song et al., 2024). This distinction can be explained by the differing cognitive demands of the two reading skills. Foundational skills like skimming and scanning rely on information retrieval and pattern recognition, processes at which chatbots excel. These findings corroborate evidence that AI tools are particularly effective in supporting lower-level linguistic processing and vocabulary acquisition (Mohamed et al., 2026a; Pan et al., 2024). By matching keywords, providing immediate factual responses, and streamlining interaction, the chatbot effectively reduced intrinsic cognitive load, leading to significant gains in these areas. This is in line with cognitive load

theory, which posits that reducing unnecessary processing enhances learning efficiency (Mayer, 2024; Sweller, 1988).

On the contrary, critical reading involves complex activities such as evaluation, inference, and synthesis, which impose a significant mental load. The skills required are adaptive, dialogic, and context-sensitive, and they support scaffolding that addresses a learner's tacit reasoning and provides feedback not only on the content but also on the metacognitive process of analysis. Empirical research confirms that such higher-order skills require sustained metacognitive scaffolding and interactive dialogue, which current chatbot systems struggle to provide (Elizabeth Bunga et al., 2024; Olifant, 2024). The chatbot was likely interactive, but not very deep as it lacked nuance and contingent responsiveness, which are structural characteristics of human mentorship or more complicated forms of adaptive learning. It could answer what, and hard to direct the how and why of critical analysis. This limitation echoes findings that students often perceive AI responses as useful yet insufficiently reliable or nuanced for complex academic reasoning (Dahlkemper et al., 2023). This aligns with the qualitative feedback, in which students gave the practice a high rating but noted that the chatbot is not very good at responding to complex questions, which are considered essential in nursing, where a misinterpreted message from the chatbot can have clinical consequences. Such concerns are particularly critical in nursing education, where accurate interpretation and critical judgment directly impact clinical decision-making (Mitchell, 2024).

Additionally, another variable of significance in the explication of these findings is the 3-month intervention period. Perhaps this period will suffice to acquire and practice the lower levels of comprehension, but it is unlikely to be sufficient to develop the process of critical literacy, which will take years of scaffolded exposure to more difficult texts. This temporal limitation is supported by studies indicating that critical thinking development is a longitudinal process requiring sustained instructional interventions (Basco-Prado et al., 2024; Phimphimon et al., 2024). The latter applies particularly to nursing education, where evidence-based practice is conflated with critical reading. Learning how to analyze research approaches, bias, and clinical validity cannot be acquired with a single short-lived technological tool; it takes time and professional judgment. Therefore, integrating AI tools should be viewed as complementary rather than substitutive within long-term pedagogical frameworks (Simms, 2024).

Qualitative data provides a secondary explanatory dimension in terms of the affective and motivational theories. The self-determination theory was supported, as students reported higher levels of motivation and less anxiety when using tools that foster autonomy and perceived ability to achieve goals, as self-determination theory postulates, which in turn complicates the tools that help learners engage more in the learning process. This finding aligns with recent research demonstrating that AI tools can enhance learner motivation, engagement, and reduce anxiety in language learning contexts

(Yılmaz & Aydın, 2025; Zheng, 2024). This benefit of affect is not a trivial consequence of the fact that reduced anxiety can open cognitive resources to the learning process. Nevertheless, this beneficial impact was intertwined with the experiences of cognitive load in cases when the responses provided by the chatbot were erroneous or ambiguous in the context of difficult tasks, which serves as an example of a delicate balance that technology has to maintain to prevent the appearance of extraneous load and possibly augment the latter in instances of confusion during communication. Such findings reinforce the importance of designing AI systems that minimize misinformation and cognitive overload (Kiili et al., 2024; Yuan, 2025).

The results indicate a subtle position of AI in nursing education. Chatbots are most effectively applied in a supportive capacity to strengthen vocabulary, limited understanding, and fluency, while leaving instructors to address higher-order critical dialogue. This hybrid human-AI instructional model is increasingly recommended in the literature as a balanced approach to maximize learning outcomes (Jackson et al., 2024; Molenaar, 2022). The curriculum must be organized in a progressive series of such tools, with the earlier ones preceding the latest and serving as preparatory rather than primary tools for clinical reasoning or evidence appraisal. From a practical standpoint, educators should design staged integration models where AI tools scaffold early learning while instructors facilitate critical discourse and reflective practice. More importantly, training of educators and orientation of students should promote critical AI literacy, which is the appraisal of AI production, knowledge of their weaknesses, and the application of this knowledge to professional practice based on human judgment, ethics, and experience. Developing AI literacy is essential to ensure safe and ethical application in healthcare education (AlAli & Wardat, 2024; Duran, 2024). It is not intended to replace teachers but rather to increase efficiency and involvement in the learning process, ultimately supporting the human-centered, critical nature of safe nursing care. Future research should examine longer interventions, more adaptive AI systems, and metacognitive scaffolding to better support critical reading in specialized fields such as nursing. However, the findings should be interpreted with caution due to the short intervention period and single-institution context, which may limit generalizability and the observation of long-term skill development. Additionally, the study focused only on chatbot-supported guided reading of medical texts and did not address other language skills or alternative AI scaffolding approaches.

Conclusion

This paper demonstrates that chatbot-based learning plays a two-fold role in nursing education. Large effect sizes were found in chatbot development for foundational reading, including scanning and skimming. In qualitative terms, students were found to be more motivated (55.6%) and had less anxiety (51.9%), attributing it to the availability of the chatbot, ease of explanations, and the non-judgmental attitude of the chatbot.

Nevertheless, there were no significant gains in critical reading, and some interviewees found the chatbot unclear and inaccurate. It means that AI is efficient at facilitating understanding but is not viable for supporting higher levels of analytical thinking required in nursing practice. The conclusions highlight the necessity of hybrid teaching that combines AI effectiveness with human support for working with complex tasks.

The results are highly applicable to emerging AI-based medical nursing education. Even though chatbots have the potential to facilitate basic learning, they should be integrated into scaffolded models that will assist instructors in transforming AI-facilitated knowledge acquisition into critical thinking. Future research will provide unambiguous models of integration that will preserve the motivational benefits of AI and improve clinical decision-making and higher-order skills. The research concludes that technological innovation and human experience are the best approaches to integrating technology into the teaching process, thereby facilitating the provision of the literacy skills needed by nurses in the new healthcare system.

This study has several limitations. The intervention was too limited to identify long-term effects or developmental patterns. Only medical texts were used in the research conducted within a single institution, limiting generalizability. The key mediators were explored qualitatively, including cognitive load and metacognitive involvement. To enhance measurement, valid psychometric scales should be employed in future studies.
In order to present additional research, a longitudinal study will be required to track the process of skill improvement as AI is maintained. In an experimental study, the type of chatbot scaffolding most effective in supporting critical thinking could be isolated and contrasted with metacognitive prompts and corrective feedback. Notably, research is needed on hybrid pedagogical methods that strategically align chatbot-assisted foundational practice with instructor-led seminars grounded in deep critical inquiry and clinical practice.

ORCID
https://orcid.org/0009-0008-5598-9043
https://orcid.org/0000-0002-9627-1774
https://orcid.org/0009-0007-2368-4651
https://orcid.org/0009-0004-4045-8195
https://orcid.org/0009-0005-9900-1437
https://orcid.org/0000-0003-3891-963X

Acknowledgements
The authors would like to express their sincere appreciation to the nursing students for their time and cooperation.

Funding Statement
This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2602).

CRediT Authorship Contribution Statement
Nada S. Al Ruwaili: Conceptualization, Methodology, Investigation, Data Curation, Writing – Original Draft, Project Administration
Abdul Aziz Mohamed Mohamed Ali El Deen: Writing – Review & Editing, Validation, Formal Analysis, Visualization, Revision Management
Shahad Z. Aldandan: Investigation, Resources, Data Curation, Writing – Review & Editing
Noura A. Alhudaib: Investigation, Resources, Data Curation
Mohammed A. Alrashed: Methodology, Formal Analysis, Supervision, Writing – Review & Editing
Amr M. Mohamed: Conceptualization, Methodology, Supervision, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Project Administration

Generative AI Use Disclosure Statement
The authors confirm that no Gen-AI applications were used in the conception, design, analysis, or writing of this manuscript. However, Grammarly was employed solely for language editing and readability.

Ethics Declarations
World Medical Association (WMA) Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Participants
Ethical approval was granted by North Private College of Nursing, in line with the Declaration of Helsinki. All participants gave written informed consent and were assured they could withdraw at any time without penalty.

Competing Interests
The authors declare that they have no competing interests.

Data Availability
The datasets are available from the corresponding author upon reasonable request.

References
AlAli, R., & Wardat, Y. (2024). Enhancing classroom learning: ChatGPT's integration and educational challenges. *International Journal of Religion*, *5*(6), 971–985. https://doi.org/10.61707/znwnxd43

Alfaleh, M., Albasis, H. H., & Mohamed, A. M. (2025). EFL learners' motivation and engagement with oral reading activities and AI tools: Exploring the effect of gender, study level and English language proficiency. *Cogent Education*, *12*(1), Article 2542395. https://doi.org/10.1080/2331186X.2025.2542395

Ali El Deen, A. M. M., Emran, A. Q., Alenazi, M. E., Mohamed, A. M., & Ibrahim, A. M. (2026). Leveraging generative artificial intelligence (GenAI) to transform EFL instructional practices in the context of 21st-century skills. *Scientific Culture, 12*, 3015–3027. https://doi.org/10.5281/zenodo.18840865

Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences*, *14*(6), 643. https://doi.org/10.3390/educsci14060643

AlTwijri, L., & Abdelhalim, S. M. (2026). Enhancing reading engagement in adult EFL classrooms through AI-supported instruction: A mixed-methods study. *System*, *136*, Article 103906. https://doi.org/10.1016/j.system.2025.103906

Barrot, J. S. (2024). Leveraging ChatGPT in the writing classrooms: Theoretical and practical insights. *Language Teaching Research Quarterly, 43*, 43–53. https://doi.org/10.32038/ltrq.2024.43.03

Basco-Prado, L., Biurrun-Garrido, A., Zuriguel-Pérez, E., Roldán-Merino, J., & Mestres-Soler, O. (2024). Critical thinking level in nursing degree students according to sociodemographic and academic variables: a correlational study. *Nurse Education in Practice*, *78*, Article 103955. https://doi.org/10.1016/j.nepr.2024.103955

Baskara, F. R., Puri, A. D., & Mbato, C. L. (2024). Exploring the use of generative AI in student-produced EFL podcasts: A qualitative study. *Language Teaching Research Quarterly*, *43*, 81–101. https://doi.org/10.32038/ltrq.2024.43.05

Brown, C. A., Danvers, K., & Doran, D. T. (2016). Student perceptions on using guided reading questions to motivate student reading in the flipped classroom. *Accounting Education*, *25*(3), 256–271. https://doi.org/10.1080/09639284.2016.1165124

Cheng, X., Yin, L., Lin, C., Shi, Z., Zheng, H., Zhu, L., … Dong, R. (2024). Chatbot dialogic reading boosts comprehension for Chinese kindergarteners with higher language skills. *Journal of Experimental Child Psychology, 240*, Article 105842. https://doi.org/10.1016/j.jecp.2023.105842

Dahlkemper, M. N., Lahme, S. Z., & Klein, P. (2023). How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT. *Physical Review Physics Education Research, 19*(1), Article 010142. https://doi.org/10.1103/PhysRevPhysEducRes.19.010142

Davis, A. J. (2024). AI rising in higher education: opportunities, risks and limitations. *Asian Education and Development Studies*, *13*(4), 307–319. https://doi.org/10.1108/AEDS-01-2024-0017

Dong, L. (2024). ChatGPT in language writing education: Reflections and a research agenda for a ChatGPT feedback engagement framework. *Language Teaching Research Quarterly, 43*, 121–131. https://doi.org/10.32038/ltrq.2024.43.07

Duke, N. K., Ward, A. E., & Pearson, P. D. (2021). The science of reading comprehension instruction. *The Reading Teacher*, *74*(6), 663–672. https://doi.org/10.1002/trtr.1993

Duran, V. (2024). Analyzing teacher candidates' arguments on AI integration in education via different chatbots. *Digital Education Review, 45*, 68–83. https://doi.org/10.1344/der.2024.45.68-83

Elizabeth Bunga, D. U., Nurkamto, J., Suryani, N., & Gunarhadi, G. (2024). Enhancing critical reading through metacognitive scaffolding in a flipped classroom. *World Journal of English Language, 14*(6), 297. https://doi.org/10.5430/wjel.v14n6p297

Goktas, P., Kucukkaya, A., & Karacay, P. (2024). Utilizing GPT 4.0 with prompt learning in nursing education: A case study approach based on Benner's theory. *Teaching and Learning in Nursing*, *19*(2), e358–e367. https://doi.org/10.1016/j.teln.2023.12.014

Griche, M. S., & Bennis, M. (2026). Generative artificial intelligence integration in foreign language education in higher education: A systematic literature review. *Technology in Language Teaching & Learning*, *8*, Article 102827. https://doi.org/10.29140/tltl.2026.102827

Huang, H. L., Hwang, G. J., Chen, P. Y., & Hwang, G. H. (2026). Effects of graphical summarization and AI-based chatbots on learners' reading comprehension achievement, motivation, self-efficacy, critical thinking and cognitive load. *Educational Technology Research and Development*, 1–27. Advance online publication. https://doi.org/10.1007/s11423-025-10583-x

Jackson, I., Ivanov, D., Dolgui, A., & Namdar, J. (2024). Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation. *International Journal of Production Research*, *62*(17), 6120–6145. https://doi.org/10.1080/00207543.2024.2309309

Jmaiel, H. A., Abukhait, R. O., Mohamed, A. M., Shaaban, T. S., Al-khresheh, M. H., & Al-Qadri, A. H. (2025). The role of ChatGPT in enhancing EFL students' ESP writing skills: an experimental study of gender and major differences. *Discover Education, 4*(1), 1–19. https://doi.org/10.1007/s44217-025-00700-6

Kiili, K., Siuko, J., & Ninaus, M. (2024). Tackling misinformation with games: a systematic literature review. *Interactive Learning Environments, 32*(10), 7086–7101. https://doi.org/10.1080/10494820.2023.2299999

Kitsili, N. A., & Murray, S. R. (2024). The challenge of incorporating new methods: The case of group guided reading in South Africa. *South African Journal of Childhood Education*, *14*(1), 1487. https://hdl.handle.net/10520/ejc-sajce_v14_n1_a1487

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Kurzeja, O., Flynn, S., Grindle, C. F., Sutherland, D., & Hastings, R. P. (2024). Teaching reading skills to individuals with autism and/or intellectual disabilities using computer-assisted instruction: A systematic review. *Review Journal of Autism and Developmental Disorders*. Advance online publication. https://doi.org/10.1007/s40489-024-00465-y

Le, H. V., & Nguyen, L. Q. (2024). Promoting L2 learners' critical thinking skills: The role of social constructivism in reading classes. *Frontiers in Education, 9*, Article 1241973. https://doi.org/10.3389/feduc.2024.1241973

Le, H. V., Nguyen, T. A. D., Le, D. H. N., Nguyen, P. U., & Nguyen, T. T. A. (2024). Unveiling critical reading strategies and challenges: a mixed-methods study among English major students in a Vietnamese higher education institution. *Cogent Education*, *11*(1), Article 2326732. https://doi.org/10.1080/2331186X.2024.2326732

Lee, S., & Bolt, D. M. (2018). Asymmetric item characteristic curves and item complexity: Insights from simulation and real data analyses. *Psychometrika, 83*(2), 453–475. https://doi.org/10.1007/s11336-017-9586-5

Liu, C. C., Chen, W. J., Lo, F. Y., Chang, C. H., & Lin, H. M. (2024). Teachable Q&A agent: The effect of Chatbot training by students on reading interest and engagement. *Journal of Educational Computing Research*, *62*(4), 1122–1154. https://doi.org/10.1177/07356331241236467

Liu, C. C., Liao, M. G., Chang, C. H., & Lin, H. M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, *189*, Article 104576. https://doi.org/10.1016/j.compedu.2022.104576

Makhlouf, E., Alenezi, A., & Shokr, E. A. (2024). Effectiveness of designing a knowledge-based artificial intelligence chatbot system into a nursing training program: a quasi-experimental design. *Nurse Education Today*, *137*, Article 106159. https://doi.org/10.1016/j.nedt.2024.106159

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, *36*(1), 8. https://doi.org/10.1007/s10648-023-09842-1

Mitchell, K. M. (2024). Does reading ability predict student success? A scoping review. *Nurse Education Today*, *136*, Article 106150. https://doi.org/10.1016/j.nedt.2024.106150

Mohamed, A. M., Alenezi, N. A., Darrag, Y., Al Ruwaili, N. S. A., Houfan, R. A., & Berdida, D. J. E. (2026a). The role of AI chatbots in nursing students' autonomous learning in mastering medical vocabulary: A quasi-experimental study. *Teaching and Learning in Nursing*, 21(1), e76–e85. https://doi.org/10.1016/j.teln.2025.08.007

Mohamed, A. M., Emran, A. Q. E., Abdelrady, A. H., & Ali El Deen, A. M. M. (2026b). Giftedness and academic motivation in GenAI contexts: The moderating and mediating role of gender and AI anxiety. *Frontiers in Education, 11*, Article 1762707. https://doi.org/10.3389/feduc.2026.1762707

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Montreuil, J., Lacasse, M., Audétat, M. C., Boileau, É., Laferrière, M. C., Lafleur, A., ... & Steinert, Y. (2025). Interventions for undergraduate and postgraduate medical learners with academic difficulties: A BEME systematic review update: BEME Guide No. 85. *Medical Teacher, 47*(1), 18–36. https://doi.org/10.1080/0142159X.2024.2331041

Nghia, N., & Yen, H. (2018). ESP nursing students' problems in reading medical academic texts and their proposed solutions. In *Proceedings of the International Conference on Language and Learning Today* (pp. 316–341).

Olifant, F. M. (2024). Breaking the cycle of poor critical reading comprehension: A strategy-based intervention. *Literator-Journal of Literary Criticism, Comparative Linguistics and Literary Studies*, *45*(1), 2080. https://hdl.handle.net/10520/ejc-literat_v45_n1_a2080

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Pan, M., Guo, K., & Lai, C. (2024). Using artificial intelligence chatbots to support English-as-a-foreign-language students' self-regulated reading. *RELC Journal, 56*(3), 781–793. https://doi.org/10.1177/00336882241264030

Phimphimon, N., Intasena, A., Srimunta, T., & Khantasiri, P. (2024). Improving critical reading abilities in 10th graders: An active learning approach. *International Journal of Learning, Teaching and Educational Research, 23*(3), 186–198. https://doi.org/10.26803/ijlter.23.3.12

Rosenbaum, P. R. (1987). Comparing item characteristic curves. *Psychometrika, 52*(2), 217–233. https://doi.org/10.1007/BF02294236

Shafiee Rad, H. (2025). Reinforcing L2 reading comprehension through artificial intelligence intervention: refining engagement to foster self-regulated learning. *Smart Learning Environments*, *12*(1), 23. https://doi.org/10.1186/s40561-025-00377-2

Simms, R. C. (2024). Work with ChatGPT, not against: 3 teaching strategies that harness the power of artificial intelligence. *Nurse Educator*, *49*(3), 158–161. https://doi.org/10.1097/NNE.0000000000001634

Song, Y., Weisberg, L. R., Zhang, S., Tian, X., Boyer, K. E., & Israel, M. (2024). A framework for inclusive AI learning design for diverse learners. *Computers and Education: Artificial Intelligence*, Article 100212. https://doi.org/10.1016/j.caeai.2024.100212

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Teng, M. F., Mizumoto, A., & Takeuchi, O. (2024). Understanding growth mindset, self-regulated vocabulary learning, and vocabulary knowledge. *System*, *122*, Article 103255. https://doi.org/10.1016/j.system.2024.103255

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

Wardhono, A., Asari, S., Kalista, A., Dumiyati, D., & Anwar, K. (2024). Reading automated feedback (RAF) in English for nursing: Augmenting perceptions and reading ability. *International Journal of Learning, Teaching and Educational Research, 23*(10), 544–563. https://doi.org/10.26803/ijlter.23.10.26

Xerri, D., Kamali, J., & Mohebbi, H. (2025). Technology-enhanced language teacher education: Opportunities, challenges, and futures. *Technology in Language Teaching & Learning*, *7*(2), Article 103213. https://doi.org/10.29140/tltl.v7n2.103213

Yetişensoy, O., & Karaduman, H. (2024). The effect of AI-powered chatbots in social studies education. *Education and Information Technologies*, *29*(13), 17035–17069. https://doi.org/10.1007/s10639-024-12485-6

Yılmaz, Ö. K., & Aydın, S. (2025). The impact of the use of artificial intelligence–generated materials on reading motivation among EFL learners. *Reading Research Quarterly*, *60*(3), Article e70016. https://doi.org/10.1002/rrq.70016

Yuan, H. (2025). Artificial intelligence in language learning: biometric feedback and adaptive reading for improved comprehension and reduced anxiety. *Humanities and Social Sciences Communications*, *12*(1), 1–16. https://doi.org/10.1057/s41599-025-04878-w

Zheng, S. (2024). The effects of chatbot use on foreign language reading anxiety and reading performance among Chinese secondary school students. *Computers and Education: Artificial Intelligence*, *7*, Article 100271. https://doi.org/10.1016/j.caeai.2024.100271

Appendix A

Figure A1

ICC Plot for Items 17 to 20