Castledown

OPEN ACCESS

Technology in Language Teaching & Learning
ISSN 2652-1687
https://www.castledown.com/journals/tltl/

Technology in Language Teaching & Learning, 8, 103893 (2026)
https://doi.org/10.29140/tltl.2026.103893

The Effect of Using Bearly AI on Developing Reading Comprehension Skills among Female Tenth-grade EFL Students

SABA ALAMOUSH[a] (Corresponding author)
MAHMOUD SULAIMAN BANI ABDELRAHMAN[b]

[a]*Department of Curriculum and Instruction, Al-Hussein Bin Talal University, Jordan*
sabaomosh@gmail.com

[b]*Department of Curriculum and Instruction, Al-Hussein Bin Talal University, Jordan*
Mahmoud.Sulaiman@ahu.edu.jo

Abstract

This study examined the association between the use of the artificial intelligence (AI) platform Bearly AI and the development of reading comprehension skills among female tenth-grade EFL students in Jordan. The sample comprised 54 female tenth-grade students, and a quasi-experimental design was employed, with an experimental group receiving AI-supported reading instruction and a control group following conventional textbook-based instruction. Validated pre- and post-tests were administered to assess changes in students' reading comprehension. The data were analyzed using descriptive statistics and inferential procedures, including ANCOVA and MANCOVA. The results showed that students who engaged in AI-supported reading activities demonstrated higher post-test performance in overall reading comprehension and across five reading sub-skills compared with those who received conventional instruction. The findings suggest that integrating AI-based tools such as Bearly AI into EFL reading instruction is associated with improved reading outcomes and may support more differentiated, data-informed instructional practices in secondary school contexts.

Keywords: AI in education; artificial intelligence, EFL learners; English language skills; Jordan; language learning technology

Introduction

Language learning is among the most important domains of education, as language serves not only as a means of communication but also as a fundamental enabler of learning, cultural exchange, and participation in a globalized world (Rao, 2019). To become proficient in English, learners must develop competence in its four core skills: listening, speaking, reading, and writing. Of these, reading has a central role because it facilitates the acquisition of linguistic knowledge and broader comprehension skills (Sucompa, 1998).

Many learners of English as a foreign language (EFL) struggle with reading comprehension in conventionally taught classes (Al-salameen & Abdelrahman, 2023). EFL learners encounter numerous challenges within conventional teaching frameworks, which neither account for differences among individual students nor address varying levels of reading comprehension within a group (Fareh, 2010). This can raise psychological and emotional barriers for individual students, including the anxiety and embarrassment that may arise from making mistakes in front of others; such barriers can hinder participation and students' self-confidence (Saadi & Saadat, 2015).

Another key factor influencing comprehension is prior knowledge. Students with stronger background knowledge tend to perform better on reading tasks than those with limited prior experience (Abdelaal & Sase, 2014). A lack of grammar, vocabulary, and practical skills makes learning even more difficult, particularly in conventional classroom settings (Etaat, 2024). This highlights the need to provide learners with supplementary resources to bridge knowledge gaps, as textbooks alone often fail to meet the diverse learning needs of students. Electronic and digital resources can help fill this gap by offering flexible, accessible, and tailored learning opportunities (Saadi & Saadat, 2015). In addition, despite its importance, teachers often find that providing individual corrective feedback in conventional ways takes up a considerable amount of lesson time, especially in large classes with low proficiency levels under a teacher-centered approach (Ferris, 2007). These challenges illustrate the necessity for stimulating and dynamic learning environments that can adapt to the difficulties faced by EFL students. In recent decades, the integration of technology into education has significantly transformed teaching and learning worldwide (De Houwer et al., 2013). The use of technology has enhanced accessibility, interaction, and effectiveness in education, allowing teachers and students to engage in dynamic and personalized learning experiences (Haleem et al., 2022).

Recent studies have emphasized the growing potential of artificial intelligence (AI) in teaching English (Chan, 2025; Rahman & Watanobe, 2023; Tlili et al., 2023). According to Fernandez-Rio (Fernández-Río, 2016), curriculum, teachers, students, and a suitable learning environment constitute interrelated elements in education. AI can help manage these elements. By integrating AI into the classroom, educators can meet the varied requirements of their students, make better use of their time, and create an atmosphere conducive to learning (Griche & Bennis, 2026; Hamad et al., 2022; Hwang et al., 2021).

Although reading comprehension is a core skill and fundamental to learning English, many EFL students—particularly at the secondary school level—face persistent challenges in this area. Tenth-grade students encounter additional barriers due to the transitional nature of this stage in their educational journey. Given the rapid advancement of educational technology, it has become both timely and necessary to leverage modern tools to enhance English language instruction in general and reading comprehension instruction in particular. Nevertheless, despite the increasing prevalence of AI applications in education, a significant dearth of empirical studies exists concerning their efficacy in teaching reading comprehension in EFL contexts. In particular, insufficient attention has been directed toward the impact of AI tools such as Bearly AI on enhancing reading comprehension skills among tenth-grade

EFL students. Understanding the academic outcomes of AI integration is crucial for guiding future teaching methods and enhancing student engagement and success in EFL contexts.

Bearly AI is a free, AI-powered reading support platform designed to assist learners in text processing through adaptive prompts, vocabulary support, summarization, and comprehension guidance. Compared with other freely available AI platforms, Bearly AI offers a focused educational interface that supports differentiated instruction, real-time feedback, and learner-initiated inquiry, making it particularly suitable for classroom-based EFL reading comprehension instruction. In addition, it allows controlled, text-based interaction aligned with curricular reading objectives, without requiring advanced technical skills or paid subscriptions.

The present study aimed to investigate the effect of using the AI platform Bearly AI on the development of reading comprehension skills among female tenth-grade EFL students. Specifically, it sought to identify any statistically significant differences between the mean reading comprehension scores of the control group (conventionally taught) and the experimental group (receiving AI-integrated instruction) that could be attributed to the teaching method.

Theoretical Framework

The present study is primarily grounded in constructivist learning theory (Vygotsky & Cole, 1978) and differentiated instruction (Tomlinson, 2001), which informed the design of the AI-supported reading activities and the structuring of the pre-reading, during-reading, and post-reading phases. From a constructivist perspective, reading comprehension is viewed as an active process of meaning construction through interaction with texts, guided support, and iterative reflection. Differentiated instruction informed the proficiency-based grouping of learners and the adaptation of instructional tasks to varying levels of readiness. Krashen's Affective Filter Hypothesis and Input Hypothesis (Krashen, 1982) were used as interpretive lenses to contextualize the potential affective and cognitive affordances of AI-supported learning environments (e.g., private interaction and individualized pacing). However, affective variables such as anxiety and motivation were not directly measured in the present study and, therefore, are not treated as empirically tested mechanisms. Similarly, the Technological Pedagogical Content Knowledge (TPACK) framework (Koehler et al., 2014) is referenced to situate the pedagogical integration of AI within technology-enhanced language teaching. Teacher TPACK development was not measured and is therefore discussed as a contextual pedagogical consideration rather than an outcome variable.

Materials and Methods

Study Design and Setting

The study followed a quasi-experimental design. In educational settings where random assignment is impractical, this approach can be used to investigate how an independent variable affects dependent variables.

The study was carried out in a public school governed by the Zarqa Directorate of Education in Jordan during the second semester of the 2024/25 school year. The study was conducted in tenth-grade classrooms, which offered an official, structured environment and complied with the national curriculum.

Sampling and Participants

Purposive sampling was used to select tenth-grade participants from a secondary school for girls because the students shared similar educational and linguistic characteristics, making the sample suitable for a quasi-experimental design and ensuring comparable instructional contexts. The school comprised three intact tenth-grade classes (Classes A, B, and C), each originally enrolling 27 students. To ensure instructional consistency and avoid cross-group contamination, intact classes were retained and not mixed across conditions. Following random assignment at the class level, Class A (n = 27) was designated as the experimental group and received the Bearly AI-based intervention, whereas Class B (n = 27) served as the control group and received conventional textbook-based instruction. Class C (n = 27) was used exclusively for piloting the research instruments and did not take part in the intervention. All instructional procedures were implemented in the students' regular classrooms, and no class was divided or merged during the study.

The participants were female students aged approximately 15–16 years, all of whom had studied English as a foreign language for about ten years. The classes shared similar educational, linguistic, and curricular backgrounds, which supported the internal consistency of the quasi-experimental design. Participation was voluntary and was conducted with parental consent.

Teaching Methods

The experimental group received AI-assisted instruction using Bearly AI, whereas the control group received conventional textbook-based instruction. Both groups were taught following the same instructional phases (pre-reading, during-reading, and post-reading) and learning objectives, grounded in differentiated instruction theory (Tomlinson, 2001). The two conditions differed only in the mode of instructional support (AI-mediated vs. conventional non-AI), while time-on-task, curricular content, lesson sequencing, and assessment procedures were held constant across groups. Students in both groups were classified into three proficiency levels for the purpose of differentiated instruction: beginner, intermediate, and advanced.

Instruction was delivered by two regular English language teachers at the participating school. Each instructor taught one intact class: one teacher was responsible for the experimental group, while the other taught the control group. The instructors were not members of the research team and were not involved in the study design, data collection, or data analysis. Their role was limited to implementing the assigned instructional procedures as specified in the teacher's guide, which helped ensure instructional consistency while minimizing potential researcher bias. Both instructors followed identical lesson objectives, time-on-task, materials, and task sequencing as specified in the standardized teacher guide to reduce instructor-related variability. The instructor of the experimental group received targeted training on the pedagogical use of Bearly AI, and a structured guidebook with lesson plans and AI-supported tasks was developed to standardize implementation.

The intervention was implemented through a systematically organized sequence to ensure methodological rigor and instructional consistency. Bearly AI was selected due to its accessibility, ease of integration into school computer laboratories, and compatibility with structured instructional prompts. The school computer laboratory was equipped with the free version of Bearly AI, and individual email accounts were created for each participant in the experimental group. Reading texts from the tenth-grade *Action Pack 10* curriculum (Johnson, 2014) and proficiency-based instructional prompts were prepared to match the three proficiency levels. All instructional materials and procedures were reviewed and validated by a panel of university professors, educational supervisors, and experienced EFL teachers. The intervention lasted one and a half months, after which the same post-test was administered to both groups.

As illustrated in Figure 1, the experimental condition adopted a structured teacher–student–AI interaction model. Following proficiency-based classification, students engaged in AI-supported reading sessions aligned with the first module of *Action Pack 10*. In the pre-reading phase, Bearly AI supported the activation of prior knowledge through background information retrieval. During reading, the AI provided adaptive vocabulary support, contextualized examples, synonyms and antonyms, and guidance on identifying main and supporting ideas in response to student- and teacher-initiated prompts. In the post-reading phase, Bearly AI generated differentiated comprehension tasks, evaluated student summaries, and provided immediate formative feedback. Throughout the intervention, the teacher monitored progress, provided pedagogical scaffolding, and ensured curricular alignment, while students engaged in iterative practice, reflection, and response revision. In-session formative monitoring was supported through AI-generated feedback and was used solely for instructional purposes (e.g., summarizing students' overall strengths and areas for improvement across sessions). For example, at the end of activities, the teacher prompted Bearly AI to review students' responses and provide feedback. These AI–student interaction outputs were used to inform immediate instructional support and scaffolding but were neither stored nor analyzed as research data.

The control group followed the same instructional phases (pre-reading, during-reading, and post-reading), learning objectives, lesson duration, and curricular content as the experimental group, using the first module of *Action Pack 10*. In the pre-reading phase, prior knowledge was activated through teacher-led questioning, picture-based discussion, and brainstorming guided by the textbook and the teacher's guide. During reading, vocabulary explanation, inference-making, and identification of main and supporting ideas were conducted through direct instruction, contextual guessing strategies, board work, and guided whole-class discussion. In the post-reading phase, students completed summaries and comprehension tasks drawn from the textbook and teacher-prepared worksheets, followed by teacher feedback and class discussion. Instructional support was provided through teacher scaffolding, monitoring, and guided practice.

Research Tool and Data Collection

A single primary tool was used in the study to gather information on the research variables. A reading comprehension test (Supplementary Material, Appendix 1) was administered as both the pre-test and post-test to ensure strict equivalence in construct representation, text difficulty, and sub-skill coverage. To reduce potential practice or memory effects, the two administrations were separated by a six-week instructional period, and test items were not disclosed, reviewed, or practiced during the intervention.

The reading comprehension test (Supplementary Material, Appendix 1) consisted of two reading passages, with 20 items per passage, yielding a total of 40 items per administration. Each item was worth one point, and the maximum possible score was 40. The test was administered over two class periods (60 minutes in total). The 40 items were distributed across five reading sub-skills: vocabulary, literal comprehension, inferential comprehension, analytical skills, and evaluation skills. Within each passage, the 20 items were evenly allocated across the five sub-skills (four items per sub-skill), resulting in a total of eight items per sub-skill across the two passages.

The test also functioned as an English proficiency test to determine the general language level of the students and to ensure group equivalence prior to the intervention. Based on the pre-test scores (maximum = 40), students in both groups were classified into three proficiency levels: beginner (1–19), intermediate (20–29), and advanced (30–40). The cut-off points were informed by score distribution and classroom performance norms commonly observed at the Grade 10 level and were used pragmatically to support differentiated instruction rather than for formal proficiency labeling.

The test was developed based on three complementary assessment frameworks: Barrett's Taxonomy, Bloom's Revised Taxonomy, and the Cognitive Classification System of the Jordanian Ministry of

Control Group Participants (N=27)	Experimental Group Participants (N=27)

Pre-test Administration
(Reading comprehension test – Barret / Bloom)

Student Proficiency Classification
Beginner / Intermediate / Advanced

Instructional phase (22 sessions–40 min)
First module of Action Pack 10 (Johnson, 2014)

Conventional Reading Comprehension Session (Control Group)	AI-Supported Reading Comprehension Session (Experimental Group)
Pre-reading phase **Activating prior knowledge** -Students activated prior knowledge through teacher-led questioning, picture-based discussions, brainstorming activities, and textbook-based prompts.	**Pre-reading phase** **Activating prior knowledge** – Students used Bearly AI to search for more information about the text topic.
During-reading phase **Vocabulary support, highlighting, inference** Vocabulary explanation and inference-making were conducted through conventional instructional practices, including direct explanation, contextual guessing strategies, board work, and guided discussion, using the textbook and the official teacher's guidebook.	**During-reading phase** **Vocabulary support, highlighting, inference** • Students highlighted difficult words and used Bearly AI to learn them through interactive and personalized learning. • Students used Bearly AI to provide vocabulary words within meaningful or real-life sentences • Students used Bearly AI to offer synonyms and antonyms.
Main and supporting ideas Identification of main and supporting ideas was facilitated through teacher explanation, guided discussion, and board-based modeling, following the textbook and the official teacher's guidebook.	**Main and supporting ideas** – Students interacted with Bearly AI on how to know the main and supporting ideas and how to differentiate between them in the text.
Post-reading phase **Summary of the text and writing like** Students completed summaries and comprehension exercises drawn from the textbook and supplementary teacher-prepared materials, followed by teacher feedback and whole-class discussion.	**Post-reading phase** **Summary of the text and writing like** Students write a summary, then Bearly AI checks the answer and corrects the mistake.
Post-reading activities Follow-up activities were implemented using conventional worksheets, oral questioning, and written exercises, with feedback provided by the teacher.	**Post-reading activities** Bearly AI creates more reading comprehension questions and activities, depending on the differentiated learning by using the prompts. Bearly AI checks the answer and corrects mistakes.

Teacher–Student Interaction Teacher: scaffolding, direct instruction, monitoring, feedback Instructional materials (textbook & guidebook): structured prompts, fixed activities, non-adaptive feedback Students: interaction, response practice, reflection.	Teacher–Student–AI Interaction Teacher: scaffolding, monitoring, guidance Bearly AI: adaptive prompts, immediate feedback, differentiation learning Students: interaction, response revision, reflection.

Teacher-led Feedback and Progress Monitoring	In-session Formative Monitoring via AI Feedback

Post-test Administration
(Same reading comprehension test for both groups)

Figure 1 *Teacher–Student Interaction Model for the Control Group (Using the Conventional Teaching Method) and the Experimental Group (Using the Bearly AI-Based Reading Intervention)*

Education. Barrett's Taxonomy provided the conceptual basis for constructing items that assess literal, inferential, and evaluative comprehension relevant to EFL reading (Krismadayanti & Zainil, 2022; Smith & Barrett, 1979). Bloom's Revised Taxonomy (Anderson & Krathwohl, 2001) guided the design of items across lower- and higher-order cognitive levels. In addition, the Jordanian Ministry of Education's cognitive classification system informed the proportional weighting of cognitive levels in accordance with national assessment standards (Supplementary Table 1).

In line with the grading criteria of the Jordanian Ministry of Education for tenth-grade students, the cognitive weighting of items followed a 50% remembering, 30% understanding, and 20% higher-order thinking distribution. Accordingly, for each sub-skill across the two passages (eight items in total), four items assessed remembering, two items assessed understanding, and two items assessed higher-order thinking skills (analyzing or evaluating). A detailed test blueprint mapping sub-skills to cognitive levels and item distribution was developed to ensure construct validity and transparency (Supplementary Table 2).

Content validity was established through expert review by the same panel of university professors, educational supervisors, and experienced teachers who evaluated the instructional program. Their feedback was incorporated into the final version of the test. Revisions included rewording an ambiguous item and refining the overall distribution of comprehension questions to improve content coverage. In addition, the Ministry of Education's success and failure criteria for tenth-grade English were adopted as standardized benchmarks for interpreting student performance.

To assess the internal consistency of the reading comprehension test, a sample of 15 students was selected from the same school but from a different class (Class C). These students were excluded from the main study sample and were not involved in the experimental procedures. Their participation was limited to piloting the study instrument. The test was administered twice with a one-week interval to assess its reliability using the test–retest method and its stability over time.

The Pearson correlation coefficients between each item score and the total scores of its corresponding level and the overall test were calculated. In addition, the corrected item–total correlations between each item score and the total score of its level were computed. The Pearson correlation coefficients between item scores and both the total score of their level and the overall test were higher than the threshold value of 0.35. The corrected item–total correlations between item scores and the total scores of their levels were higher than the threshold value of 0.40. These values indicate an acceptable degree of internal consistency, supporting the construct validity of the test.

The psychometric characteristics of the test were calculated to evaluate the difficulty and discrimination indices of the reading comprehension items by administering the test to the same pilot group of 15 students. The item discrimination indices ranged from 0.44 to 0.70, indicating a high level of discrimination (higher than 0.20) (Brown, 1983). The difficulty indices ranged from 0.42 to 0.73. According to Doran (1980), items with difficulty indices between 0.20 and 0.80 are acceptable. Consequently, the difficulty and discrimination indices of the reading comprehension test items were satisfactory.

Furthermore, Cronbach's alpha and test–retest reliability coefficients for the reading comprehension test were calculated. The test exhibited acceptable to excellent reliability across its various components. The overall Cronbach's alpha coefficient (0.91) and strong test–retest coefficients (ranging from 0.82 to 0.90) indicate that the test is both internally consistent and stable over time. This dual evidence of reliability enhances confidence in using the test for both research and educational purposes, ensuring that any inferences made about learners' reading comprehension are based on consistent and reproducible measurements.

Data were collected during the second semester of the 2024/25 academic year. Data obtained from the reading comprehension tests were used to examine the effect of the Bearly AI-based instruction on students' reading comprehension abilities. The tests were administered before and after the intervention.

Ethics and Data Sharing

The free version of Bearly AI was used in the school computer laboratory for instructional purposes only. Students interacted with the platform to complete reading comprehension tasks aligned with the prescribed curriculum. The content shared with the platform consisted exclusively of instructional reading texts from *Action Pack 10* and task-related prompts generated by the teacher and students for learning support (e.g., vocabulary clarification, comprehension questions, and summary feedback). No personally identifiable information (e.g., names, student IDs, email addresses, grades, or demographic data) was entered into the platform. Individual email accounts were created solely to enable access to the tool and were not linked to students' real names or school records. Student outputs generated during AI-supported activities were not downloaded, stored, or analyzed by the researchers outside the classroom context, and no platform usage logs were collected for research purposes. AI-supported monitoring of overall performance was used for formative instructional purposes only (e.g., summarizing strengths and weaknesses across sessions) and was not stored or analyzed as research data. All research data were derived exclusively from anonymized pre- and post-test scores administered by the researchers. The study adhered to institutional ethical guidelines, and informed consent was obtained from students' parents/guardians before participation.

Data Management and Analysis

Data were analyzed using SPSS version 21 (IBM), with the significance level set at $\alpha = 0.05$. Descriptive statistics (means and standard deviations) were computed for pre-test and post-test scores for both the experimental and control groups.

Independent-samples t-tests were conducted to examine baseline differences between groups. In addition to null-hypothesis testing, standardized mean differences (Cohen's d) were calculated for pre-test scores to assess practical equivalence between groups at baseline. Before conducting ANCOVA and MANCOVA to examine the effect of Bearly AI on reading comprehension and its sub-skills, key statistical assumptions were examined. Normality was assessed using the Shapiro–Wilk test, homogeneity of variances using Levene's test, linearity was inspected via bivariate scatterplots, and homogeneity of regression slopes was tested using Group × pre-test interaction terms. The homogeneity of regression slopes assumption was met, supporting the use of ANCOVA/MANCOVA. Minor deviations from normality were within acceptable limits given the balanced group sizes and the robustness of these procedures. Following a significant multivariate effect, planned univariate ANCOVAs were conducted for each reading sub-skill. A Bonferroni-adjusted alpha level was applied to control for familywise error, and partial eta-squared (ηp^2) was reported as an estimate of effect size for each outcome. Effect sizes were interpreted using partial eta-squared (ηp^2), with values of 0.01, 0.06, and 0.14 indicating small, medium, and large effects, respectively.

Results

To determine the equivalence of the two study groups in reading comprehension, the reading pre-test was administered during the first week of the semester. Independent-samples t-tests were conducted for the two groups on each sub-skill and on the overall score. The results indicate that the differences between the mean scores of the experimental and control groups across the five reading comprehension sub-skills and on the overall score were not statistically significant ($p > 0.05$) (Supplementary Table 3).

In addition to independent-samples t-tests, standardized mean differences (Cohen's d) were calculated to examine practical equivalence between groups at baseline. Effect sizes for pre-test differences were small across all reading sub-skills ($|d| \leq 0.27$) and small for the total reading comprehension score ($d = 0.32$), indicating negligible practical differences between the experimental and control groups prior to the intervention.

Means and standard deviations of the pre-test and post-test reading comprehension scores for both the control and experimental groups were calculated (Table 1). The overall mean score of the experimental group (32.52 out of a possible 40) was higher than that of the control group (25.56) on the reading comprehension post-test. The test scores for the five reading comprehension sub-skills were also analyzed (Table 1). The post-test mean scores of the experimental group were higher than those of the control group across all five reading comprehension sub-skills: vocabulary, literal comprehension, inferential comprehension, analysis, and evaluation.

Table 1. *Means, Standard Deviations, and Standard Errors of the Students' Scores for the Five Reading Comprehension Sub-skills and Overall on the Pre- and Post-tests, Along with Values Adjusted using Pre-test Score as a Covariate*

Reading comprehension skill	Group (N = 27)	Pre-test		Post-test, unadjusted		Post-test, adjusted	
		Mean	SD	Mean	SD	Mean	SE
Overall	Experimental	20.15	3.483	32.52	3.599	32.314	0.748
	Control	18.96	4.033	25.56	4.432	25.760	0.748
Vocabulary	Experimental	4.30	1.877	7.04	1.126	7.001	0.244
	Control	4.70	1.409	5.19	1.469	5.221	0.244
Literal comprehension	Experimental	4.44	1.649	6.33	1.387	6.260	0.235
	Control	4.07	1.072	5.26	1.023	5.332	0.235
Inferential comprehension	Experimental	3.52	1.695	6.74	0.944	6.728	0.202
	Control	3.15	1.322	5.30	1.103	5.309	0.202
Analysis	Experimental	4.04	1.951	6.63	0.926	6.555	0.213
	Control	3.48	2.486	5.07	1.412	5.149	0.213
Evaluation	Experimental	3.85	1.680	5.78	1.311	5.785	0.255
	Control	3.56	1.553	4.74	1.403	4.733	0.255

The total overall scores are out of 40. SD, standard deviation; SE, standard error.

To examine whether Bearly AI was associated with differences in overall reading comprehension outcomes after controlling for pre-test performance, a one-way analysis of covariance (ANCOVA) was conducted (Supplementary Table 4). The results indicated statistically significant group differences in post-test reading comprehension scores. The partial eta-squared value of 0.427 indicates that 42.7% of the variance in post-test scores can be attributed to group membership (AI-assisted), suggesting that the use of Bearly AI was associated with higher reading comprehension scores.

Furthermore, the adjusted means, standard errors, and standard deviations of the overall scores for the two groups were calculated after controlling for pre-test differences (Table 1), and students' performance again differed significantly between groups. Accordingly, the Bearly AI–based method enhanced the overall reading comprehension performance of the experimental group. The five reading comprehension sub-skills were also analyzed using ANCOVA (Table 1). Differences between the two groups across the five sub-skills remained statistically significant after controlling for pre-test scores, indicating that the Bearly AI–integrated instructional method enhanced students' performance in each sub-skill.

To assess group differences in the combined set of the five reading comprehension sub-skills while controlling for pre-test scores, a one-way multivariate analysis of covariance (MANCOVA) using Hotelling's Trace was performed (Supplementary Table 5). The experimental and control groups differed significantly on the linear combination of sub-skills. The partial eta-squared value of 0.519 indicates that 51.9% of the variance in the combined sub-skill outcomes can be attributed to the instructional condition.

Following the significant multivariate effect, planned univariate ANCOVAs were conducted for each reading comprehension sub-skill using Bonferroni-adjusted alpha levels. Adjusted *p*-values and partial eta-squared values (ηp^2) are reported in Table 2. The experimental group outperformed the control group across all five sub-skills (Table 2). The largest effect size was observed for vocabulary, followed by inferential comprehension, analysis, evaluation, and literal comprehension.

Table 2. *Follow-up Univariate Analysis on the Five Reading Comprehension Sub-skills After Controlling for the Effect of Pre-test Scores*

Source	Dependent variable	Type III sum of squares	df	Mean squared	F	Sig.	Partial eta-squared
Pre vocabulary (covariate)	Vocabulary	1.727	1	1.727	1.105	0.299	0.023
Pre literal comprehension (covariate)	Literal comprehension	1.339	1	1.339	0.924	0.341	0.019
Pre inferential comprehension (covariate)	Inferential comprehension	0.041	1	0.041	0.039	0.845	0.001
Pre analysis (covariate)	Analysis	11.892	1	11.892	9.952	0.003	0.175
Pre evaluation (covariate)	Evaluation	0.126	1	0.126	0.073	0.788	0.002
Group	Vocabulary	40.633	1	40.633	26.010	< 0.001	.356
	Literal comprehension	11.049	1	11.049	7.626	0.008	0.140
	Inferential comprehension	25.853	1	25.853	24.039	< 0.001	0.338
	Analysis	25.344	1	25.344	21.209	< 0.001	0.311
	Evaluation	14.199	1	14.199	8.284	0.006	0.150

(Continued)

Source	Dependent variable	Type III sum of squares	df	Mean squared	F	Sig.	Partial eta-squared
Error	Vocabulary	73.423	47	1.562	–	–	–
	Literal comprehension	68.098	47	1.449	–	–	–
	Inferential comprehension	50.547	47	1.075	–	–	–
	Analysis	56.164	47	1.195	–	–	–
	Evaluation	80.559	47	1.714	–	–	–
Corrected total	Vocabulary	135.333	53	–	–	–	–
	Literal comprehension	92.759	53	–	–	–	–
	Inferential comprehension	82.981	53	–	–	–	–
	Analysis	106.815	53	–	–	–	–
	Evaluation	110.370	53	–	–	–	–

Discussion

The independent-samples *t*-test revealed a statistically significant difference between the experimental and control groups on the post-test of reading comprehension ($p \leq .05$), favoring the experimental group that received AI-supported instruction. While both groups demonstrated comparable performance on the pre-test, indicating initial equivalence, the experimental group showed significantly higher gains on the post-test.

These results indicate a significant association between participation in AI-supported reading comprehension sessions and improved reading comprehension performance. No statistically significant improvement was observed in the control group following conventional instruction. The findings therefore demonstrate that students who engaged in AI-supported reading activities achieved higher reading comprehension scores than those who received textbook-based instruction under the same curricular content and instructional phases.

Learners in EFL reading contexts are widely reported in the literature to encounter persistent difficulties in comprehending texts, including organizing ideas, responding to comprehension questions, identifying the overall meaning and purpose of passages, inferring the meanings of unfamiliar vocabulary, activating prior knowledge, and distinguishing between main and supporting ideas (Alkhawaldeh, 2012; Altyari, 2017; Soisuwan et al., 2022). These well-established difficulties provide a plausible interpretive framework for understanding why additional instructional support may be associated with improved reading outcomes.

In line with this interpretation, previous studies have demonstrated that AI-assisted instruction can support learners in constructing textual meaning and identifying main and supporting ideas more effectively (Alsadoon, 2021; Haristiani, 2019; Hosseini & Amirkhani, 2025; Kim, 2024). Similarly, chatbot-assisted and AI-supported reading environments have been shown to promote self-paced engagement and strategy use, which are associated with improved reading comprehension performance (Pan et al., 2025). Although such learning processes were not directly measured in the present study, these findings offer a theoretically grounded explanation for the observed performance differences between the experimental and control groups.

Another plausible explanation relates to instructional design. Conventional textbook-based instruction is often characterized by uniform pacing and limited differentiation due to constraints related to teachers' time and workload. This limitation reduces the feasibility of sustained individualized instruction, which may contribute to learners' cognitive strain and reduced confidence when tasks exceed their current proficiency levels and may not adequately address individual differences in reading proficiency (Choi et al., 2025). In contrast, AI-supported learning environments are theoretically aligned with principles of personalized learning and adaptive scaffolding and have the potential to support learners' self-efficacy, particularly among students with lower English proficiency, through the generation of tailored content, such as leveled reading materials and practice quizzes aligned with learners' linguistic proficiency (Mollick & Mollick, 2023). Prior research indicates that personalized instructional support can enhance learners' academic engagement and achievement (Al Nabhani et al., 2025; Chiu et al., 2023; Saadi & Saadat, 2015; Slamet, 2024). From this perspective, the structured integration of Bearly AI across the reading phases may have afforded learners differentiated forms of support that are less readily available in traditional classroom practices, which could help account for the higher post-test performance observed in the experimental group.

Prior research suggests that limited access to supplementary digital materials may constrain engagement in traditional reading classrooms, whereas electronic resources can enhance motivation and involvement (Al Nabhani et al., 2025). In addition, AI is used to provide students with instant feedback on areas such as grammar, pronunciation, speaking, and writing (Griche & Bennis, 2026). Electronic feedback has also been associated with lower anxiety and a greater willingness to engage with errors compared to public teacher correction (Saadi & Saadat, 2015). Although these affective factors were not directly measured in the present study, they provide a theoretically grounded account of how AI-supported feedback and resources may be associated with improved reading outcomes.

Taken together, the superior performance of the experimental group may be interpreted in light of established theoretical and empirical work on AI-supported instruction and differentiated learning environments. However, given the reliance on pre- and post-test measures as the sole data source, these explanations remain theoretically grounded interpretations rather than empirically verified mechanisms within the present study.

Strengths and Limitations

The present study's rigorous methodology and the caliber of the evidence it generates constitute its main strengths. The theoretical framework guiding this study draws upon several key learning and instructional theories relevant to foreign language acquisition and the integration of technology in education. Moreover, by evaluating improvements in reading comprehension objectively and reducing bias through standardized procedures, the findings provide strong empirical evidence for the positive impact of AI-assisted instruction on EFL learners' reading performance. In this context, the quasi-experimental design with pre- and post-testing provides robust quantitative evidence. In addition, the use of an expert-approved instrument—the reading comprehension test—further enhances the reliability and validity of the results. Furthermore, the study is theoretically grounded in high-quality literature and a variety of empirical investigations, which further justifies the intervention and enhances the credibility of the findings.

Moreover, because the intervention was implemented in actual classroom settings using authentic curricular materials, the study demonstrates high practical applicability, and the findings are directly relevant to school-based EFL classroom instruction. Another strength is that Bearly AI's differentiated training enhances ecological validity and learner engagement by providing authentic, tailored learning experiences. Finally, because the study examined Jordanian female tenth-grade EFL learners, a demographic group that has received limited research attention, it provides context-specific insights that extend research on AI-assisted reading instruction.

The present study has some limitations. It was conducted at a single secondary school for girls under the Zarqa Directorate of Education, with a relatively small sample (54 female tenth-grade students). Participants were selected using purposive sampling based on school access and laboratory availability, which may have introduced selection bias and limited the representativeness of the sample. Although adequate for a quasi-experimental design, the single-site context and sample size constrain the generalizability of the findings. In addition, intact classes were used with group-level assignment, which limits causal inference.

Data collection relied solely on pre- and post-tests of reading comprehension; no qualitative or process-level data (e.g., engagement, strategy use, or AI interaction logs) were collected, and explanatory interpretations are therefore theoretically informed rather than empirically verified. Using the same test for pre- and post-assessment may have introduced practice effects; although a four-week interval was maintained and test items were not reviewed during instruction, familiarity effects cannot be ruled out. Finally, the short intervention period (second semester, 2024/25) precluded examination of long-term effects, and outcomes were limited to reading comprehension, excluding other language skills.

Practice and Research Implications

The findings are expected to assist English language teachers by offering innovative and effective AI-based instructional strategies that can enhance students' reading comprehension and critical thinking abilities. Educational AI tools such as Bearly should be integrated into high school English curricula, as they have shown a positive impact on developing reading comprehension skills and motivation. Governments and educational institutions should invest in developing the necessary technological infrastructure by providing an adequate number of devices, reliable internet access, and technical support to ensure the effective use of AI tools in the classroom.

The study may encourage teachers to integrate technology into their classrooms in ways that promote engagement, autonomy, and motivation among learners. Teachers are encouraged to combine traditional instructional approaches with AI-based tools to create a balanced learning environment that blends human interaction with technological innovation. Specialized training programs should be designed to equip teachers with the necessary skills to implement AI tools effectively in classrooms. Training should focus on how to use AI to personalize instruction and enhance students' learning outcomes.

In the Jordanian context, the study may contribute to improving the achievement of tenth-grade students in English by demonstrating how AI tools can be effectively integrated into existing curricula. For students, the use of Bearly AI may provide an interactive and adaptive learning environment that supports the development of their language proficiency.

Regarding research implications, more longitudinal and large-scale experimental studies should be conducted to explore the long-term effects of using AI tools on student achievement, attitudes, and language proficiency. Future research could investigate the effectiveness of Bearly AI in developing other English language skills, such as writing, listening, and speaking. Studies may be conducted in other governorates across the Hashemite Kingdom of Jordan and at different educational levels—such as middle school or university—to evaluate the platform's impact in varied learning contexts.

Comparative studies could be carried out to examine the effectiveness of Bearly relative to other AI-based learning tools for improving reading comprehension. Further research could also focus on the perspectives of other educational stakeholders, such as policymakers, curriculum developers, and school administrators, to understand their views on integrating AI into English language teaching and learning.

Future studies are encouraged to adopt larger, multi-site samples and mixed-methods designs to enhance internal validity and extend the generalizability of findings across diverse educational contexts and language skills. The use of parallel or counterbalanced test forms is also recommended to reduce potential testing effects. In addition, integrating learning analytics and qualitative process data (e.g., interaction logs, learner reflections, and teacher observations) would enable a more fine-grained examination of how specific features of AI-supported platforms contribute to the development of reading comprehension.

Conclusion

This study investigated the effect of integrating Bearly AI into reading comprehension instruction for female tenth-grade EFL students. The findings indicate that AI-supported instruction was associated with higher reading comprehension performance compared with conventional textbook-based instruction. Students who learned through Bearly AI demonstrated greater improvement in overall reading comprehension and across key reading sub-skills, suggesting that AI integration can enhance multiple dimensions of reading development.

These outcomes can be interpreted in light of previous research highlighting the potential of AI-supported and technology-enhanced learning environments to provide personalized support, adaptive scaffolding, and timely feedback. The structured integration of Bearly AI across the reading phases appears to have offered learners additional instructional support beyond what is typically available in conventional classrooms, which may have contributed to their improved performance.

Overall, the study provides empirical support for the pedagogical value of AI tools in EFL reading instruction, particularly within secondary school contexts. Although the findings are subject to design and sampling limitations, they underscore the potential of AI-assisted learning as an effective complement to traditional teaching practices and point to the need for further research into the long-term and broader impacts of AI integration in language education.

Acknowledgments

We thank the Ministry of Education in Jordan for approving this research. We also thank the Zarqa Directorate of Education for facilitating this work. Special thanks go to the teachers and students who participated in this research.

Declaration of Interest Statement

The authors declare no conflicts of interest regarding the publication of this paper.

References

Abdelaal, N. M., & Sase, A. S. (2014). Relationship between prior knowledge and reading comprehension. *Advances in Language and Literary Studies*, *5*(6), 125–131. https://doi.org/10.7575/aiac.alls.v.5n.6p.125

Al Nabhani, F., Hamzah, M. B., & Abuhassna, H. (2025). The role of artificial intelligence in personalizing educational content: Enhancing the learning experience and developing the teacher's role in an integrated educational environment. *Contemporary Educational Technology*, *17*(2), ep573. https://doi.org/10.30935/cedtech/16089

Alkhawaldeh, A. (2012). High school students' challenges in English reading comprehension in Amman Second Directorate of Education. *Journal of Instructional Psychology*, *39*(3), 214.

Alsadoon, R. (2021). Chatting with AI Bot: Vocabulary learning assistant for Saudi EFL learners. *English Language Teaching*, *14*(6), 135–157. https://doi.org/10.5539/elt.v14n6p135

Al-salameen, A., & Abdelrahman, M. S. B. (2023). The effect of using the six thinking hats method on the development of EFL female eleventh-grade students' reading comprehension skills in Petra Directorate of Education. *Journal of Education and Practice*, *14*(12), 26–39. https://doi.org/10.7176/JEP/14-12-03

Altyari, A. W. (2017). English vocabulary uptake by Saudi Arabic-speaking students at public schools. *British Journal of English Linguistics*, *5*(1), 10–16.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Chan, K. L. R. (2025). Exploring generative artificial intelligence-aided language learning tools in language classrooms: Insights from language teachers of English in Hong Kong. *Technology in Language Teaching & Learning*, *7*(2), 103217. https://doi.org/10.29140/tltl.v7n2.103217

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Choi, K. Y., Wu, C., & Moorhouse, B. L. (2025). Exploring the use of Generative Artificial Intelligence (GenAI) in English language teaching: Voices from in-service teachers at an early-adopting Hong Kong secondary school. *Technology in Language Teaching & Learning*, *7*(3), 102516. https://doi.org/10.29140/tltl.v7n3.102516

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review*, *20*(4), 631–642. https://doi.org/10.3758/s13423-013-0386-3

Etaat, F. (2024). *The effect of AI-based applications on EFL writing skill development: An inquiry into integration of AI into language learning* [Master's thesis, UiT Norges arktiske universitet]. https://munin.uit.no/handle/10037/33935

Fareh, S. (2010). Challenges of teaching English in the Arab world: Why can't EFL programs deliver as expected? *Procedia - Social and Behavioral Sciences, Innovation and Creativity in Education*, *2*(2), 3600–3604. https://doi.org/10.1016/j.sbspro.2010.03.559

Fernández-Río, J. (2016). Student-teacher-content-context: Indissoluble ingredients in the teaching-learning process. *Journal of Physical Education, Recreation & Dance*, *87*(1), 3–5. https://doi.org/10.1080/07303084.2016.1110476

Ferris, D. (2007). Preparing teachers to respond to student writing. *Journal of Second Language Writing*, *16*(3), 165–193. https://doi.org/10.1016/j.jslw.2007.07.003

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*(1), 102827. https://doi.org/10.29140/tltl.2026.102827

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, *3*, 275–285. https://doi.org/10.1016/j.susoc.2022.05.004

Hamad, S., Tairab, H., Wardat, Y., Rabbani, L., AlArabi, K., Yousif, M., Abu-Al-Aish, A., & Stoica, G. (2022). Understanding science teachers' implementations of integrated STEM: Teacher perceptions and practice. *Sustainability*, *14*(6), 3594. https://doi.org/10.3390/su14063594

Haristiani, N. (2019). Artificial intelligence (AI) chatbot as language learning medium: An inquiry. *Journal of Physics: Conference Series*, *1387*(1), 012020. https://doi.org/10.1088/1742-6596/1387/1/012020

Hosseini, A., & Amirkhani, Z. (2025). Exploring AI chatbot affordances in EFL higher education: Impacts on language proficiency, usability, engagement, and perceived effectiveness. *International Journal of Multicultural and Multireligious Understanding*, *12*(3), Article 3. https://doi.org/10.18415/ijmmu.v12i3.6682

Hwang, W., Hoang, A., & Lin, Y. (2021). Smart mechanisms and their influence on geometry learning of elementary school students in authentic contexts. *Journal of Computer Assisted Learning*, *37*(5), 1441–1454. https://doi.org/10.1111/jcal.12584

Johnson, E. (2014). *Student's book tenth grade: Action pack 10*. York Press.

Kim, R. (2024). Effects of ChatGPT on Korean EFL Learners' Main-Idea Reading Comprehension via Top-Down Processing. *Language Research*, *60*(1), 83–106. https://doi.org/10.30961/lr.2024.60.1.83

Koehler, M. J., Mishra, P., Kereluik, K., Shin, T. S., & Graham, C. R. (2014). The technological pedagogical content knowledge framework. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 101–111). Springer. https://doi.org/10.1007/978-1-4614-3185-5_9

Krashen, S. D. (1982). *Principles and practice in second language acquisition* (1st ed). Pergamon.

Krismadayanti, A., & Zainil, Y. (2022). The level of the students' reading comprehension analyzed by using Barrett taxonomy. *Journal of Cultura and Lingua*, *3*(1), 39–48. https://doi.org/10.37301/culingua.v3i1.110

Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4391243

Rahman, Md. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, *13*(9), 5783. https://doi.org/10.3390/app13095783

Rao, P. S. (2019). The role of English as a global language. *Research Journal of English*, *4*(1), 65–79.

Saadi, Z. K., & Saadat, M. (2015). EFL learners' writing accuracy: Effects of direct and metalinguistic electronic feedback. *Theory and Practice in Language Studies*, *5*(10), 2053. https://doi.org/10.17507/tpls.0510.11

Slamet, J. (2024). Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions. *Discover Artificial Intelligence*, *4*(46), 15. https://doi.org/10.1007/s44163-024-00143-2

Smith, R. J., & Barrett, T. C. (1979). *Teaching reading in the middle grades*. Addison-Wesley Publishing Company.

Soisuwan, T., Sinturat, T., & Lekdamrongsak, P. (2022). The investigation in reading problems and reading strategies of undergraduate students. *Journal of Social Science Studies*, *9*(2), 90. https://doi.org/10.5296/jsss.v9i2.20487

Sucompa, S. (1998). *A survey study of current needs and problems in using technical English for tourism for the higher certificate level students of Rajamangala Institute of Technology* [Unpublished master's thesis]. Mahidol University.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 24. https://doi.org/10.1186/s40561-023-00237-x

Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms*. Association for Supervision & Curriculum Development.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Supplementary Material

Appendix 1

Reading Comprehension Skill Test

Al-Hussein Bin Talal University	**جامعة الحسين بن طلال**
Deanship of Scientific Research & Graduate Studies	**عمادة البحث العلمي والدراسات العليا**
Faculty of Educational sciences	**كلية العلوم التربوية**
Philosophy in General Curricula and Teaching Methods	**الفلسفة العامة وطرق التدريس**

Reading Comprehension Skill Test

Class: 10th Grade **Time: one hour**

Name: **Date:**

ملاحظة: ـ أجيبي عن جميع الأسئلة وعددها (40) و عدد الصفحات (8) مع العلم أن الإجابة على نفس الورقة

Note: Please answer all the following questions: -

. . . .

Question Number One (20 Points)

-Read the following text carefully and then answer the questions that follow. Your answers should be based on the following text:

> Earth's climate has always been in flux, and this is not a recent phenomenon. Over millions of years, Earth has gone through various significant shifts in its climate. For example, numerous ice ages, in which vast glaciers spread across the planet, have occurred. In addition, Earth's climate has been influenced by catastrophic events, such as massive volcanic eruptions and asteroid collisions. These incidents sent ash and debris into the atmosphere, blocking sunlight and cooling the Earth, which caused darkness and lower temperatures. Many species of plants and animals likely perished during these times. Scientists, particularly geologists, study the Earth's layers of rock to understand past climate conditions. The rock layers hold a detailed history, indicating different climate periods. Fossils of ancient plants and animals, preserved within these layers, provide further insight. For instance, in present-day Antarctica, researchers have found fossilized tree remains, proving that the region was once much warmer 100 million years ago. While human activities today may be contributing to global warming, future extreme climate changes could occur if nothing is done to address the situation. No one can say for certain what the future holds, but Earth, which has weathered significant climatic events in the past, may face further challenges ahead.

1. **What does the word *flux* mean in the text?**
   - A) Stability
   - B) Change
   - C) Balance
   - D) Growth
2. **What does the word *catastrophic* mean in the text?**
   - A) Gradual
   - B) Disastrous
   - C) Harmless
   - D) Predictable
3. **What does the word *fossilized* mean in the passage?**
   - A) Preserved in rock layers
   - B) Destroyed by climate change
   - C) Created by scientists
   - D) Frozen in ice
4. **What does the word *perished* suggest in the passage?**
   - A) Many plants and animals died
   - B) Species moved to other areas
   - C) Animals adapted to climate change
   - D) Scientists studied new species
5. **What has influenced Earth's climate in the past?**
   - A) Solar panels
   - B) Volcanic eruptions
   - C) Human activities
   - D) Ocean currents
6. **Where were fossilized trees found?**
   - A) Antarctica
   - B) Amazon Rainforest
   - C) Sahara Desert
   - D) The Arctic
7. **What do the fossilized trees in Antarctica show?**
   - A) Antarctica was once warmer
   - B) Antarctica was always cold
   - C) Antarctica never had plants
   - D) Antarctica was covered in ice
8. **What is the main idea of the passage?**
   - A) Earth's climate has changed many times
   - B) Scientists control climate change
   - C) Climate never changes naturally
   - D) The Earth's future is predictable
9. **What may happen if Earth's climate continues to change?**
   - A) Extreme changes may occur
   - B) The climate will remain the same
   - C) Earth will get colder
   - D) Scientists will stop studying climate
10. **How do scientists study past climates?**
    - A) By studying rock layers
    - B) By predicting the weather
    - C) By collecting air samples
    - D) By using telescopes

11. What happened after volcanic eruptions?
- A) Ash blocked sunlight and cooled Earth
- B) The Earth became warmer
- C) More plants grew
- D) Animals thrived

12. What does the passage suggest about Earth's future climate?
- A) It could change again significantly
- B) It will stay the same forever
- C) Human activities have no effect
- D) Scientists can fully control it

13. What caused climate changes in the past?
- A) Ice ages and volcanic eruptions
- B) Human activities
- C) Solar energy
- D) Pollution

14. What information do rock layers provide?
- A) History of Earth's climate
- B) Locations of earthquakes
- C) Predictions about the weather
- D) Causes of volcanic eruptions

15. How do fossils help scientists understand the past?
- A) They show past climate conditions
- B) They explain modern weather patterns
- C) They describe new animal species
- D) They show future climate changes

16. Why is studying past climate changes important?
- A) To understand future climate challenges
- B) To create new fossils
- C) To control volcanic eruptions
- D) To prevent asteroid collisions

17. What proves that Earth's climate has changed before?
- A) Fossils and rock layers
- B) Modern weather reports
- C) Historical documents
- D) Ocean currents

18. What natural events affected Earth's climate?
- A) Volcanic eruptions and asteroid collisions
- B) Human activities and pollution
- C) Solar panel usage
- D) Agriculture

19. Why do scientists study fossils?
- A) To understand past climate conditions
- B) To discover new planets
- C) To create new species
- D) To predict tomorrow's weather

20. What does the text suggest about Earth's ability to handle climate change?
- A) Earth has faced challenges before but may face more in the future
- B) Earth will not experience climate change again
- C) Climate change is a new problem
- D) Earth's climate will improve without action

Question Number Two (20 Points)

-Read the following text carefully and then answer the questions that follow. Your answers should be based on the following text:

Jordan is globally renowned for its rich history and archaeological wonders, making it a premier destination for cultural tourism. Petra, a UNESCO World Heritage Site, is one of the most famous landmarks in Jordan. Known as the "Rose City" due to the color of its stone, Petra was carved into the rock over 2,000 years ago and was once a bustling trade hub for the Nabataeans. Visitors marvel at its intricate architecture, such as the Treasury and the Monastery, which showcase the advanced engineering skills of ancient civilizations.

Another significant historical site is Jerash, located north of the capital, Amman. It is one of the best-preserved Roman cities in the world. Jerash features grand colonnaded streets, impressive amphitheaters, and ancient temples, offering tourists a glimpse into life during Roman times.

Jordan's historical treasures are complemented by its cultural experiences. Tourists can engage in traditional Bedouin hospitality, sample delicious local cuisine, and explore bustling bazaars. With its unique blend of history, culture, and stunning landscapes, Jordan attracts millions of visitors annually, leaving them with unforgettable memories.

1. **What does the phrase *archaeological wonders* in the passage mean?**
   - A) Ancient ruins and artifacts
   - B) Modern buildings
   - C) Natural landscapes
   - D) Scientific discoveries
2. **What does the word *intricate* mean in the text?**
   - A) Simple
   - B) Detailed
   - C) Ordinary
   - D) Fragile
3. **What does the word *hospitality* refer to in the passage?**
   - A) Welcoming guests warmly
   - B) Building hotels
   - C) Cooking traditional food
   - D) Selling souvenirs
4. **What does the word *bustling* suggest about the trade hub of Petra?**
   - A) It was quiet and isolated
   - B) It was full of activity and life
   - C) It was small and unknown
   - D) It was only used by locals
5. **What is Petra known as?**
   - A) The Golden City
   - B) The Rose City
   - C) The Desert City
   - D) The Nabataean City

6. **Which city is one of the best-preserved Roman cities in the world?**
    - A) Amman
    - B) Petra
    - C) Jerash
    - D) Aqaba
7. **Why do tourists visit Jerash?**
    - A) To see Roman architecture
    - B) To relax on the beach
    - C) To shop in modern malls
    - D) To experience modern culture
8. **What makes Jordan a premier destination for cultural tourism?**
    - A) Its combination of history, culture, and landscapes
    - B) Its modern architecture and shopping malls
    - C) Its agricultural industry
    - D) Its vast deserts and empty spaces
9. **What attracts visitors to Petra?**
    - A) Its ancient architecture and historical significance
    - B) Its sandy beaches and resorts
    - C) Its modern buildings and nightlife
    - D) Its agricultural sites
10. **What is one feature of Jerash mentioned in the text?**
    - A) Colonnaded streets
    - B) Underground tunnels
    - C) Waterfalls
    - D) Palaces
11. **Why is Petra called the "Rose City"?**
    - A) Because of the pinkish color of its stone
    - B) Because roses grow there
    - C) Because it was named after a queen
    - D) Because of its perfume markets
12. **Based on the passage, how does Jordan create unforgettable memories for visitors?**
    - A) Through its historical sites and cultural experiences
    - B) Through its business and trade opportunities
    - C) Through its shopping malls and resorts
    - D) Through its educational institutions
13. **Which ancient civilization built Petra?**
    - A) The Nabataeans
    - B) The Romans
    - C) The Greeks
    - D) The Ottomans
14. **What are some of the main attractions in Petra?**
    - A) The Treasury and the Monastery
    - B) Modern hotels and malls
    - C) Beaches and resorts
    - D) Libraries and schools
15. **How do visitors experience Bedouin hospitality?**
    - A) By engaging in cultural traditions and cuisine
    - B) By staying in modern hotels
    - C) By visiting international restaurants
    - D) By shopping in supermarkets

16. Why do tourists find Jordan appealing as a cultural destination?
- A) It offers a mixture of history, culture, and landscapes
- B) It is a new, unexplored country
- C) It has no modern developments
- D) It only has ancient ruins

17. What makes Petra an important historical site?
- A) Its role as an ancient trade hub
- B) Its modern design and facilities
- C) Its location by the sea
- D) Its advanced transportation system

18. What features make Jerash a unique site for tourists?
- A) Its well-preserved Roman architecture
- B) Its position in the desert
- C) Its modern shopping centers
- D) Its natural hot springs

19. How does Jordan's cultural tourism benefit from Bedouin traditions?
- A) They provide authentic cultural experiences for visitors
- B) They modernize Jordan's economy
- C) They focus only on desert life
- D) They are not popular with tourists

20. What can be concluded about Jordan's appeal to tourists?
- A) It offers rich historical and cultural experience
- B) It is mainly known for shopping
- C) It focuses only on modern attractions
- D) It attracts visitors only for business purposes

The End

Supplementary Table 1. Specification table for the reading comprehension test aligned with the Jordanian Ministry of Education's grading criteria (tenth grade)

Table of specifications						
Skill	**Percentage of skill**	**No. of items**	**Levels of cognition based on Bloom's taxonomy**			
		No. of quest × **per of skills**	Remembering	Understanding + application	Higher thinking levels	Total 100%
		100%	50%	30%	20%	
		20	No. of quest. of a skill × percentage 100%			
Vocabulary	**Gr. (10) 20%**	**4**	2	1.2	0.8	**4**
Literal comprehension	**Gr. (10) 20%**	**4**	2	1.2	0.8	**4**
Inferential comprehension	**Gr. (10) 20%**	**4**	2	1.2	0.8	**4**
Analysis	**Gr. (10) 20%**	**4**	2	1.2	0.8	**4**
Evaluation	**Gr. (10) 20%**	**4**	2	1.2	0.8	**4**
Total	100%					**20**

Supplementary Table 2. Blueprint for the reading comprehension test

Sub-skill	**Cognitive Level (Bloom/MoE)**	**Items per Passage**	**Total Items (Two Passages)**	**Item Numbers (Example)**
Vocabulary	Remembering (LOTS)	2	4	Q1, Q2; Q21, Q22
	Understanding (MOTS)	1	2	Q3; Q23
	HOTS (Analyzing/Evaluating)	1	2	Q4; Q24
Literal Comprehension	Remembering (LOTS)	2	4	Q5, Q6; Q25, Q26
	Understanding (MOTS)	1	2	Q7; Q27
	HOTS (Analyzing/Evaluating)	1	2	Q8; Q28
Inferential Comprehension	Remembering (LOTS)	2	4	Q9, Q10; Q29, Q30
	Understanding (MOTS)	1	2	Q11; Q31
	HOTS (Analyzing/Evaluating)	1	2	Q12; Q32
Analytical Skills	Remembering (LOTS)	2	4	Q13, Q14; Q33, Q34
	Understanding (MOTS)	1	2	Q15; Q35
	HOTS (Analyzing/Evaluating)	1	2	Q16; Q36
Evaluation Skills	Remembering (LOTS)	2	4	Q17, Q18; Q37, Q38
	Understanding (MOTS)	1	2	Q19; Q39
	HOTS (Analyzing/Evaluating)	1	2	Q20; Q40
Total	—	20 per passage	40 total items	—

Supplementary Table 3. Results of the t-test for the equivalence of the two groups in reading comprehension levels after controlling for the pre-test scores

Level	Group	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)
vocabulary	Experimental	27	4.30	1.877	0.902	52	0.371
	Control	27	4.70	1.409			
literal comprehension	Experimental	27	4.44	1.649	0.979	52	0.332
	Control	27	4.07	1.072			
inferential comprehension	Experimental	27	3.52	1.695	0.895	52	0.375
	Control	27	3.15	1.322			
analysis	Experimental	27	4.04	1.951	0.913	52	0.365
	Control	27	3.48	2.486			
evaluation	Experimental	27	3.85	1.680	0.673	52	0.504
	Control	27	3.56	1.553			
Reading Comprehension Skill Test	Experimental	27	20.15	3.483	1.156	52	0.253
	Control	27	18.96	4.033			

Supplementary Table 4. One-way ANCOVA for the effect of teaching method on the overall reading comprehension after controlling for the effect of the pre-test scores

Source	Type III sum of squares	df	Mean squared	F	Sig.	Partial eta-squared
Pre-test (covariate)	87.708	1	87.708	5.888	0.019	0.104
Group	565.455	1	565.455	37.960	< 0.001	0.427
Error	759.700	51	14.896	–	–	–
Corrected total	1501.926	53	–	–	–	–

Supplementary Table 5. One-way MANCOVA for the effect of teaching method on the post-test scores on the five reading comprehension sub-skills.

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial eta-squared
Hotelling's trace	1.080	9.285	5.000	43.000	< 0.001	0.519