Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence, No. 3. 2026. Pp.232- 250
DOI: https://dx.doi.org/10.24093/awej/AI3.15

GenAI, Prompt Literacy, and Early Academic Writing Skills in Saudi EFL Students: A Quasi-Experimental Study

Heena Saifullah Amjad
English Language Instructor/Scientific Research Committee Member
AlKhaleej Training and Education at King Saud University, Riyadh, Saudi Arabia
Corresponding Author: hsaifullah.c@ksu.edu.sa

Hanady Abdulaziz Alfalig
Assistant Professor of Applied Linguistics,
English Language Skills Department,
Common First Year, King Saud University, Riyadh, Saudi Arabia.
Email: halfalig@ksu.edu.sa

Received:10/19 /2025 Accepted:12/17/ /2025 Published:01/24/2026

Abstract

Generative Artificial Intelligence is increasingly used in language education; however, its effectiveness in supporting early academic writing has not been thoroughly evaluated through classroom-based research. This study investigates whether guided text-to-image prompting facilitates vocabulary acquisition, short-term retention, and the development of descriptive writing among Saudi foundation-year EFL students. The research emphasizes responsible, learner-authored applications of Generative Artificial Intelligence, which remain insufficiently explored, particularly in foundational EFL contexts where vocabulary and sentence-level development present ongoing challenges. Two B1-level classes (N = 50) at a Saudi public university were assigned to experimental and control groups for a six-week intervention. The experimental group engaged in vocabulary and writing tasks using text-to-image prompting, while the control group completed parallel text-based activities. A mixed-methods approach was adopted, incorporating both quantitative and qualitative data. Quantitative measures included pre-, post-, and delayed vocabulary tests, as well as sentence-level assessments of lexical diversity, clause density, and modifier use. Qualitative analysis focused on mid- and post-intervention writing to evaluate linguistic integration and descriptive development. Results demonstrate that the experimental group outperformed the control group in vocabulary improvement, retention, sentence elaboration, and descriptive richness. Additionally, learners in the experimental group more consistently integrated target vocabulary into cohesive texts. The study concludes that guided text-to-image prompting serves as a pedagogically responsible multimodal supplement to traditional instruction and recommends further research on GenAI designs that enhance learner-generated language without replacing student writing.

Keywords: academic writing development, Generative Artificial Intelligence, prompt literacy, Saudi EFL, vocabulary learning

How to cite: Amjad, H. S., & Alfalig, H. A. (2026).GenAI, Prompt Literacy, and Early Academic Writing Skills in Saudi EFL Students: A Quasi-Experimental Study. *Arab World English Journal, Special Issue on Artificial Intelligence,* (3):232- 250. https://dx.doi.org/10.24093/awej/AI3.15

Introduction

Generative Artificial Intelligence (GenAI) is increasingly integrated into language-learning contexts, with recent studies emphasizing its role in facilitating feedback, task completion, and writing development (Law, 2024; Balraj, 2025). Tools such as ChatGPT, Poe, and Copilot are frequently employed to enhance vocabulary acquisition, structural organization, and clarity in student writing (Pryma et al., 2025; Kim et al., 2024). Unlike static visuals, text-to-image GenAI systems require learners to construct linguistically precise prompts, emphasizing lexical selection, syntactic accuracy, and descriptive detail. As a result, prompt construction is recognized as an emerging literacy practice within AI-mediated language learning (Tour & Zadorozhnyy, 2025).

In Saudi EFL contexts, vocabulary instruction predominantly relies on bilingual dictionaries, translation, and memorization (Aljurbua, 2021). These approaches offer limited opportunities for contextualized vocabulary use, multimodal engagement, and extended writing practice (Abduljawad, 2024). Current AI-based interventions have primarily emphasized text generation or corrective feedback, raising concerns regarding the preservation of learner authorship (Abduljawad, 2024). Text-to-image prompting presents an alternative by enabling learners to visualize meanings and revise language when generated images do not align with intended messages (Xia et al., 2024). Despite its multimodal potential, empirical classroom-based evidence for GenAI approaches in supporting early academic writing in Saudi EFL settings remains limited (Mahdi et al., 2024; Qian, 2025).

While scholarly interest in GenAI has increased, significant research gaps persist. The majority of existing studies focus on text-based AI tools (Pryma et al., 2025; Tour & Zadorozhnyy, 2025), with limited investigation of GenAI applications that require student-generated language rather than AI-produced text. Research on text-to-image prompting is still emerging and has rarely been explored through controlled classroom studies, particularly in Saudi EFL contexts. To date, no study has examined whether GenAI-mediated image generation can enhance vocabulary retention or descriptive writing among foundation-year learners.

Given these considerations, the present study investigates whether GenAI-mediated text-to-image prompting facilitates vocabulary acquisition, short-term retention, and descriptive writing among Saudi foundation-year EFL students. Additionally, the study compares GenAI-mediated prompting with traditional instructional methods to determine whether multimodal tasks promote lexical diversity, sentence elaboration, and descriptive richness, while preserving student authorship and academic integrity and clarifying the pedagogical role of AI in writing development.

In this study, early academic writing refers to the foundational stage during which learners develop lexical control, sentence-level clarity, and elaboration necessary for later academic genres such as essays or reports (Staples et al., 2016). Descriptive writing, examined through sentence-level elaboration and vocabulary use, is a central element of this stage, emphasizing the linguistic precision required for extended academic writing. While descriptive writing is not classified as an

academic genre, it contributes to early academic writing by developing the sentence-level clarity and lexical precision necessary for more extended academic texts.

To achieve this aim, the study sets out to:

1. Compare the effects of GenAI-mediated text-to-image prompting and traditional instruction on learners' vocabulary acquisition and short-term retention.
2. Examine how GenAI-mediated prompting influences sentence elaboration and descriptive writing, particularly in terms of lexical diversity, syntactic complexity, and descriptive richness.

Accordingly, the research addresses the following questions:

1. How does GenAI-mediated text-to-image prompting influence learners' vocabulary acquisition and short-term retention compared with traditional instruction?
2. In what ways do GenAI-mediated prompting and traditional instruction differentially affect lexical diversity, syntactic complexity, and descriptive richness in students' descriptive writing?
3.

Literature Review

GenAI in Language Learning and Academic Writing

GenAI has gained prominence in higher education for supporting the development of grammar, vocabulary, and academic writing skills (Ali et al., 2025; Karataş et al., 2024; Pegrum, 2025; Pryma et al., 2025). However, much of the existing research is perception-based rather than grounded in measurable linguistic outcomes (Kim et al., 2025; Wang, 2025). Performance-based studies typically address broad writing features (Balraj, 2025), and their findings are mixed. Performance-based studies typically focus on global measures of writing quality, such as fluency, accuracy, or overall scores, and their findings remain mixed (Balraj, 2025). There are concerns that AI-generated feedback may promote only superficial revision (Kasneci et al., 2023). These limitations highlight the need for controlled, classroom-based investigations that focus on specific linguistic outcomes, such as vocabulary acquisition and descriptive writing.

GenAI in Saudi EFL Higher Education

Recent studies in Saudi universities have documented the role of ChatGPT in supporting English language learning. Ali et al. (2025) reported improvements in grammar among students receiving AI-assisted instruction. Abduljawad (2024) observed improvements in writing fluency, accuracy, and vocabulary among learners using ChatGPT, whereas Mugableh (2024) found that AI-generated exercises led to greater vocabulary growth than traditional tasks. However, Saudi research has primarily focused on AI tools that generate or revise language for learners, providing limited insight into approaches that maintain learner authorship or enhance the form–meaning mapping essential for descriptive writing. This gap is especially pronounced in foundation-year programs, where persistent challenges include vocabulary use, sentence construction, and elaborated description.

Vocabulary Development Mechanisms

Research on vocabulary learning highlights the significance of repeated retrieval, generative processing, and contextualized use in developing sound lexical knowledge (Laufer & Girsai, 2008; Nakata & Webb, 2016; Schmitt, 2020). Traditional Saudi EFL practices, such as the use of bilingual dictionaries, translation, and isolated-sentence production, provide limited opportunities for deeper learning processes. GenAI has demonstrated potential for vocabulary development, as evidenced by studies utilizing AI-generated exercises or feedback (Mugableh, 2024). However, these interventions frequently provide learners with pre-constructed language rather than engaging them in productive lexical decision-making. Approaches that require learners to independently generate and refine vocabulary, such as text-to-image prompting, remain underexplored.

Writing Development and Multimodality

Multimodal approaches to language learning emphasize that integrating visual and verbal representations enhances comprehension, retention, and form–meaning mapping (Mayer, 2014; Mahdi et al., 2024). L2 writing development similarly involves a progression from clause chaining to more controlled phrasal and thematic organization, as indicated by measures such as type-token ratio (TTR), clause density, and modifier use (Laufer & Nation, 1995; Lu, 2010; Staples et al., 2016; Biber et al., 2022). GenAI tools can facilitate lexical refinement and structural accuracy through iterative revision (Kim et al., 2025). Text-to-image prompting further enables learners to visualize the meaning encoded in their lexical and syntactic choices and to adjust language when generated images do not align with their intended meaning (Xia et al., 2024). Despite these instructional possibilities, empirical research on multimodal GenAI remains limited (Qian, 2025). In contrast to text-generating AI, text-to-image prompting preserves learner authorship while providing visual scaffolding.

Prompt Literacy and Multiliteracies as Conceptual Foundations

Prompt literacy has emerged as a key construct for understanding how learners plan, monitor, and revise prompts to guide the generation of GenAI output (Tour & Zadorozhnyy, 2025). Research indicates that effective prompting is a learnable, iterative skill that depends on linguistic precision and strategic design (Kim et al., 2025; Oppenlaender, 2024). In GenAI-mediated tasks, prompt literacy positions learners as active designers of language, focusing attention on lexical choice, syntactic control, and meaning negotiation. Refining prompts in response to discrepancies between intended meaning and generated images promotes deeper processing, which supports syntactic elaboration and descriptive clarity.

Multiliteracies pedagogy complements this perspective by conceptualizing literacy as design across linguistic, visual, and digital modes (New London Group, 1996; Cope & Kalantzis, 2009). In text-to-image prompting, learners express meaning through language, interpret visual feedback, and revise their wording when the generated image diverges from their intended

message, reflecting the core processes of multiliteracies. Collectively, prompt literacy and multiliteracies highlight the pedagogical value of multimodal GenAI tasks for enhancing form–meaning mapping. At the same time, few studies have operationalized prompt literacy or multiliteracies in controlled classroom settings to assess vocabulary or writing development (Qian, 2025).

Across the reviewed studies, three gaps are evident: GenAI research remains largely focused on text-based tools and perception studies rather than specific linguistic outcomes; Saudi studies focus on AI-generated text and seldom address pedagogical approaches that preserve learner authorship or target vocabulary acquisition and descriptive writing; and multimodal GenAI tools such as text-to-image prompting lack empirical testing in foundation-year EFL contexts where vocabulary and sentence-level development are key instructional goals.

The present study addresses these gaps by comparing GenAI-mediated text-to-image prompting with traditional instruction to examine its effects on vocabulary acquisition, short-term retention, and descriptive writing, grounded in the prompt literacy and multiliteracies framework.

Methodology

Research Design

This study employed a quasi-experimental mixed-methods design to examine the effectiveness of Generative AI in vocabulary instruction and descriptive writing. Two instructional conditions were compared: an experimental group using text-to-image GenAI tools and a control group receiving traditional teacher-led instruction. In the experimental group, students generated prompts using Poe and Copilot Image Creator and refined them by adding contextual, sensory, and spatial details until the generated images matched their intended meanings. The control group followed a Presentation–Practice–Production (PPP) approach typical of grammar-translation-based vocabulary teaching, involving teacher explanation, dictionary consultation, pair discussion, and example-sentence writing (Richards & Rodgers, 2014; Zhu et al., 2024).

The quantitative strand assessed vocabulary gains through pre-, post-, and delayed-retention tests and tracked sentence development using weekly measures of lexical diversity, syntactic complexity, and descriptive elaboration. The qualitative strand analyzed mid- and post-intervention writing samples (Weeks 4 and 7) to identify changes in vocabulary use, sentence elaboration, and descriptive expression. The design was informed by Multiliteracies Pedagogy (New London Group, 1996; Cope & Kalantzis, 2009) and Prompt Literacy, which views prompting as an iterative, metalinguistic process (Tour & Zadorozhnyy, 2025).

Participants

The study was conducted at a public university in Saudi Arabia, with 50 foundation-year EFL students at the CEFR B1 level, as determined by institutional placement testing. Participants were selected from two intact course sections on the female campus, with ages ranging from 18 to 20, reflecting the typical foundation-year demographic. Convenience sampling identified the two

sections, which were subsequently randomly assigned to experimental and control groups using cluster-level randomization. Both groups were taught by the same instructors and followed the same syllabus and coursebook to ensure equivalent instructional conditions, with the experimental group receiving GenAI-mediated tasks. The instructional intervention lasted six weeks, followed by a descriptive writing task in Week 7 and a delayed retention test in Week 9. All students provided informed consent, and the study adhered to institutional ethical procedures.

Instructional Procedures

Both groups used the same target vocabulary and completed meaning-identification and sentence-production tasks during regular lessons throughout the six-week intervention. In the GenAI group, students used their own sentences as text-to-image prompts using Poe and Copilot Image Creator. These tools were accessed via free browser-based interfaces: Poe generated images through diffusion models (Stable Diffusion XL and Playground v2.5), while Copilot Image Creator utilized DALL·E 3. Following the teacher's guidance, students revised their prompts by incorporating contextual, spatial, and sensory details to ensure that the images aligned with their intended meanings. Neither the teacher nor the AI generated sentences or paragraphs for the students. The traditional group completed the same tasks without AI support, following teacher modelling and controlled practice within a Presentation–Practice–Production sequence. Peer review and teacher feedback were provided in both groups; however, iterative prompt–image–revision cycles were exclusive to the GenAI condition. All other instructional elements, including syllabus, pacing, and materials, remained consistent across groups.

Research Instruments

Data was collected from three sources. The first consisted of pre-, post-, and delayed-retention vocabulary tests designed to measure learners' acquisition and short-term retention of the 120 target vocabulary items. Each test included sentence-completion and recognition items and followed established L2 vocabulary-testing procedures (Laufer & Girsai, 2008; Nakata & Webb, 2016; Zhu et al., 2024). Two instructors reviewed test items to ensure alignment with the instructional vocabulary list and content validity.

The second instrument comprised the weekly sentences that students produced during regular lessons, providing an ongoing record of lexical use and syntactic development throughout the intervention. These sentences were coded for type–token ratio, sentence length, clause density, modifier use, and inclusion of target vocabulary through a semi-automated NLP tagging process with subsequent manual verification. Features of descriptive writing were evaluated using an analytic rubric adapted from Laufer and Nation (1995), Lu (2010), and Cope and Kalantzis (2009).

The third instrument involved two descriptive-writing tasks completed at mid-intervention and post-intervention. These samples were analyzed using the same rubric to trace changes in learners' vocabulary use, structural complexity, and descriptive elaboration. To ensure scoring reliability, 30% of the samples were double-rated after a calibration session, yielding strong inter-rater consistency (ICC = .86), and discrepancies were resolved through discussion. A small pilot

study with eight students was also conducted to confirm task clarity, test functionality, and refine scoring procedures; minor adjustments were made based on this feedback.

Data Collection

Data was collected over nine weeks. Vocabulary learning was assessed using a pre-test (Week 1), post-test (Week 6), and delayed-retention test (Week 9). Weekly sentence samples were gathered during the six instructional weeks. Descriptive-writing data was collected twice: students revised sentences using that week's vocabulary list following teacher feedback (Week 4) and later produced a full descriptive paragraph (Week 7). In the GenAI condition, students wrote in response to randomly selected images they had previously generated, whereas the control group completed parallel tasks based on textbook activities. All other instructional conditions (materials, pacing, and syllabus coverage) remained identical across groups.

Research Procedures

A mixed-methods approach was used to integrate quantitative and qualitative evidence on the effects of GenAI-mediated instruction. Quantitative analyses examined changes in vocabulary acquisition and writing performance across groups. Descriptive statistics (means, standard deviations, and percentages) were calculated for all tests and writing measures. Differences in pre-, post-, and delayed scores were analyzed using paired-sample and independent-sample *t*-tests (Field, 2018). Writing measures (sentence length, clause density, TTR, modifier use, and target-word inclusion) were analyzed in the same way to track development over the six weeks.

For the quantitative analysis of weekly sentence data, each group produced approximately 65–70 sentences per week. A stratified sample of 40 sentences per group per week (≈240 per group) was coded for lexical, structural, and descriptive features using an analytic rubric adapted from Laufer and Nation (1995), Lu (2010), and Cope and Kalantzis (2009). Coding combined a semi-automated NLP pipeline (spaCy) with manual verification.

Qualitative analysis of Week 4 and Week 7 writing tasks followed Braun and Clarke's (2006) thematic guidelines, focusing on vocabulary integration, clause expansion, and contextual detail. Findings from both strands were integrated to provide a comprehensive understanding of how GenAI-mediated prompting shaped learners' vocabulary development and writing.

Ethical Considerations

Ethical approval was obtained from the university's Institutional Review Board. Informed consent was collected from all participants. Data was anonymized at the point of collection and used solely for research purposes.

Findings

This section reports quantitative results on vocabulary learning, sentence elaboration, and descriptive writing, and qualitative insights from Week 4 and Week 7 samples showing how GenAI-mediated prompting shaped vocabulary use, syntactic development, and descriptive detail.

Vocabulary Test Results

Pre-, post-, and delayed vocabulary tests assessed the effects of the instructional conditions on vocabulary learning and retention. Table 1 presents the descriptive statistics, and Table 2 summarizes the corresponding inferential results. Descriptive Statistics for Pre-, Post-, and Delayed Vocabulary Tests

Table 1. *Descriptive statistics for pre-, post-, and delayed- vocabulary tests of experimental and control groups*

Group	N	Pre-Test Mean (SD)	Post-Test Mean (SD)	Delayed-Test Mean (SD)	Mean Gain (Post – Pre)	Retention Change (Delayed – Post)	Retention Ratio (%)
Experimental	25	13.12 (2.19)	24.80 (1.98)	24.40 (2.08)	11.68	− 0.40	98.39
Control	25	13.56 (1.92)	22.48 (1.83)	20.48 (1.90)	8.92	−2.00	91.10

Note. Scores out of 30. Mean Gain = Post – Pre. Retention Difference = Delayed – Post (a smaller negative value indicates stronger retention). Retention Ratio = (Delayed ÷ Post) × 100.

Table 1 shows substantial pre-to-post-test improvement in both groups. The experimental group increased from 13.12 (SD = 2.19) to 24.80 (SD = 1.98), and the control group from 13.56 (SD = 1.92) to 22.48 (SD = 1.83). At delayed testing, the experimental group maintained a mean of 24.40 (SD = 2.08), a minor decline of 0.40 points, whereas the control group declined by 2.00 points to 20.48 (SD = 1.90). Accordingly, the experimental group showed a larger mean gain (+11.68 vs. +8.92) and a higher retention ratio (98.39% vs. 91.10%).

Table 2. *Paired- and independent-sample t-test results for vocabulary performance*

Comparison	t	df	P	Cohen's d
Pre- vs Post Test (Experimental; Paired)	24.58	24	< .001	5.60
Pre- vs Post Test (Control; Paired)	16.87	24	< .001	4.76
Post Test (Exp vs Ctrl; Independent)	4.31	48	< .001	1.22
Post vs Delayed Test (Experimental; Paired)	–1.68	24	0.106	- 0.20
Post vs Delayed Test (Control; Paired)	–4.40	24	< .001	- 1.07
Delayed Test (Exp vs Ctrl; Independent)	6.96	48	< .001	1.97

Note. $p < .05$ indicates statistical significance. Cohen's *d* values of 0.2 = small, 0.5 = medium, 0.8 = large (Field, 2018).

As shown in Table 2, paired *t*-tests confirmed significant pre-to-post gains for both groups ($p < .001$), with a larger effect size for the experimental group ($d = 5.60$) than for the control group ($d = 4.76$). Independent-sample *t*-tests demonstrated that the experimental group outperformed the control group on both the post-test ($t(48) = 4.31$, $p < .001$, $d = 1.22$) and delayed test ($t(48) = 6.96$, $p < .001$, $d = 1.97$). Between post- and delayed testing, the experimental group's slight decline was

not statistically significant (p = .106), while the control group showed a significant decline (p < .001).

Overall, the results indicate that the GenAI-mediated condition produced larger immediate vocabulary gains and stronger short-term retention than traditional instruction.

Vocabulary Use and Sentence Development

Weekly sentence samples were analyzed to track lexical and structural development during the six-week intervention. A stratified sample of 40 sentences per group per week was coded for type–token ratio (TTR), sentence length, clause density, modifier use, and target-word integration. Summary measures appear in Tables 3 and 4.

Table 3. *Lexical and structural measures during the six-week intervention*

Week	N (Exp)	N (Ctrl)	TTR (Exp)	TTR (Ctrl)	Mean Sentence Length (Exp)	Mean Sentence Length (Ctrl)	Clauses per Sentence (Exp)	Clauses per Sentence (Ctrl)
1	40	40	0.543	0.501	8.44 (1.55)	8.38 (1.39)	1.27 (0.45)	1.25 (0.44)
2	40	40	0.544	0.512	10.51 (2.92)	8.15 (2.82)	1.59 (0.84)	1.15 (0.43)
3	40	40	0.540	0.544	10.46 (2.75)	8.32 (2.09)	1.63 (0.66)	1.25 (0.49)
4	40	40	0.470	0.548	13.61 (3.69)	8.68 (3.41)	2.22 (0.85)	1.25 (0.49)
5	40	40	0.446	0.541	21.51 (8.01)	10.47 (1.84)	3.49 (2.25)	1.40 (0.55)
6	40	40	0.444	0.488	23.68 (9.62)	10.57 (2.06)	3.59 (2.12)	1.55 (0.64)

Note. Values are means with standard deviations in parentheses. TTR = type–token ratio; N = number of student texts per group. TTR values represent aggregated group-level means for each week; standard deviations are not applicable.

Table 3 shows that Type–token ratios (TTR) values remained relatively stable in both groups (Exp: 0.54 → 0.44; Ctrl: 0.50 → 0.49), indicating that increased sentence complexity did not reduce lexical diversity, a pattern of controlled lexical recycling consistent with higher-order lexical sophistication (Laufer & Nation, 1995). More pronounced differences appeared in structural measures. Mean sentence length in the experimental group increased from 8.4 to 23.7 words, compared with a smaller rise from 8.4 to 10.6 words in the control group. Clause density showed a similar pattern, rising from 1.27 to 3.59 clauses per sentence in the experimental group and from 1.25 to 1.55 in the control group, indicating comparatively greater syntactic elaboration in the GenAI-mediated condition. Independent-samples t-tests showed significant between-group differences in sentence length and clause density from Week 2 onward (p < .001), with large effect sizes (d > 1.0). These results indicate markedly greater syntactic elaboration under the GenAI-mediated condition.

Table 4. *Descriptive and expressive measures during the six-week intervention*

Week	N (Exp)	N (Ctrl)	Adj/Adv per Word (Exp)	Adj/Adv per Word (Ctrl)	% Sentences with ≥ 1 Target (word/phrase level) (Exp)	% Sentences with ≥ 1 Target (word/phrase level) (Ctrl)	Target Words Matched per Sentence (Exp)	Target Words Matched per Sentence (Ctrl)
1	40	40	0.14 (0.09)	0.165 (0.123)	100 (0)	97.5 (15.8)	1.00 (0)	1.05 (0.52)
2	40	40	0.10 (0.07)	0.160 (0.116)	100 (0)	95.0 (22.1)	1.05 (0.31)	1.02 (0.44)
3	40	40	0.09 (0.08)	0.159 (0.137)	97.6 (15.6)	92.5 (26.7)	1.00 (0.22)	0.97 (0.50)
4	40	40	0.10 (0.06)	0.154 (0.168)	100 (0)	82.5 (38.5)	1.27 (0.5)	0.90 (0.55)
5	40	40	0.11 (0.06)	0.141 (0.132)	100 (0)	77.5 (42.3)	2.15 (0.79)	0.88 (0.52)
6	40	40	0.09 (0.05)	0.130 (0.109)	100 (0)	75.0 (43.9)	2.20 (0.90)	0.85 (0.50)

Note. *% Sentences with ≥ 1 Target* shows task alignment with assigned vocabulary, while *Target Words Matched per Sentence* indicates the average number of target items used per sentence.

As shown in Table 4, modifier density (adjectives/adverbs per word) fluctuated across weeks in both groups, with modest increases in Weeks 4 and 5 for the experimental group and relatively stable values in the control group, suggesting a gradual increase in descriptive modification and elaboration (Staples et al., 2016). Task alignment with target vocabulary distinguished the groups more clearly: the experimental group consistently achieved or approached 100% of sentences containing at least one target item, whereas the control group declined from 97.5% in Week 1 to 75% in Week 6. The number of target words per sentence increased steadily in the experimental group from approximately 1.0 to 2.2, but remained around 1.0 in the control group, indicating broader and more integrated use of assigned lexis. A temporary rise in Week 4 for both groups reflects the planned recycling task for that week's vocabulary set. Independent-samples *t*-tests showed significant group differences in target-word inclusion and target words per sentence in Weeks 4–6 ($p < .001$), with moderate-to-large effect sizes.

Overall, the experimental group demonstrated more substantial and more sustained gains in sentence complexity, target-word integration, and descriptive elaboration than the control group.

Descriptive Writing Task Results

Extended descriptive writing from Weeks 4 and 7 was analyzed to examine how gains in vocabulary and sentence-level development were reflected in longer texts. Writing samples were evaluated using the analytic rubric (See Instrument and Data Collection Section), focusing on lexical sophistication, syntactic elaboration, and descriptive richness.

Descriptive Writing Development in the Experimental Group

The experimental group completed a list-based descriptive exercise in Week 4 and an image-prompt description in Week 7. Table 5 summarizes changes in writing performance.

Table 5. *Development of descriptive writing measures in the experimental group (weeks 4 and 7)*

Measure	Week 4 M (SD), N = 25	Week 7 M (SD), N = 25
Text length (words)	28.78 (7.46)	60.32 (12.89)
Target words per text	1.17 (1.15)	1.00 (1.15)
% texts ≥ 1 target word	66.67 %	52.63 %
Clause density (clauses per sentence)	2.41 (1.57)	1.59 (0.94)
Adj/Adv density	0.09 (0.06)	0.11 (0.07)
Type–Token Ratio (TTR)	0.83 (0.11)	0.70 (0.07)

Note. *Values represent group means with standard deviations in parentheses. Clause density, adjective/adverb density, and type–token ratio (TTR) are reported as standardized indices of syntactic and lexical elaboration.*

Week 7 texts were substantially longer than Week 4 texts (28.78 → 60.32 words), indicating increased output fluency. Modifier density rose slightly, suggesting greater descriptive elaboration, while clause density declined (2.41 → 1.59), a pattern associated with writers shifting from clause chaining toward denser phrasal packaging typical of descriptive genres (Biber et al., 2022). As noted earlier, this reflects a redistribution of complexity rather than a reduction. The decline in target-word frequency was expected, as Week 7 required responding to an image rather than a vocabulary list.

Qualitative patterns were consistent with these quantitative trends. Week 4 texts appeared to be concise and clause-heavy, often organized around teacher-supplied vocabulary. Sentences produced by students, such as *"After several months of struggle, he finally found a job with good wage that made him support himself and his family"* and *"He got first-hand experience of driving a sports car from his cousin who drives in annual sports event"* show reliance on subordination with limited descriptive layering. By Week 7, image-based prompts supported richer idea generation and more contextually grounded lexis. For example, a student's sentence *"This photo shows a crowd stadium with many fans watching a live football game… being part of a live event like this is the most exciting experience!"* demonstrates the emergence of evaluative and situational detail. Learners increasingly employed phrasal elaboration, as shown in the expression *"This picture shows a cozy and warm living room… The sunlight peaking through the window gives this room such a warm sense,"* where modifier layering, *"cozy," "warm," "classy"* reflects genre-appropriate concision and denser packaging (Biber et al., 2022; Staples et al., 2016).

Cohesion and sentence control also improved. Shorter declaratives connected through lexical repetition, such as *"The hospital had enough resources. The resources were very helpful…,"* show more precise segmentation and thematic continuity, consistent with developing academic writing practices (Biber et al., 2022). Although mean target-word use declined, the vocabulary that did appear was typically integrated more naturally and contextually, as exhibited in the expression: *"The image shows two girls… They are so beautiful and have a cute smile."*

Taken together, the Week 7 texts show a shift from clause-bound, list-responsive writing toward longer, visually grounded, and elaborative phrasal descriptions.

Descriptive Writing Development in the Control Group

The control group also wrote short descriptive texts in Week 4 using teacher-provided word lists. In Week 7, they completed reflective writing tasks on personal goals and routines, without visual prompts or AI-supported practice. Table 6 summarizes the results.

Table 6. *Development of descriptive writing measures in the control group (Weeks 4 and 7)*

Measure	Week 4 M (SD), N = 25	Week 7 M (SD), N = 25
Text length (words)	14.28 (2.99)	36.05 (19.54)
Target words per text	1.28 (1.02)	0.20 (0.52)
% texts ≥ 1 target word	77.78 %	15.00 %
Clause density (clauses per sentence)	2.00 (0.77)	3.90 (1.52)
Adj/Adv density	0.01 (0.02)	0.02 (0.02)
Type–Token Ratio (TTR)	0.95 (0.06)	0.83 (0.11)

Note. *Values represent group means with standard deviations in parentheses.*

The control group exhibited moderate gains in text length (14.28 → 36.05 words), indicating improved fluency. However, target-word use declined sharply (77.78% → 15%), driven partly by the shift away from list-based writing. Modifier density increased only minimally, and clause density rose sharply (2.00 → 3.90), suggesting reliance on additive clause chaining rather than controlled syntactic development. The decrease in TTR reflects longer outputs rather than expanded lexical breadth.

Qualitatively, sentences produced by students in Week 4 texts were short and formulaic, *"My brother is tall and kind"* and *"I like my city because it is big"*, typical of early-stage descriptive production. In Week 7, longer sentences were formulated but often through repeated coordination: *"Yesterday I went to the park and I saw many people and they were playing football and I was happy."* Such chaining reflects gains in fluency but limited hierarchical sentence organization (Staples et al., 2016). Lexical repetition of high-frequency adjectives (e.g., *good, nice, big*) suggests partial retention but limited semantic range, consistent with research showing that decontextualized reviews produce weak integration in extended discourse (Zhu et al., 2024).

Some students attempted extended causal linking, for example, *"I woke up late because I was tired and then I went to work and after that I slept early"*, producing longer but loosely organized narratives. The control group's Week 7 texts were longer but relied heavily on clause chaining and high-frequency adjectives.

Overall, development in the control group centered on increased output quantity rather than lexical elaboration or refined syntactic structure.

Discussion

This study investigated whether GenAI-mediated text-to-image prompting enhances vocabulary acquisition, short-term retention, and descriptive writing among Saudi foundation-year EFL students compared to traditional instruction. The results indicate that the GenAI group

achieved greater and more sustained vocabulary gains and produced writing characterized by increased syntactic elaboration and descriptive richness. These outcomes suggest that text-to-image prompting strengthens form–meaning connections and supports early academic writing development while maintaining learner authorship.

Regarding the first research question on vocabulary acquisition and short-term retention, both groups showed improvement; however, the GenAI-mediated group achieved greater gains during the instructional period and maintained these gains on the delayed test. This pattern indicates that refining prompts to generate accurate images required repeated retrieval, generative use, and fine-grained semantic differentiation. These processes are recognized for strengthening lexical representations through elaborative encoding, spaced retrieval, and meaning-oriented processing (Laufer & Girsai, 2008; Nakata & Webb, 2016; Schmitt, 2020). In contrast, the control group's reliance on translation, dictionary use, and isolated sentences provided fewer opportunities for contextualized activation, aligning with critiques of decontextualized vocabulary practice in EFL writing (Zhu et al., 2024; Richards & Rodgers, 2014).

At the sentence level, the experimental group exhibited clear syntactic progression, including higher clause density, longer, more structurally developed sentences, and greater integration of target words, while maintaining stable type–token ratios. This pattern aligns with descriptions of L2 writing development in which increased length is accompanied by more controlled structuring and integration of lexis, rather than simple linear expansion (Staples et al., 2016; Biber et al., 2022). In contrast, the control group showed only modest increases in sentence length and primarily relied on additive coordination, with a decline in target-word use. These differences reflect findings that GenAI-supported tasks enhance accuracy, fluency, and lexical integration when learners actively generate and refine their own language (Ali et al., 2025; Karataş et al., 2024; Mugableh, 2024). Unlike approaches in which ChatGPT provides or edits text (Abduljawad, 2024; Kim et al., 2025), the current design requires learners to select and adjust vocabulary independently, supporting more deliberate syntactic control and consistent integration of target lexis.

A clearer distinction between the groups emerged in extended writing, which addressed the second research question concerning lexical diversity, syntactic complexity, and descriptive richness. While both groups produced longer texts by Week 7, the experimental group generated more cohesive and descriptively focused writing, characterized by increased use of modifiers and more contextually integrated vocabulary. This transition from list-based, clause-heavy structures reflects developing control over sentence construction and lexical selection, consistent with descriptions of descriptive writing development that progress from clause chaining to denser phrasal packaging and more precise lexis (Staples et al., 2016; Biber et al., 2022). Although the control group also increased text length, their writing relied on additive coordination, exhibited higher clause density, limited descriptive modification, frequent use of high-frequency adjectives, and linear coordination. These patterns suggest gains in fluency without corresponding growth in structural control or descriptive precision. Such contrasts align with research indicating that

GenAI-mediated tasks support lexical refinement, cohesion, and clarity when learners actively review and adjust their language, whereas traditional practice tends to increase output without promoting controlled complexity in early academic writing (Kim et al., 2025; Mayer, 2014; Mahdi et al., 2024). The results suggest that iterative prompting and visual feedback encouraged the experimental group to reorganize structure and integrate vocabulary more purposefully, while the absence of such scaffolding left the control group dependent on coordination rather than on expanded lexical or structural choices.

The multimodal design of the GenAI tasks accounts for these observed differences. Text-to-image prompting required learners to express meanings linguistically, evaluate generated images, and revise prompts when the visual output did not align with their intended message. This iterative process embodies key principles of Multiliteracies Pedagogy, in which learners construct meaning across linguistic and visual modes and adjust their representations in response to feedback (New London Group, 1996; Cope & Kalantzis, 2009). The experimental group's greater gains in descriptive richness and syntactic elaboration indicate that combining visual feedback with repeated lexical and structural decisions facilitated the transition from isolated word knowledge to more integrated written expression. This finding is consistent with multimodal research linking visual–verbal integration to deeper comprehension and retention (Mayer, 2014; Mahdi et al., 2024).

The findings further highlight the role of prompt literacy in shaping the experimental group's development. Learners' consistent integration of target vocabulary and more controlled syntactic elaboration reflect engagement with the planning, monitoring, and evaluative processes identified in prompt literacy research (Tour & Zadorozhnyy, 2025; Oppenlaender, 2024). By evaluating whether generated images corresponded to their intended meanings and adjusting their wording accordingly, learners focused on communicative precision and meaning negotiation. This iterative prompting operationalized prompt literacy in practice, supporting more precise and contextually grounded descriptive writing.

This study also contributes to ongoing debates on the responsible use of AI, which emphasize concerns that excessive reliance on AI-generated text may undermine learner authorship, critical engagement, and independent linguistic decision-making (Kasneci et al., 2023; Pegrum, 2025; Balraj, 2025). Saudi studies similarly report gains achieved through ChatGPT's direct text provision or reformulation (Ali et al., 2025; Abduljawad, 2024; Mugableh, 2024). The present study addresses these concerns by restricting GenAI to image generation and requiring students to independently produce and refine all language. The findings demonstrate that GenAI can enhance vocabulary acquisition, short-term retention, and descriptive writing without replacing student output, illustrating a pedagogically responsible model for AI integration.

In summary, the findings indicate that text-to-image prompting effectively supports vocabulary learning and early academic writing in Saudi foundation-year classes while preserving learner authorship. The study further demonstrates that structured prompting tasks can be integrated as a multimodal complement to traditional instruction, particularly to reinforce

vocabulary use, syntactic elaboration, and descriptive detail without transferring writing responsibilities to AI.

Limitations and Future Directions

This study presents several limitations that future research could address. The intervention was brief and focused exclusively on descriptive writing; therefore, subsequent studies should investigate additional genres and extended instructional periods. As the study involved intact classes at a single institution, the findings may not be generalizable to other proficiency levels or educational contexts. GenAI support was restricted to text-to-image prompting using Poe and Copilot, so alternative AI tools may produce different outcomes. The investigation of responsible AI use was limited to image-based prompting rather than text-generating tools, indicating a need for comparative studies of various forms of AI mediation. Additionally, the study assessed only the written products, not the entire process of learner–AI interaction, which limits understanding of how students' prompting strategies develop over time. These limitations define the scope of the current research and suggest directions for future research.

Pedagogical Implications

The findings indicate that AI-mediated prompting can enhance classroom writing practice by enabling learners to refine meaning through intentional lexical and syntactic choices. Visual output provides immediate feedback on the accuracy of meaning, which encourages learners to engage in iterative revision. When applied to image generation rather than text production, this method preserves learner authorship and facilitates the responsible, transparent integration of AI into writing instruction.

AI-based prompting is most effective when integrated into a structured curriculum in which teachers model prompt construction, guide interpretation of visual output, and support iterative revisions. Prompt templates can assist lower-proficiency learners, whereas open-ended tasks enable advanced learners to experiment with phrasing and elaboration. Integrating AI tools with teacher and peer feedback can enhance learners' metalinguistic awareness and facilitate the transfer of revision strategies to independent academic writing.

In Saudi foundation-year contexts, where vocabulary learning frequently depends on translation-based strategies, structured prompting tasks can provide a practical bridge to more contextually grounded and independent academic writing.

Conclusion

This study examined the impact of GenAI-mediated text-to-image prompting on vocabulary acquisition, short-term retention, and descriptive writing among Saudi foundation-year EFL learners. The results demonstrated that integrating GenAI text-to-image prompting led to improved vocabulary retention, increased syntactic elaboration, and more cohesive discourse compared to traditional text-based practice. These findings suggest that iterative prompting, which

requires learners to articulate meaning, evaluate visual output, and refine language, strengthens form–meaning connections and supports essential aspects of early academic writing development.

Since students produced all written language and GenAI generated only images, learner authorship was maintained while still enabling meaningful revision. This approach exemplifies a responsible and transparent model of AI integration, positioning AI as a supportive tool rather than a replacement. Furthermore, the findings contribute to the growing body of research on AI-assisted academic writing pedagogy by demonstrating that guided prompting can promote precision, elaboration, and coherence in ways accessible to foundation-year learners.

Overall, this study provides empirical evidence that multimodal prompting tasks can enhance vocabulary learning and descriptive writing in foundation-year educational contexts. The results highlight the potential of text-to-image prompting as a practical supplement to traditional instruction and suggest promising directions for future research on learner-authored applications of GenAI.

Funding: This research is not funded.
Acknowledgments: Not applicable
Conflicts of Interest: The authors declare no conflict of interest.
Authenticity: This manuscript is an original work
Artificial Intelligence Statement: AI and AI-assisted technologies were not used to generate or rewrite the manuscript text. AI was used only as part of the instructional intervention described in the study.

About the Authors:
Dr. Heena Saifullah Amjad is an English Language Instructor in the English Language Skills Department, CFY at King Saud University, Saudi Arabia, under AlKhaleej Training and Education. She holds a PhD in English Linguistics and specializes in corpus linguistics, critical discourse analysis, and digital discourse studies. ORCID ID: https://orcid.org/0009-0007-5490-2033
Dr. Hanady Abdulaziz Alfalig is an Assistant Professor of Applied Linguistics at King Saud University. She holds a PhD in Applied Linguistics from Ball State University. Her research interests include pragmatics, sociolinguistics, and English language teaching. She has also held leadership roles and served in several committees. ORCID ID: https://orcid.org/0000-0002-4766-5451

References

Abduljawad, S. A. (2024). Investigating the Impact of ChatGPT as an AI Tool on ESL Writing: Prospects and Challenges in Saudi Arabian Higher Education. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 14*(1), 1-19. https://doi.org/10.4018/IJCALLT.367276

Aljurbua, M. F. (2021). Memory Strategies for Vocabulary Learning Employed by Saudi Undergraduate EFL Learners. (Published Master's Thesis). Retrieved from *Arab World English Journal* (ID Number: 276) August 2021: 1-100. https://dx.doi.org/10.24093/awej/th.276

Ali, Y.A., Ali, M,M., Alharbi, W., Hamid, S., & Abbas, S. (2025). The Influence of ChatGPT on English Grammar Among Saudi ESL Learners. *Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence*:358-371. https://dx.doi.org/10.24093/awej/AI.21

Balraj, B. M. (2025). Exploring the Use of ChatGPT in Academic Writing: A Systematic Literature Review on Undergraduates' Perceptions. *Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence*:348 -357. https://dx.doi.org/10.24093/awej/AI.20

Biber, D., Gray, B., Staples, S., & Egbert, J. (2022). *The register-functional approach to grammatical complexity: Theoretical foundation, descriptive research findings, application*. Routledge. https://doi.org/10.4324/9781003087991

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cope, B., & Kalantzis, M. (2009). Multiliteracies: New Literacies, New Learning. *Pedagogies: An International Journal*, *4*(3), 164–195. https://doi.org/10.1080/15544800903076044

Karataş, F. et al. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education Information Technology (29)*, 19343–19366. https://doi.org/10.1007/s10639-024-12574-6

Kasneci, E. et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies 30*, 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Laufer, B., & Girsai, N. (2008). Form-focused Instruction in Second Language Vocabulary Learning: A Case for Contrastive Analysis and Translation. *Applied Linguistics*, *29*(4), 694–716. https://doi.org/10.1093/applin/amn018

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Law, L. (2024). Application of generative artificial intelligence in language teaching and learning: A scoping review. *Computers & Education Open*, *6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Lu, X. (2010). A Corpus-based Evaluation of Syntactic Complexity Measures as Indices of College-level ESL Writers' Language Development. *TESOL Quarterly, 45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Mahdi, H. S., Mohsen, M. A., & Almanea, M. (2024). Multimedia glosses and second language vocabulary learning: A second-round meta-analysis. *Acta Psychologica, 248,* 104341. https://doi.org/10.1016/j.actpsy.2024.104341

Mayer, R. E. (Ed.). (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Mugableh, A. I. (2024). The Impact of ChatGPT on the Development of Vocabulary Knowledge of Saudi EFL Students. *Arab World English Journal (AWEJ) Special Issue on ChatGPT,* April 2024: 265-281. https://dx.doi.org/10.24093/awej/ChatGPT.18

Nakata, T., & Webb, S. (2016). Does Studying Vocabulary in Smaller Sets Increase Learning?: The Effects of Part and Whole Learning on Second Language Vocabulary Acquisition. *Studies in Second Language Acquisition, 38*(3), 523–552. https://doi.org/10.1017/S0272263115000236

Zhu, T., Zhang, Y. & Irwin, D. Second and Foreign Language Vocabulary Learning through Digital Reading: A Meta-Analysis. *Education and Information Technologies 29,* 4531–4563. https://doi.org/10.1007/s10639-023-11969-1

New London Group. (1996). A Pedagogy of Multiliteracies: Designing Social Futures. *Harvard Educational Review, 66*(1), 60–92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Oppenlaender, J. (2024). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology, 43*(15), 3763–3776. https://doi.org/10.1080/0144929X.2023.2286532

Pegrum, M. (2025). From revolution to evolution: What generative AI really means for language learning. *Language Teaching.* Published Online: 1-17 https://doi.org/10.1017/S0261444825000151

Pryma, V., Pelivan, O., Teletska, T., Tsobenko, O., & Zagrebelna, N. (2025). AI Writing Assistants and Student Competence: A Linguistic Aspect. *Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence*: 319-329. https://dx.doi.org/10.24093/awej/AI.18

Qian, Y. (2025). Prompt Engineering in Education: A Systematic Review of Approaches and Educational Applications. *Journal of Educational Computing Research, 63*(7-8), 1782-1818. https://doi.org/10.1177/07356331251365189

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and Methods in Language Teaching* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781009024532

Schmitt, N. (2020). Understanding vocabulary acquisition, instruction, and assessment: A research synthesis. *Annual Review of Applied Linguistics, 40,* 1–19. https://doi.org/10.1017/S0267190520000028

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic Writing Development at the University Level. *Written Communication, 33,* 149 - 183. https://doi.org/10.4324/9781003087991-20

Tour, E., & A. Zadorozhnyy. (2025). Conceptualizing and Operationalizing Prompt Literacy for English Language Learners. *Journal of Adolescent & Adult Literacy, 69*(3), e70020. https://doi.org/10.1002/jaal.70020

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*, (12), 621. https://doi.org/10.1057/s41599-025-04787-y

Xia, Q. et al. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education,* (21)40. https://doi.org/10.1186/s41239-024-00468-z