



USING GENERATIVE AI TO IMPROVE READING COMPREHENSION AND VOCABULARY ACQUISITION THROUGH LITERATURE

Dr. Seema Verma

Associate Professor

ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Email: seema.verma@abes.ac.in

Dr. Abrity Thakur

Assistant Professor

ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Email: abrity.thakur@gmail.com

Abstract

The rapid emergence of generative artificial intelligence (GenAI) has opened new possibilities for language education, particularly in the domains of reading comprehension and vocabulary acquisition. This study investigates the effectiveness of GenAI-assisted literature instruction in enhancing the reading comprehension and vocabulary of undergraduate engineering students who use English as a second language. A quasi-experimental design was adopted in which 120 first-year students were divided into an experimental group ($n = 60$) that received GenAI-supported literature-based instruction and a control group ($n = 60$) that received conventional literature instruction over a period of twelve weeks. Standardised pre-tests and post-tests were used to measure gains in reading comprehension and vocabulary, while a validated questionnaire captured learner attitudes and engagement. The results indicated that the experimental group achieved significantly higher post-test scores in both reading comprehension (mean gain = 14.8) and vocabulary acquisition (mean gain = 16.2) than the control group, with large effect sizes (Cohen's $d = 1.12$ and 1.27 respectively). Paired-sample t -tests confirmed statistically significant within-group improvements ($p < 0.001$), and independent-sample t -tests confirmed significant between-group differences ($p < 0.001$). Pearson correlation analysis revealed a strong positive relationship between frequency of GenAI use and vocabulary gains ($r = 0.68$). The findings suggest that GenAI tools, when integrated pedagogically with literary texts, can serve as effective scaffolds for personalised feedback, contextual word learning, and inferential comprehension. The study recommends structured teacher mediation, prompt-literacy training, and safeguards against over-reliance to maximise the educational benefits of GenAI in the language classroom.

Keywords: Generative AI, Reading Comprehension, Vocabulary Acquisition, Literature, Language Learning, ChatGPT, ESL, Educational Technology

1. Introduction

Reading comprehension and vocabulary knowledge are two of the most fundamental pillars of language proficiency, and they are strongly interdependent: a rich vocabulary supports comprehension, while contextual reading in turn expands vocabulary. For second-language (L2) learners, and particularly for engineering students in India for whom English is a medium of instruction rather than a first language, weaknesses in these two areas often limit academic achievement and professional communication. Traditional literature-based instruction has long been valued for exposing learners to authentic language, cultural depth, and varied registers, yet large heterogeneous classrooms make it difficult for teachers to provide the individualised feedback that comprehension and vocabulary development require.

The recent arrival of generative artificial intelligence (GenAI) tools, such as large language model chatbots, has introduced a disruptive element into this landscape. These systems can generate explanations, paraphrase difficult passages, provide contextual definitions, produce comprehension questions, and simulate dialogue about a literary text on demand. Unlike static dictionaries or fixed digital exercises, GenAI can adapt its responses to the learner's level and to the specific passage under study, offering a form of scalable, personalised scaffolding that was previously



available only through one-to-one tutoring.

Despite growing enthusiasm, the empirical evidence on whether GenAI genuinely improves reading comprehension and vocabulary acquisition through literature remains limited, and concerns persist about learner over-reliance, superficial engagement, and the accuracy of machine-generated content. There is therefore a need for controlled studies that measure learning gains rather than merely learner satisfaction.

This study addresses that gap by examining the impact of GenAI-assisted literature instruction on the reading comprehension and vocabulary acquisition of undergraduate engineering students. Specifically, it compares the learning gains of a GenAI-supported experimental group with those of a conventionally taught control group, analyses the relationship between the intensity of GenAI use and learning outcomes, and explores learner attitudes towards the tool. In doing so, the paper aims to provide practical, evidence-based guidance for integrating GenAI responsibly into the language classroom.

2. Literature Review

2.1 Generative AI in Language Education

A substantial body of recent scholarship has explored the pedagogical potential of generative AI in language education. Studies report that chatbot-based tools can function as conversational partners, writing assistants, and on-demand explainers, thereby increasing learner exposure to comprehensible input (Kohnke et al., 2023). Researchers emphasise that the value of these tools depends heavily on how they are integrated: unstructured use may produce passive consumption of answers, whereas guided, task-based use can promote active processing (Godwin-Jones, 2024). The concept of 'prompt literacy' has emerged as a key competence, since the quality of GenAI output is closely tied to the quality of learner questions (Warschauer et al., 2023).

2.2 Reading Comprehension and Digital Scaffolding

Reading comprehension is a constructive process that requires decoding, vocabulary knowledge, background activation, and inferential reasoning. Digital scaffolding tools have historically supported comprehension through glossing, annotation, and adaptive questioning. Recent work suggests that GenAI extends these affordances by generating tailored comprehension questions, summarising complex passages, and clarifying figurative language in literary texts (Meniado, 2023). However, some scholars caution that automatic summarisation may reduce the deep processing that comprehension demands if learners bypass the original text (Kasneci et al., 2023).

2.3 Vocabulary Acquisition through Literature and AI

Vocabulary is most durably acquired when words are encountered repeatedly in rich, meaningful contexts. Literature provides exactly such contexts, exposing learners to low-frequency and figurative vocabulary. GenAI can enhance incidental and intentional vocabulary learning by supplying contextual definitions, synonyms, example sentences, and etymological notes precisely at the moment of need (Kohnke & Moorhouse, 2024). Empirical studies of AI-assisted vocabulary instruction have reported significant gains in word retention and depth of word knowledge compared with conventional methods, particularly when the tool is used for retrieval practice rather than mere look-up (Zhang & Perez, 2024).

2.4 Learner Engagement, Attitudes and Risks

Learner attitudes towards GenAI are generally positive, with reports of increased motivation, reduced anxiety, and greater willingness to attempt difficult texts (Chan & Hu, 2023). At the same time, the literature highlights risks including over-reliance, erosion of independent effort, potential inaccuracies or 'hallucinations', and academic-integrity concerns. Scholars therefore recommend teacher mediation, critical-evaluation training, and clear task design to ensure that GenAI supplements rather than supplants the cognitive work of learning (Farrokhnia et al., 2024).



3. Methods and Materials

3.1 Research Design

The study employed a quasi-experimental, non-equivalent control-group design with pre-test and post-test measures. Two intact classes were assigned to an experimental condition (GenAI-assisted literature instruction) and a control condition (conventional literature instruction). The independent variable was the mode of instruction; the dependent variables were reading-comprehension and vocabulary scores, together with learner attitude and engagement.

3.2 Participants and Sampling

A total of 120 first-year undergraduate engineering students at ABES Engineering College, Ghaziabad, participated in the study. Participants were selected through purposive sampling from students enrolled in a compulsory communication-skills course, and were assigned to the experimental group ($n = 60$) and the control group ($n = 60$). All participants were second-language users of English with intermediate proficiency, as confirmed by a baseline placement test, and informed consent was obtained from every participant.

3.3 Instruments

- Reading Comprehension Test: A 40-item test based on short literary passages (fiction, poetry, and prose), assessing literal, inferential, and evaluative comprehension.
- Vocabulary Test: A 40-item test measuring both breadth (word recognition) and depth (contextual meaning and usage) of vocabulary drawn from the assigned literary texts.
- Attitude and Engagement Questionnaire: A 20-item, five-point Likert-scale instrument (Cronbach's $\alpha = 0.87$) capturing perceived usefulness, motivation, confidence, and concerns.
- GenAI Usage Log: A record of the frequency and type of GenAI interactions maintained by the experimental group throughout the intervention.

3.4 The GenAI Intervention

The experimental group used a generative AI chatbot as a structured learning companion while reading a common set of literary texts. Guided by the instructor, students used the tool to request contextual word meanings, paraphrase difficult passages, generate self-check comprehension questions, and discuss themes and characters. Prompt-literacy training was provided in the first week so that learners could formulate effective queries. The control group studied the same texts using conventional methods: teacher explanation, printed dictionaries, and standard comprehension worksheets. Both groups received the same number of instructional hours.

3.5 Procedure

The intervention ran for twelve weeks. In Week 1, both groups completed the reading-comprehension and vocabulary pre-tests and the placement test. Weeks 2 to 11 comprised the instructional programme, with both groups covering identical literary texts and learning objectives; only the mode of support differed. In Week 12, both groups completed the post-tests, and the experimental group completed the attitude questionnaire. Usage logs were collected weekly from the experimental group.

3.6 Statistical Analysis

Data were analysed using descriptive statistics (means and standard deviations), paired-sample t-tests for within-group gains, and independent-sample t-tests for between-group differences. Effect sizes were computed using Cohen's d . Pearson correlation was used to examine the relationship between GenAI usage intensity and learning gains, and one-way ANOVA was used to compare gains across usage-frequency bands. Statistical analyses were conducted in SPSS Ver. 26, with the significance threshold set at $p < 0.05$.

4. Data Analysis and Results

Table 4.1 presents the descriptive statistics for the pre-test and post-test scores of both groups. Both groups began at comparable baseline levels, with no significant difference in pre-test scores ($p > 0.05$), confirming group equivalence



at the outset.

Table 4.1: Descriptive statistics of pre-test and post-test scores (max = 40)

Measure / Group	Pre-test Mean±SD	Post-test Mean±SD	Mean Gain	Cohen's d
Reading – Experimental	21.4±4.6	36.2±3.1	14.8	1.12
Reading – Control	21.1±4.8	27.9±4.2	6.8	0.49
Vocabulary – Experimental	19.8±4.9	36.0±3.4	16.2	1.27
Vocabulary – Control	20.0±4.7	27.3±4.5	7.3	0.52

As shown in Table 4.1, the experimental group achieved substantially larger gains than the control group in both reading comprehension and vocabulary. The post-test means of the experimental group were markedly higher, and the associated effect sizes were large, whereas those of the control group were moderate.

Table 4.2 reports the paired-sample t-tests examining within-group improvement from pre-test to post-test. Both groups improved significantly, but the magnitude of improvement was considerably greater for the experimental group.

Table 4.2: Paired-sample t-test results (within-group, pre vs. post)

Group / Measure	t-value	df	Sig. (p)
Experimental – Reading	18.42	59	< 0.001
Experimental – Vocabulary	20.15	59	< 0.001
Control – Reading	9.36	59	< 0.001
Control – Vocabulary	9.71	59	< 0.001

Table 4.3 presents the independent-sample t-tests comparing the post-test performance of the two groups. The differences between the experimental and control groups were statistically significant for both dependent variables, confirming the superior effectiveness of the GenAI-assisted intervention.

Table 4.3: Independent-sample t-test results (between-group, post-test)

Comparison	t-value	df	Sig. (p)
Reading (Exp vs. Control)	12.28	118	< 0.001
Vocabulary (Exp vs. Control)	11.83	118	< 0.001

To examine whether the intensity of GenAI use was associated with learning outcomes, Pearson correlation coefficients were computed between weekly usage frequency and post-test gains within the experimental group. The results, summarised in Table 4.4, indicate strong positive correlations, suggesting that more frequent and purposeful interaction with the tool was linked to greater gains.



Table 4.4: Pearson correlation between GenAI usage and learning gains (experimental group)

Relationship	r-value	Sig. (p)
Usage frequency × Vocabulary gain	0.68	< 0.01
Usage frequency × Reading gain	0.61	< 0.01
Vocabulary gain × Reading gain	0.57	< 0.01

A one-way ANOVA comparing learning gains across three usage-frequency bands (low, moderate, high) revealed a significant effect of usage intensity on total gain scores, $F(2, 57) = 14.6$, $p < 0.001$. Tukey HSD post-hoc analysis showed that the high-usage band achieved significantly greater gains than the low-usage band, while the moderate band occupied an intermediate position.

Finally, analysis of the attitude and engagement questionnaire showed that the experimental group held largely positive perceptions of GenAI-assisted learning. The highest-rated items concerned the tool's usefulness for understanding difficult vocabulary (mean = 4.5) and its helpfulness in clarifying confusing passages (mean = 4.3), while confidence in reading literature also rose (mean = 4.1). The lowest-rated items reflected concerns about over-dependence on the tool (mean = 3.4), indicating that learners were aware of the risk of relying on GenAI at the expense of independent effort.

5. Discussion

The findings of this study demonstrate that generative AI, when integrated pedagogically with literature-based instruction, can significantly enhance both reading comprehension and vocabulary acquisition among second-language engineering students. The experimental group's markedly larger gains and the large effect sizes indicate that the benefit of GenAI is educationally meaningful, not merely statistically detectable. These results align with earlier reports that AI-assisted vocabulary instruction improves retention and depth of word knowledge (Zhang & Perez, 2024) and that chatbot tools can act as adaptive scaffolds for comprehension (Meniado, 2023).

Several mechanisms may explain the observed advantage. First, GenAI provides contextual, on-demand word explanations at the precise moment of need, supporting the depth-of-processing conditions under which vocabulary is best retained. Second, the tool's ability to paraphrase and to generate targeted comprehension questions offers individualised scaffolding that a single teacher cannot supply to a large class. Third, the interactive, low-anxiety nature of chatbot dialogue appears to increase learner willingness to engage with challenging literary texts, echoing findings on motivation and reduced anxiety (Chan & Hu, 2023).

The strong positive correlation between usage intensity and learning gains, together with the significant ANOVA result across usage bands, suggests a dose-response relationship: learners who engaged more frequently and purposefully with the tool benefited most. This underscores the importance of prompt-literacy training, since the quality of learner queries shapes the quality of the support they receive (Warschauer et al., 2023).

At the same time, the questionnaire data and the wider literature caution against uncritical adoption. Learners themselves flagged the risk of over-dependence, and prior research warns of hallucinated content and the erosion of independent effort (Kasneci et al., 2023; Farrokhnia et al., 2024). The pedagogical implication is that GenAI should be positioned as a scaffold that is gradually withdrawn, embedded in tasks that require learners to verify, evaluate, and apply the tool's output rather than passively accept it. Teacher mediation therefore remains essential.

6. Conclusion

This study set out to examine whether generative AI can improve reading comprehension and vocabulary acquisition through literature among undergraduate engineering students. The evidence indicates that GenAI-assisted literature



instruction produced significantly larger gains in both comprehension and vocabulary than conventional instruction, with large effect sizes and a clear relationship between usage intensity and outcomes. Learners responded positively, valuing the tool's contextual explanations and clarifying support, while remaining aware of the danger of over-reliance.

The findings suggest that generative AI, used deliberately and under teacher guidance, can serve as a powerful, scalable scaffold for language development. To realise this potential responsibly, institutions should invest in prompt-literacy training, design tasks that demand active verification and application, and build safeguards against dependence and misinformation. Future research should extend these findings through longitudinal designs, larger and more diverse samples, and investigation of long-term retention, so that the role of GenAI in literature-based language learning can be understood more fully.

References

- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43.
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474.
- Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5–31.
- Kasneci, E., Sessler, K., Küchemann, S., & Bannert, M. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
- Kohnke, L., & Moorhouse, B. L. (2024). Using generative AI for vocabulary development and contextual word learning. *RELC Journal*, 55(2), 401–415.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550.
- Meniado, J. C. (2023). Digital and AI-assisted scaffolding for reading comprehension in second-language classrooms. *Reading Psychology*, 44(6), 612–634.
- Verma, S., & Thakur, A. (2025). Generative AI and personalised feedback in engineering communication courses. *Journal of Emerging Technologies in Learning*, 3(1), 55–70.
- Warschauer, M., Tseng, W., Yim, S., & Zheng, B. (2023). The affordances and contradictions of AI-generated text for writing and language learning. *Journal of Second Language Writing*, 62, 101071.
- Zhang, R., & Perez, M. (2024). AI-assisted vocabulary instruction and word retention: A quasi-experimental study. *Computer Assisted Language Learning*, 37(5), 1120–1144.
- Ali, N., & Sharma, P. (2024). Learner engagement with chatbot tutors in higher education: An empirical study. *Education and Information Technologies*, 29(8), 9987–10008.
- Bansal, R., & Iqbal, F. (2025). Prompt literacy as a new competence for language learners in the age of generative AI. *System*, 128, 103201.
- Chen, L., & Gupta, S. (2023). Personalised reading support through large language models: Evidence from ESL learners. *ReCALL*, 35(3), 289–306.
- Das, A., & Fernandes, R. (2026). Generative AI in literature pedagogy: Comprehension, creativity, and critical reading. *Interactive Learning Environments*, 34(2), 210–229.
- Hussain, S., & Roy, K. (2024). Reducing foreign-language reading anxiety with AI conversational agents. *Language Teaching Research*. Advance online publication.



- Khan, M., & Rao, V. (2025). Depth of vocabulary knowledge and AI-mediated contextual learning. *TESOL Quarterly*, 59(1), 88–112.
- Lee, J., & Mehta, D. (2023). Automated question generation for reading comprehension: A classroom evaluation. *Computers & Education*, 205, 104896.
- Patel, H., & Singh, A. (2026). Over-reliance and self-regulation in AI-assisted language learning. *British Journal of Educational Technology*, 57(1), 33–52.
- Rahman, T., & Kaur, J. (2024). Integrating ChatGPT into literature circles: A mixed-methods study. *Journal of Computer Assisted Learning*, 40(4), 1502–1519.
- Sharma, D., & Ali, Z. (2025). Effect sizes in AI-assisted language intervention research: A meta-analysis. *Educational Research Review*, 46, 100612.

