
THE EFFECT OF AI PORTRAIT PROMPT WRITING ON VOCABULARY DEVELOPMENT OF SECONDARY ESL LEARNERS

Dr. Jam Khan Muhammad

Assistant professor, Centre of English language and Linguistic (CELL) Mehran university of Engineering and Technology, Jamshoro

Syed Khuram Shahzad*

PhD Scholar in English Linguistics, Institute of English Language and Literature (IELL) University of Sindh, Jamshoro

*Corresponding Author

Abstract

English vocabulary knowledge is considered as the prerequisite for second language competence and in Pakistan the instruction of English vocabulary in public secondary schools is mostly text book-based, formal and rote learning which results in superficial and superficial language knowledge in second language learners (ESL). Generative text-to-image AI creates a novel pedagogical task, AI portrait prompt writing, where learners create text-based prompts to be translated into visual portraits, potentially involving both comprehensible input, pushed output, and tool-mediated scaffolding, and dual verbal-visual coding. The aim of the present study was to examine the impact of using AI-generated writing prompts for portraits on the acquisition of the vocabulary of secondary English learners and its comparative effectiveness with traditional vocabulary teaching. A true experimental pre test – post test control group design was used in which the experimental group ($n = 30$) and the control group ($n = 30$) Grade 9 learners were randomly selected from a government secondary school in Tehsil Sadiqabad, District Rahim Yar Khan. The experimental group was taught for 8 weeks using AI portrait prompt writing whereas the same content was taught to the control group using traditional ways. The researcher derived 40-item Vocabulary Achievement Test (VAT) was used to assess vocabulary development with the test administered both before and after intervention. The two groups were statistically equivalent at pretest, $t(58) = 0.162$, $p > .05$, but significantly different by the posttest, $t(58) = 7.3$, $p < .001$, with a large effect size (Cohen's $d \approx 1.9$). The results show that AI-generated portrait prompt writing results in a substantial increase in vocabulary compared to traditional methods, suggesting its viability for use in ESL vocabulary instruction in disadvantaged public schools.

Keywords: AI Portrait Prompt Writing; Vocabulary Development; Computer-Assisted Language Learning (CALL); True Experimental Design; Secondary ESL Learners

Introduction

English is extremely important not only for its role as a subject, but also as a medium of instruction in higher education and professional education in Pakistan. Vocabulary knowledge is known to be a central catalyst of language skills in L2 listening, reading, speaking and writing (Nation, 2001; Schmitt, 2000). Vocabulary instruction in public secondary schools of Pakistan, however, largely continues to be based on vocabulary memorization, bilingual lists of words, and isolated and recyclable exercises that tend to fix words in short-term memory, if they are remembered at all, instead of enabling meaningful use of words (Schmitt, 2008). Language acquisition theories offer a better basis for good second language learning of vocabulary. According to Krashen's (1985) Input Hypothesis, students learn best by listening to comprehensible input, and according to Swain's (1985) Comprehensible Output Hypothesis, students enhance their vocabulary knowledge by producing comprehensible output. Likewise, the Sociocultural Theory by Vygotsky (1978) emphasizes the need for guided learning in the ZPD, and the Dual Coding Theory by Paivio (1971, 1986) indicates that vocabulary retention is improved when both verbal and visual information are provided.

AI portrait prompt writing is a new method for mastering vocabulary using generative AI, specifically text-to-image models. Learners need to use descriptive vocabulary with accuracy to create accurate images, relating to appearance, feelings and setting. The produced image instantly gives learners visual feedback and allows them to review and edit words by practicing repeatedly. This process is interactive, involving language production, visual learning and

ongoing revision, and seems to be an effective alternative to vocabulary instruction. The literature of Computer-Assisted Language Learning (CALL) has shown the advantages of technology-enhanced language learning (TELL) (Godwin-Jones, 2022; Golonka et al., 2014), and more recent research has provided insights into the possibilities of generative AI in personalized language learning. However, there is a limited empirical study on the use of AI prompt writing for vocabulary learning in true experimental studies with secondary school English second language learners in Pakistan's public schools. This is an important research gap which is addressed in the present study.

Statement of the Problem

In spite of years of formal English teaching, the ESL students of secondary schools in Tehsil Sadiqabad, District Rahim Yar Khan are still showing low levels of productive vocabulary, especially in the domains of description (physical appearance, emotions, spatial and lighting description). In these schools, the conventional approach to vocabulary has largely been a textbook-centred, teacher-led and translation-based method with little room for meaningful productive use of new lexical items (Schmitt, 2008; Warschauer & Healey, 1998). Meanwhile, generative AI tools that can convert a written prompt into a visual portrait are becoming widely available, with no solid experimental evidence to date suggesting that using a structured AI-guided approach for writing prompts leads to a statistically significant gain in vocabulary development when compared to traditional methods of writing instruction for this population of learners. This study thus aims at gap filling the lack of an empirical study regarding the effectiveness of AI prompt writing for vocabulary development for secondary ESL learners.

Research Objectives

1. To investigate the effect of AI portrait prompt writing on the vocabulary development of secondary ESL learners.
2. To compare the vocabulary achievement of secondary ESL learners taught through AI portrait prompt writing with those taught through conventional vocabulary instruction.

Research Questions

1. What is the effect of AI portrait prompt writing on the vocabulary development of secondary ESL learners?
2. Is there a significant difference in vocabulary achievement between secondary ESL learners taught through AI portrait prompt writing and those taught through conventional vocabulary instruction?

Hypotheses

H0₁: There is no statistically significant difference between the pretest and posttest vocabulary scores of the experimental group.

H1₁: There is a statistically significant difference between the pretest and posttest vocabulary scores of the experimental group.

H0₂: There is no statistically significant difference in posttest vocabulary scores between the experimental group and the control group.

H1₂: There is a statistically significant difference in posttest vocabulary scores between the experimental group and the control group, favoring the experimental group.

Significance of the Study

This study is expected to contribute both theoretically and practically to the field of Computer-Assisted Language Learning. Theoretically, it extends the application of Krashen's (1985) Input Hypothesis, Swain's (1985) Output Hypothesis, Vygotsky's (1978) sociocultural theory, and Paivio's (1971) Dual Coding Theory to a novel instructional task, namely AI-mediated prompt writing, thereby offering an empirical test of these frameworks in a generative-AI

context that did not exist at the time the theories were formulated. Practically, the findings are expected to inform ESL teachers, curriculum developers, and education policymakers in Pakistan's public secondary school system regarding the feasibility and effectiveness of integrating generative AI tools into vocabulary instruction, particularly in under-resourced rural settings such as Tehsil Sadiqabad. The study may also serve as an evidence base for teacher training initiatives and for future large-scale adoption decisions regarding AI-based language learning tools within government schools.

Delimitations of the Study

- The study is delimited to government secondary schools located in Tehsil Sadiqabad, District Rahim Yar Khan, Punjab, Pakistan.
- The intervention is delimited to an eight-week instructional period.
- Vocabulary development is measured using a 40-item parallel-form researcher-developed vocabulary test comprising matching, multiple choice, fill-in-the-blank, sentence completion, and picture labeling sections.
- The vocabulary domains under investigation are delimited to appearance adjectives, emotion and expression vocabulary, and spatial and lighting language, as these are the domains most directly activated by AI portrait prompt writing tasks.

Definition of Key Terms

AI Portrait Prompt Writing: The instructional task in which learners compose written descriptive prompts, using targeted vocabulary, which is submitted to a generative text-to-image AI system to produce a corresponding visual portrait.

Vocabulary Development: A measurable, positive change in a learner's receptive and productive knowledge of targeted lexical items, as evidenced by improved performance from pretest to posttest on a standardized vocabulary instrument (Nation, 2001).

True Experimental Design: A research design characterized by random assignment of participants to experimental and control groups, together with pretest and posttest measurement, enabling causal inference regarding the effect of an intervention.

Secondary ESL Learners: Students enrolled at the secondary level (grades 9–10) in government schools, for which English is learned as a second or foreign language rather than a first language.

Computer-Assisted Language Learning (CALL): An approach to language education in which computer or digital technology is used to support and enhance language teaching and learning processes (Warschauer & Healey, 1998).

REVIEW OF RELATED LITERATURE

Introduction

This chapter reviews the theoretical and empirical literature relevant to the effect of AI portrait prompt writing on the vocabulary development of secondary ESL learners. The chapter is organized into eight sections. It begins with the theoretical framework underpinning the study, drawing on second language acquisition and cognitive learning theory. It then reviews literature on vocabulary acquisition in ESL contexts, the trajectory of Computer-Assisted Language Learning (CALL) research, and the more recent body of scholarship on generative artificial intelligence in language education. The chapter subsequently examines multimodal and image-based approaches to vocabulary instruction, situates the study within the ESL education context of Pakistan, reviews prior true experimental studies of a comparable design, and concludes by identifying the specific research gap the present study addresses and presenting the conceptual framework guiding the investigation.

Theoretical Framework

Vygotsky's Sociocultural Theory and the Zone of Proximal Development

According to Vygotsky (1978), the sociocultural theory, learning is considered to be primarily

social and mediated; it takes place best when a learner is in his or her Zone of Proximal Development (ZPD), the space that exists between what he or she can already do and what he or she cannot do independently. Technology-mediated learning can be a mediating artifact in the ZPD (Lantolf & Thorne, 2006). The present study is not just a presentation of information, however, but a dynamic interaction with the learner's own descriptive language, which provides a scaffolded space for experimentation with words and for the learner to see the immediate consequences of their word choice.

2Paivio's Dual Coding Theory

The Dual Coding Theory proposed by Paivio (1971, 1986) states that humans learn through two channels: verbal and nonverbal (visual/imagery), and that information presented in both is remembered for a longer time than information presented in either one channel alone. The empirical research on vocabulary instruction has shown that verbal imagery is more effective at retaining vocabulary than verbal instruction alone (Sadoski & Paivio, 2013). Rather than having the teacher provide the word-image pairing, AI portrait prompt writing is a learner-generated pair, with the text description being generated by the learner themselves and the image being generated by AI: This could increase the word-image association beyond what can be achieved when the learner is just given a picture and a pre-selected word to pair with it.

Vocabulary Acquisition in ESL Learning

It has long been known that vocabulary is a key component of L2 proficiency (Nation, 2001; Schmitt, 2000). It is possible to be able to recognize and comprehend a word but not know them well enough to use it correctly in speech and writing, or to recall it, and those two skills are known as receptive vocabulary knowledge versus productive vocabulary knowledge, respectively. Nation (2001) notes that productive knowledge is usually more intensive, repeated, and meaningful in terms of the involvement with the lexical item but also develops later. The Involvement Load Hypothesis (ILH) of Laufer and Hulstijn (2001) also suggests that the retention of vocabulary is directly proportional to the extent of the cognitive involvement that a task elicits; involving tasks will result in better retention than less involving tasks, like simple word-list memorization. Descriptive prompt writing is likely to involve a high load of involvement because learners have to compare a series of candidate words to the communicative task that they are tasked with performing in this case, creating an image to match the communicative intention.

The shortcomings of traditional ESL vocabulary teaching in the context of public schools in Pakistan, where the dominant approach to language teaching is grammar translation based, textbook-oriented and memorization of bilingual word lists (Warsi, 2004; Coleman, 2010), have been well documented. These have been linked to low level productive vocabulary and low levels of motivation in the learner, especially in secondary schools where the demands of the curriculum rise, yet the pedagogical methods are not that different from what was used from the previous grades.

Computer-Assisted Language Learning and Vocabulary Development

The CALL research has passed through several conceptual stages: the first was represented by the use of the drill and practice software; the second was the communicative stage that focuses on meaningful interaction and dialogue; and the third was referred to as integrative and sociocognitive CALL in which technology is used as part of real communicative tasks, not a vehicle for presenting isolated exercises. Levy's (1997) definition still remains a basic definition in the field of CALL, although the range of tools under this definition has increased significantly, for instance with the arrival of mobile-assisted language learning (Kukulska-Hulme, 2012), and, more recently, generative AI.

A meta-analysis study of Golonka et al. (2014) analyzed a wide variety of CALL technologies and concluded that while the effects are generally positive, the effectiveness is dependent upon

the pedagogical integration of the technology and not necessarily the technology itself, as seen in the case of language outcomes such as vocabulary. This result is in line with a vast amount of vocabulary-specific CALL research (Nation & Newton, 2009) that suggests digital media that demand active, generative learning by the students are more likely to lead to better vocabulary learning, as opposed to passive exposure.

Generative Artificial Intelligence and Language Learning

With the swift development of the generative AI systems like large language models and the text-to-image systems, the research interests of CALL are revitalized. Godwin-Jones (2022) sees generative AI writing assistance as a qualitative change in CALL, because it can generate contextually appropriate and individualized input and feedback at a scale and speed that was not possible with the previous generation of CALL software. Kohnke, Zou, and Chen (2023) also explore how generative AI chatbots might be used to facilitate second language vocabulary and writing practice by enabling an interactive and individualized approach, highlighting a lack of learning outcome research evidence, but not the learner's perceptions. Zhang & Zou (2022) reviewed the existing studies about the Vocabulary Learning Tools under the aid of AI, and reported the overall positive results on the retention of vocabulary. However, the majority of available studies used quasi-experimental designs instead of true experimental designs, which restricts the strength of causality that can be inferred.

In this nascent body of work, generative text-to-image AI systems are relatively under-studied. Previous literature on language learning and image-generation AI has tended to emphasize the use of the technology in visual arts and design, instead of language-specific goals like increasing vocabulary acquisition (Godwin-Jones, 2022). The instructional task at the center of the present study, in which learners write sentences that prompt an AI system to generate a portrait, lies at the nexus of two fields of research in the field of language learning—the study of generative AI scholarship and multimodal vocabulary pedagogy—that have never been studied in tandem.

Multimodal and Image-Based Approaches to Vocabulary Instruction

The pedagogical use of image based vocabulary instruction is fairly long-standing, even before the emergence of digital technology such as picture dictionaries and flashcards, and has always been supported by the dual coding research (Sadoski & Paivio, 2013). This tradition has been continued by digital multimodal tools, such as picture-supported CALL software and more recently by augmented and virtual reality applications, which, as a rule, report positive effects on the retention of the target words when they are presented in a meaningful way in the context of a picture (Yang & Chen, 2023). The difference between receptive multimodal tasks (where the learners are given an image and word) and generative multimodal tasks (where the learners create both the image and the word) is highlighted as an important difference between pedagogies in this literature. The latter is more consistent with the Involvement Load Hypothesis (Laufer & Hulstijn, 2001) and also follows Swain's (1985) focus on output because the learner is expected to involve himself in the act of building the word-image association instead of simply being exposed to it. It is a generative multimodal task in this sense, and existing theory would suggest that it would lead to better vocabulary retention than traditional picture-vocabulary matching tasks, but there has not been a prior experimental test.

Research Gap

Three strands of theoretical support emerge from the reviewed literature, which developed from the input, output, sociocultural, and dual coding theories: (1) the expectation that a generative, image-producing AI task should support vocabulary development, (2) the expectation that the task will support students' learning by having them compare and contrast the prompts, and (3) the expectation that the task will support vocabulary development by providing students with a second language representation. It also concludes that, as of now, CALL and generative AI

vocabulary research has focused primarily on quasi-experimental research methods and seldom has included text-to-image prompt writing as a vocabulary intervention, rather than an AI chatbot-based writing intervention. In addition, there are no reported studies that have investigated this task in the context of the Pakistani government secondary schools in which research on generative AI education has been conducted, where there are significant differences in resource availability and culturally shared norms of instruction. The aim of this present true experimental study is to empirically investigate the effect of eight weeks of AI portrait prompt writing intervention on the achievement of secondary ESL learner's vocabulary development with the help of the random assignment approach and pretest-posttest design in Tehsil Sadiqabad of Rahim Yar Khan District..

Research Approach

The study is based on quantitative research, which is in line to the positivistic approach of the study. Quantitative research is defined by the systematic gathering of quantitative (numerical) data, the use of structured and standardized tools, and the application of statistical techniques to test the hypotheses and to determine generalizable patterns. The effectiveness of the AI portrait prompt writing as a vocabulary-development intervention is assessed through the numerical test scores of the experimental and control groups pre- and post-test scores. A quantitative approach was chosen over a qualitative or mixed methods approach because if the instructional method A produces a statistically significant increase in vocabulary scores over instructional method B, and if the question is most effectively and rigorously addressed by controlled measurement and inferential statistical testing, rather than descriptive or interpretive methods, then the quantitative approach is most appropriate.

Research Design

For establishing causal inference, the study uses True Experimental Pretest-Posttest Control Group Design which is one of the most methodologically rigorous designs in educational research. The design of this project is a randomized design consisting of two groups: experimental and control. Both groups are statistically similar with respect to measured and unmeasured characteristics before the intervention begins. Both groups undergo the same vocabulary pretest at the outset of the intervention, to detect any differences between the groups and to statistically control for these differences. The experimental group is taught for the next eight weeks using AI portrait prompt writing, and the control group is taught by the traditional method of using textbooks for vocabulary instruction, but the same lexical content is taught. Both groups take an identical vocabulary post test at the end of the intervention, which the researcher can use to compare the performance of the two groups and to isolate the effect that can be attributed to the writing intervention using the AI prompt for writing the portrait. The logic of the design is conventionally represented using the notation shown in the table below, where O denotes an observation (that is, a test administration) and X denotes the experimental treatment:

Group	Pretest	Treatment	Posttest
Experimental	O ₁	X	O ₂
Control	O ₃	—	O ₄

Here, O₁ represents the pretest score of the experimental group, O₂ represents its posttest score, O₃ and O₄ represent the corresponding pretest and posttest scores of the control

group, and X represents the AI portrait prompt writing intervention administered exclusively to the experimental group. Random assignment to groups, combined with the pretest-posttest structure, allows the researcher to control for maturation, history, testing effects, instrumentation, and regression to the mean, which are among the most common threats to internal validity in educational intervention research.

Variables of the Study

Every experimental study is organized around the relationship between an independent variable, which the researcher manipulates, and a dependent variable, which the researcher measures in order to detect the effect of that manipulation. In the present study, the independent variable is AI portrait prompt writing, the pedagogical intervention through which experimental-group learners are taught to compose, refine, and use descriptive prompts to generate AI portrait images as a vehicle for vocabulary acquisition. The dependent variable is vocabulary development, operationalized as the learners' performance on the researcher-developed Vocabulary Achievement Test, administered as a pretest before the intervention and as a posttest after it. The study is designed so that any statistically significant difference in vocabulary development between the experimental and control groups, over and above their baseline pretest performance, can reasonably be attributed to the manipulation of the independent variable.

Population of the Study

The population of the study comprises all Grade 9 secondary-level ESL learners enrolled in government secondary schools in Tehsil Sadiqabad, District Rahim Yar Khan, Punjab, Pakistan. These learners follow the Punjab curriculum, under which English is studied as a compulsory subject from an early grade through to secondary level, making Grade 9 learners an appropriate population for a study concerned with vocabulary development in English as a second language. The choice of this population is also grounded in practical and contextual considerations: government secondary schools in this tehsil are broadly representative, in terms of resources, class size, and instructional practice, of public-sector secondary education across rural and semi-urban Punjab, which enhances the ecological validity of the findings and their relevance to policy and pedagogy within the wider public education system.

Sample of the Study

This population is randomly sampled to yield a sample of sixty Grade 9 students, half of which are assigned to the experimental group (30 students) and the other half to the control group (30 students). In the methodological literature on experimental and quasi-experimental research, a sample size of $n=30$ is a reasonable minimum sample size for the meaningful use of parametric inferential statistics (independent samples t-test, paired samples t-test) because the sample size required for the sampling distribution of the mean to approximate the normal distribution, by the central limit theorem, is close to $n=30$ even if the underlying distribution in the population is not perfectly normal. Equal group sizes also make it easier to interpret between-group comparisons and minimize the potential for unequal variance to affect the analysis of the t-tests.

Sampling Technique

A multistage sampling technique is employed to select the final sample from the wider population, proceeding through three distinct stages. In the first stage, one government secondary school located in Tehsil Sadiqabad is selected from among the eligible schools using simple random sampling, ensuring that every school in the population has an equal probability of selection and thereby minimizing selection bias at the level of the institution. In the second stage, two intact Grade 9 sections within the selected school, matched as closely as possible in terms of prior academic performance and general English proficiency, are identified for

participation in the study; using intact sections rather than reconstituting entirely new groups reflects the practical realities of conducting research within an operating school timetable while still preserving comparability between groups. In the third stage, students within these two sections are randomly assigned to the experimental and control groups using simple random sampling, for example through a lottery or random-number procedure, so that assignment to treatment condition is determined by chance rather than by any characteristic of the student, the teacher, or the researcher. This final stage of random assignment is the defining feature of a true experimental design and is what distinguishes the present study from a quasi-experimental design employing intact, non-randomized groups; it is this random assignment that most effectively equalizes the two groups on both known and unknown confounding variables prior to the intervention.

Participants

Participants in the study are Grade 9 ESL learners between fourteen and sixteen years of age, drawn from sections with a broadly similar level of English proficiency and a broadly similar educational background, and who maintain regular attendance at school throughout the period of the study. Restricting participation to students within this narrow age band and proficiency range reduces the influence of extraneous variables such as differing cognitive maturity or markedly different levels of prior English exposure, both of which could otherwise confound the relationship between the intervention and vocabulary gain. The requirement of regular attendance is likewise important, since the eight-week intervention is cumulative in nature, and irregular attendance would expose a participant to only a partial version of the treatment, thereby weakening the internal validity of the comparison between groups.

Research Instrument

The researcher developed instrument, the Vocabulary Achievement Test, is used in this study. The instrument is administered in the same way on the pretest and posttest sessions so that gain scores can be validly compared between the two test sessions. The test is structured into 5 sections with each targeting a different aspect of vocabulary knowledge: multiple choice questions (MCQ), matching, fill in the blanks, vocabulary in context and AI portrait prompt writing, where the latter is the focus of the experimental intervention. The instrument consists of five sections for a total of fifty marks and is administered in a 50 minute testing session which is deemed to be of a realistic time to allow Grade 9 learners to complete all five sections without excessive time pressure but be short enough to fit into one class.

Validity of the Instrument

Face Validity

The face validity of the instrument, that is, the extent to which it appears, on inspection, to measure what it purports to measure, was established through expert review. The draft instrument was examined by a panel comprising three ESL experts, two practicing English teachers, and one educational assessment specialist, each of whom evaluated the clarity, wording, grammatical accuracy, and appropriateness of item difficulty for Grade 9 learners. Feedback from this panel was systematically incorporated into a revised version of the instrument, resulting in improvements to item wording, the elimination of ambiguous distractors, and adjustments to ensure that the level of difficulty was appropriate for the target age group and proficiency level.

Content Validity

The content validity of the instrument, that is, the extent to which its items adequately sample the domain of vocabulary knowledge that the study intends to measure, was established through the development of a Table of Specifications (TOS). This table maps each item and section of the test against the cognitive levels of Bloom's Revised Taxonomy, ranging from remembering and understanding through to applying and analyzing, against the vocabulary-related learning

objectives specified in the secondary English curriculum. The use of a Table of Specifications ensured that the instrument was not skewed toward any single cognitive level or vocabulary domain, but instead represented a balanced and curriculum-aligned sample of the vocabulary knowledge and skills that the intervention was designed to develop.

Pilot Study

Prior to full-scale administration, the Vocabulary Achievement Test was piloted with twenty Grade 9 students drawn from a government secondary school other than the one selected for the main study, in order to prevent any contamination between the pilot sample and the actual study participants. The pilot study served four principal objectives: to determine the clarity of the instructions and items as perceived by learners at the target level, to estimate the actual time required for completion under realistic testing conditions, to identify any items that were ambiguous, confusing, or subject to multiple defensible interpretations, and to generate a preliminary dataset from which the reliability of the instrument could be statistically calculated. Insights gained from the pilot study informed minor revisions to item wording and confirmed that the fifty-minute administration time was appropriate for the instrument in its final form.

Reliability of the Instrument

The reliability of Vocabulary Achievement Test (VAT) was estimated by Cronbach's Alpha which is the most commonly used index of reliability for a multi-item test that includes multiple types of items. The coefficient score of 0.80 or higher was determined to be acceptable based on the conventional practice in educational and psychological measurement, where scores of 0.70 or higher are considered acceptable, and 0.80 or higher are considered good for instruments designed for research purposes. Once this has been established then it can be confidently stated that the test scores derived from the instrument in the main study are reliable and consistent and that any differences observed between the experimental and control group would be due to the intervention and not measurement error associated with an unreliable instrument.

Treatment (Intervention)

The pedagogical intervention central to this study extends over a period of eight weeks, during which the experimental and control groups receive parallel but distinct forms of vocabulary instruction covering the same lexical content, so that any difference in outcome can be attributed to instructional method rather than to differences in content coverage.

Experimental Group Activities

The students in the experimental group are provided with instruction on vocabulary acquisition using the technique of writing short, structured prompts of written descriptions to feed into an AI image-generation platform to produce an image that matches the description. The activities used in this treatment will provide students with progressively greater opportunities to develop receptive and productive and contextualized vocabulary acquisition. Each unit starts with learners writing draft portrait prompts, describing human appearance and using the vocabulary that has been introduced; they are encouraged to think carefully about the descriptive adjectives they choose to use to describe facial features, clothes, colour and expression, becoming more and more precise. Learners then type these prompts into an AI image creation tool and compare the image that is generated with the one they meant to describe, thus getting almost instant feedback on the effectiveness of their vocabulary in communicating the intended meaning, in a visual and highly motivating way. A mismatch of the intended description with the created image leads the learners to make adjustments to and edits of their prompts, deliberately using more precise and/or more suitable language, which goes through iterative adjustments and improvement to both word meanings and word choice. Within this cycle, structured synonym practice is integrated to encourage learners to go beyond the first word that springs to mind to increase their use of a variety of adjectives and descriptive phrases. Peer discussion sessions

are conducted, where the learner explains his prompts and then his images to other learners, discusses with other learners other words that could be used, and receives informal feedback from other learners on their images, which helps reinforce the vocabulary through social negotiation of meaning. Finally, the classroom teacher delivers specific feedback on each learners' prompts, focusing on particularly effective vocabulary choices and offering suggestions for improvement; in this way, the intervention does not consist solely of the AI tool, but also includes a strong component of expert pedagogical guidance.

Control Group Activities

The students in the control group are taught the same set of vocabulary items and vocabulary domains as the students in the experimental group, but receive conventional (non-technology-mediated) instruction in the same vocabulary items and the same vocabulary domains using textbook exercises that require the learners to carry out standard comprehension tasks and complete the standard vocabulary exercises in the prescribed textbook, teacher explanation of target vocabulary items and their meaning in the target language, memorization tasks involving word lists and their meaning, translation exercises to render the English vocabulary items into their first language and vice versa, dictionary work that involves students looking up words they are unfamiliar with and jotting down their meaning, and structured worksheets that provide additional practice with the target vocabulary items and their meaning. The control group is not asked to use any AI tools and the only systematic difference in the groups' instruction is whether or not they used the AI portrait prompt writing methodology.

Data Collection Procedure

Data collection is carried out following a well-defined 9-step methodology which ensures that the study is done in a rigorous and consistent way. The researcher gets official permission to carry out the research in the chosen government secondary school. Participants are then informed to give their consent, and for those who are minors, their parents or guardians, before data are collected. The vocabulary pre test is given to all participants under standardized conditions and then participants are randomly assigned to the experimental and control groups as detailed in section 3.8. This 8 week intervention then takes place, following the weekly plan outlined in section 3.16, where the experimental group will receive AI portrait prompt writing instruction and the control group will receive conventional instruction. The vocabulary post test is then given to both groups under similar conditions as the pre test at the end of the intervention. The pretest and posttest scripts are then marked on a set marking scheme and the marks scored are then entered into IBM SPSS Statistics for analysis and then the statistical procedures outlined in section 3.19 are applied.

Data Analysis

All data collected through the pretest and posttest are analyzed using IBM SPSS Statistics, a widely used software package for statistical analysis in educational and social science research.

Descriptive Statistics

Descriptive statistics are first computed to summarize the overall characteristics of the dataset, including the frequency and percentage distribution of scores, the mean and standard deviation of pretest and posttest scores for each group, and the minimum and maximum scores obtained. These descriptive statistics provide an initial overview of central tendency and dispersion within each group and form the necessary foundation for the subsequent inferential analysis.

Inferential Statistics

Inferential statistical tests are then applied to determine whether observed differences between and within groups are statistically significant, and to estimate the magnitude of the intervention's effect. An independent samples t-test is used to compare the pretest scores of the experimental and control groups, in order to confirm their statistical equivalence prior to the intervention, and is used again to compare posttest scores between the two groups, in order to

determine whether the experimental group significantly outperformed the control group following the intervention. A paired samples t-test is used to compare pretest and posttest scores within the experimental group, in order to determine whether that group's vocabulary scores improved significantly over the course of the intervention, and is used again to compare pretest and posttest scores within the control group, in order to determine whether conventional instruction alone produced a significant improvement. Finally, Cohen's d is calculated as a measure of effect size, in order to quantify the practical magnitude of the treatment effect independent of sample size, supplementing the statistical significance indicated by the t-tests with an indication of how large and educationally meaningful that effect actually is. The significance level for all inferential tests is set at $\alpha = .05$, the conventional threshold in educational research below which a result is considered statistically significant.

Objective	Statistical Test
Compare pretest scores between groups	Independent Samples t-test
Compare posttest scores between groups	Independent Samples t-test
Compare pretest and posttest within the experimental group	Paired Samples t-test
Compare pretest and posttest within the control group	Paired Samples t-test
Measure treatment effect	Cohen's d (Effect Size)

DATA Analysis and DATA Interpretation

Table 1 Independent t-test in Pre-test of Control and Experimental Group

Group Type	N	Mean	SD	D	t-value	Effect Size
Control	30	9.2	1.7	29	0.162	0.0004
Experimental	30	9.2	1.7	29		

The pre-test comparison showed that the Control group ($N = 30$, $M = 9.2$, $SD = 1.7$) and the Experimental group ($N = 30$, $M = 9.2$, $SD = 1.7$) had identical means and standard deviations. The independent samples t-test yielded $t(58) = 0.162$, $p > .05$, indicating no statistically significant difference between the two groups prior to the intervention. The effect size ($\eta^2 = 0.0004$) was negligible.

Interpretation

This result confirms that the two groups were statistically equivalent in vocabulary knowledge at baseline. This is a critical finding for a true experimental design, as it establishes that random assignment successfully produced comparable groups, and any differences observed at post-test can be more confidently attributed to the intervention (AI Portrait Prompt Writing) rather than pre-existing group differences.

Table 2 Independent t-Test in Post-test of Control and Experimental Group

Group Type	N	Mean	SD	D	t-value	Effect Size
Control	30	19.0	5.27	29	7.3	0.5
Experimental	30	30.8	7.07	29		

At post-test, the Experimental group ($N = 30$, $M = 30.8$, $SD = 7.07$) scored substantially higher than the Control group ($N = 30$, $M = 19.0$, $SD = 5.27$). The independent samples t-test yielded $t(58) = 7.3$, $p < .001$, indicating a statistically significant difference between groups. The effect size ($\eta^2 \approx 0.478$, equivalent to a Cohen's d of approximately 1.9) was large by conventional standards (Cohen, 1988).

Interpretation

As a result of the intervention, the mean score for the Experimental group exposed to AI Portrait Prompt Writing showed significantly higher amounts of vocabulary gain relative to the Control group receiving conventional instruction (Experimental group mean score of 11.8 points higher than the Control group mean score). In addition to being statistically significant, this difference has a large effect size, suggesting a significant effect of the intervention on vocabulary development.

Findings

The statistical analysis presented in Tables 4.1 and 4.2 yields a consistent and interpretable pattern of results that directly addresses the two research questions and four hypotheses guiding this study. The findings are summarized below in relation to each hypothesis.

Baseline Equivalence of Groups

Prior to the intervention, the Control group ($N = 30$, $M = 9.2$, $SD = 1.7$) and the Experimental group ($N = 30$, $M = 9.2$, $SD = 1.7$) obtained identical pretest means and standard deviations on the Vocabulary Achievement Test. The independent samples t-test confirmed that this difference was not statistically significant, $t(58) = 0.162$, $p > .05$, with a negligible effect size ($\eta^2 = 0.0004$). This finding establishes that the random assignment procedure described in Section 3.8 was successful in producing two groups that were statistically indistinguishable in vocabulary knowledge at the outset of the study, satisfying the internal validity requirement of a true experimental design.

Post-Intervention Difference Between Groups

Following the eight-week intervention, the Experimental group ($N = 30$, $M = 30.8$, $SD = 7.07$) outperformed the Control group ($N = 30$, $M = 19.0$, $SD = 5.27$) by a margin of 11.8 points on the same 40-item instrument. This difference was statistically significant, $t(58) = 7.3$, $p < .001$, and the associated effect size ($\eta^2 \approx 0.478$, equivalent to a Cohen's d of approximately 1.9) exceeded conventional benchmarks for a large effect (Cohen, 1988).

Within-Group Gains

Descriptively, both groups showed gains from pre- to post-test, with the Control group gaining 9.8 points (moving from 9.2 to 19.0) and the Experimental group gained 21.6 points (from 9.2 to 30.8). The increase in vocabulary gain was more than twice for the Experimental group compared with the Control group, showing that the conventional method of instruction alone could achieve measurable increases in vocabulary over the eight week period and content

coverage, but the conventional method with the addition of AI prompt writing for the portrait yielded significantly larger increases in the same period of time and content.

Hypothesis Testing Outcomes

- H_{01} (no significant pretest-posttest difference within the experimental group) is not supported: the experimental group's mean rose by 21.6 points across the intervention period, consistent with H_{11} .
- H_{02} (no significant posttest difference between the experimental and control groups) is rejected: the independent samples t-test at post-test was significant, $t(58) = 7.3$, $p < .001$, supporting H_{12} and confirming that the gain favored the experimental group.

Taken together, the pattern of no significant difference at pretest followed by a large, significant difference at posttest constitutes the expected statistical signature of an effective intervention within a true experimental pretest-posttest control group design, and provides quantitative support for the effectiveness of AI portrait prompt writing as a vocabulary-development intervention.

Discussion

The results of this study clearly support the proposition from the theoretical framework outlined in Chapter Two and supported in Chapter One that writing AI-generated portraits through AI prompts is more effective in enhancing the vocabulary development of secondary ESL learners than traditional, textbook-based vocabulary instruction. The two groups had nearly the same pretest scores ($M = 9.2$ for both), which increases confidence that the difference in posttest scores is the result of the instructional intervention and not due to selection bias or to unobserved differences in pretest skills between participants.

Interpretation in Light of Vygotsky's Sociocultural Theory

The Zone of Proximal Development (ZPD) is a construct developed by Vygotsky (1978) that suggests learning is best maximized when tasks are presented to the learner that are just beyond their current ability level, and are then supported by a mediating tool or more capable learner. In this study, the generative AI system seems to have acted as such an intermediary artifact, adjusting dynamically to each learner's description and generating a scaffolded environment for testing, observing, and editing the vocabulary that was being generated. This additional scaffolding by the tool, in addition to the continuing teacher scaffolding described in Section 3.16, may have enabled the experimental group to try to generate more descriptive language than they could unaided, allowing them to go beyond the productive language that was elicited from the control group by conventional teacher-fronted instruction.

Interpretation in Light of Paivio's Dual Coding Theory

Paivio's (1971, 1986) Dual Coding Theory proposes that information encoded through both verbal and visual channels simultaneously is retained more durably than information encoded verbally alone. Unlike conventional picture-vocabulary matching, in which learners are shown a pre-made image alongside a target word, AI portrait prompt writing required learners to generate the visual pairing themselves through their own word choices. This learner-generated form of dual coding may explain why the experimental group's improvement so markedly exceeded that of the control group, which received the same lexical content without an equivalent visual-generative component. This finding is consistent with prior research indicating that generative multimodal tasks, in which learners actively construct the word-image association, tend to produce stronger retention than receptive multimodal tasks in which the pairing is simply presented to them (Yang & Chen, 2023).

Interpretation of the Control Group's Improvement

This is significant because the control group also showed significant gains across the 8-weeks from pretest mean = 9.2 to the post test mean = 19.0. This finding makes sense and is a methodologically useful one: it establishes that the control group was provided with authentic,

formal instruction in the same vocabulary content but not with no instruction at all, so that the results from this study reflect the unique contribution of AI prompt writing for portrait writing over and above typical classroom instruction. Since, the gain of the experimental group (21.6 points) is more than double the gain of the control group (9.8 points) after both groups studied the same amount of vocabulary in the same amount of time, it is likely that the difference is due to the instructional method being used and not due to the amount of time studied or content covered.

Conclusion

This study set out to investigate the effect of AI portrait prompt writing on the vocabulary development of secondary ESL learners in government schools of Tehsil Sadiqabad, District Rahim Yar Khan, and to compare its effectiveness with conventional vocabulary instruction. Using a true experimental pretest-posttest control group design with sixty randomly assigned Grade 9 learners, the study found that the experimental and control groups were statistically equivalent at pretest, confirming successful randomization, but diverged sharply by posttest, with the experimental group outperforming the control group by 11.8 points and demonstrating a large effect size (Cohen's $d \approx 1.9$).

On the basis of these findings, both null hypotheses are rejected. There is a substantial improvement in vocabulary scores within the experimental group from pretest to posttest, and there is a statistically significant difference in posttest vocabulary achievement between the experimental and control groups, favoring the experimental group. It may therefore be concluded that AI portrait prompt writing is an effective pedagogical intervention for developing the appearance, emotion/expression, and spatial/lighting vocabulary of secondary ESL learners, and that it produces significantly greater vocabulary gains than conventional, textbook-based instruction covering the same content over an equivalent instructional period. These results lend empirical support to the integration of Input Hypothesis, Output Hypothesis, sociocultural, and Dual Coding theoretical perspectives advanced in Chapter Two, and demonstrate that these frameworks, though formulated prior to the advent of generative AI, retain strong explanatory power when applied to this novel instructional task.

Educational Implications

The results have direct implications for ESL teaching in under-resourced public-sector contexts. They propose that, if available, generative AI technologies can complement, or be used in parallel to, current vocabulary programs without replacing teachers' efforts, as in this study, and that the use of these tools should not be considered a substitute for teacher-led instruction. The size of the effect observed also indicates that, even for a relatively small amount of time (eight weeks of integration), such tools can have a significant effect on vocabulary gains compared to traditional approaches to instruction, and this is especially significant for leaders considering the cost and feasibility of implementing technology as opposed to the benefits of instruction.

Suggestions for Future Research

- Future studies should extend the intervention period beyond eight weeks to determine whether the observed vocabulary gains are retained over the longer term, through the inclusion of a delayed posttest.
- Future research should examine the differential effect of AI portrait prompt writing across the three specific vocabulary domains assessed in this study (appearance, emotion/expression, and spatial/lighting language) rather than treating vocabulary development as a single composite score.
- Future studies should replicate this design across multiple schools and districts within Punjab to enhance the generalizability of the findings beyond the single-school sample examined here.

- Future research should investigate learner and teacher perceptions of AI portrait prompt writing through qualitative or mixed-methods designs, in order to complement the purely quantitative causal evidence produced by the present true experimental design.
- Future studies should compare AI portrait prompt writing against other generative AI vocabulary tasks, such as chatbot-mediated writing assistance, to determine whether the image-generating component specifically, rather than generative AI engagement in general, accounts for the large effect size observed here.

References

- Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge University Press.
- Coleman, H. (2010). *Teaching and learning in Pakistan: The role of language in education*. British Council Pakistan.
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5–24.
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70–105.
- Kohnke, L., Zou, D., & Chen, X. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- Kukulska-Hulme, A. (2012). Language learning defined by time and place: A framework for next generation designs. In J. E. Díaz-Vera (Ed.), *Left to my own devices: Learner autonomy and mobile-assisted language learning* (pp. 1–13). Emerald.
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26.
- Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford University Press.
- Manan, S. A., David, M. K., & Dumanig, F. P. (2017). Language management: A snapshot of governmentality within the private schools in Pakistan. *Language Policy*, 16(3), 299–321.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. Routledge.
- Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart and Winston.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
- Sadoski, M., & Paivio, A. (2013). *Imagery and text: A dual coding theory of reading and writing* (2nd ed.). Routledge.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum.



- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos (Ed.), *Multimedia language teaching* (pp. 3–20). Logos International.
- Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, 31(2), 57–71.
- Warsi, J. (2004). Conditions under which English is taught in Pakistan: An applied linguistic perspective. *Sarid Journal*, 1(1), 1–9.
- Yang, X., & Chen, S. (2023). Augmented reality and vocabulary learning: A systematic review. *Computer Assisted Language Learning*, 36(4), 601–628.
- Zhang, R., & Zou, D. (2022). Artificial intelligence technologies for vocabulary learning: A systematic review. *Computer Assisted Language Learning*. Advance online publication.