Exploring the Impact of AI-Assisted Learning-Oriented Assessment on Vocabulary Acquisition Among Iranian EFL Learners

Hassan Alizadeh Mahmoud Alilo*

Department of English, Tabriz Branch, Islamic	*Corresponding Author
Azad University, Tabriz, Iran	Hassan Alizadeh Mahmoud Alilo, Department of English, Tabriz Branch, Islamic
	Azad University, Tabriz, Iran.

Submitted: 2025, Apr 15; Accepted: 2025, May 19; Published: 2025, May 26

Citation: Alilo, H. A. M. (2025). Exploring the Impact of AI-Assisted Learning-Oriented Assessment on Vocabulary Acquisition Among Iranian EFL Learners. *AI Intell Sys Eng Med Society, 1*(1), 01-06.

Abstract

Given that technology-enhanced assessment has transformed language education by providing adaptive and interactive evaluation methods, the current study examined the effect of AI-assisted learning-oriented assessment (LOA) on Iranian EFL learners' vocabulary learning. A quasi-experimental design was employed where 40 male intermediate-level learners were non-randomly distributed into experimental (AI-supported Nearpod platform) and control groups. The homogeneous treatment groups were verified using the Oxford Placement Test, and the Vocabulary Knowledge Scale assessed vocabulary development before and after intervention. Statistically significant differentiations were estimated from analyses of data using an independent samples t-test in SPSS 27, indicating that the control group does not show such significant improvement in post-VKS scores compared with the experimental group. The results reveal that AI-assisted LOA with adapted feedback and scaffolding benefited vocabulary learning. The findings highlighted that the incorporation of AI tools in EFL curricula can be optimized for engagement and personalized learning; nevertheless, effective implementation of AI tools has technical and pedagogical challenges.

Keywords: Artificial Intelligence, EFL Learners, Learning-Oriented Assessment, Vocabulary Learning

1. Introduction

Artificial Intelligence integration in education caused fundamental changes in how students learn languages while assessments are performed. AI-assisted learning-oriented assessment in English Language Teaching represents a pedagogical innovation that produces adaptable learning environments with personalized feedback. The AI-powered interactive platform Nearpod has obtained wide adoption because it offers students synchronized formative assessments with customized instruction along with various interactive learning components. This is particularly relevant to vocabulary acquisition, which is widely recognized as a key determinant of linguistic proficiency and communicative competence [1].

Traditional vocabulary instruction, which often relies on textbook exercises, memorization, and paper-based testing, tends to assess vocabulary knowledge in a fragmented and superficial manner. Such methods overlook the depth of word knowledge and the learner's ability to use vocabulary flexibly across contexts [2]. In contrast, AIenhanced LOA platforms like Nearpod offer dynamic, data-driven alternatives. These systems allow for immediate, personalized feedback, enabling learners to interact with vocabulary through quizzes, polls, multimedia content, and scaffolded prompts tailored to individual responses. Nearpod's adaptive functionalities—such as contextual hints, real-time error detection, and tiered support—allow instructors to assess not just recognition but also contextual usage and word integration in meaningful settings [3,4].

Recent research underscores the effectiveness of digital tools, especially in Computer-Assisted Language Learning (CALL) environments, in promoting vocabulary retention. For example, studies have shown that AI-driven and mobile-supported platforms significantly outperform traditional methods in improving vocabulary acquisition, especially for intermediate learners [5]. Nearpod, as a learner-centered and interactive platform, contributes to this trend by fostering learner engagement and offering ongoing formative assessment. Nevertheless, research on its specific impact, particularly within the framework of AI-assisted assessment, is still in its infancy, especially in EFL contexts like Iran.

Vocabulary acquisition is inherently complex, involving not only initial word recognition but also long-term retention, contextual application, and lexical depth. Effective instruction requires tools that can repeatedly expose learners to target vocabulary in varied contexts and provide nuanced feedback aligned with learners' developmental stages. Nearpod is designed to facilitate such recursive learning processes through features like embedded quizzes, interactive readings, and layered scaffolding strategies that adapt based on learner input. These affordances enable differentiated instruction and cognitive support, which are essential for sustained vocabulary growth [6].

Reviewing the literature, recent studies have increasingly underscored the role of AI and digital platforms in enhancing vocabulary learning within EFL and ESL contexts. demonstrated that Nearpod, as an interactive and gamified tool, significantly improved vocabulary acquisition among young learners, who reported heightened engagement and motivation due to its visual and activity-based features [7]. Similarly explored Nearpod's application in a CALL framework with adult learners, showing notable gains in phrasal verb mastery and learner autonomy, highlighting its ability to facilitate immediate feedback and varied instructional strategies [8]. investigated the broader concept of LOA, revealing its effectiveness in vocabulary learning but limited influence on long-term retention, thereby suggesting the need for sustained and repeated exposure [9].

Meanwhile, examined perceptions surrounding AI in ESL vocabulary instruction, revealing a generally positive attitude from students who appreciated the personalized and immersive aspects of AI-assisted learning, although concerns about diminished teacher presence and technical challenges were also noted [10]. Complementing these findings, provided empirical evidence supporting the effectiveness of AI-powered language learning platforms and mobile applications, which not only improved vocabulary learning pathways [11]. Collectively, these studies reinforce the pedagogical value of integrating AI-based tools, such as Nearpod, into vocabulary instruction, while also emphasizing the need for balanced implementation strategies that address both learner preferences and instructional efficacy.

Despite growing interest in AI-enhanced learning, there remains a significant gap in empirical research investigating the effects of tools like Nearpod when used in a structured, assessmentintegrated instructional model. Particularly in Iranian EFL settings, few studies have examined how Nearpod—as an AI-assisted LOA tool—compares with conventional vocabulary teaching in fostering measurable vocabulary gains. Most existing studies focus either on general CALL applications or on learner perceptions, without isolating the assessment dimension of such technologies.

Against this backdrop, the current study seeks to address this gap by investigating the effect of AI-assisted learning-oriented assessment on vocabulary learning among intermediate-level Iranian EFL learners. By employing an experimental research design and validated measurement instruments, this study aims to provide empirical evidence regarding the effectiveness of AI-based assessment tools in enhancing lexical development. To direct the investigation, this study formulates the following research question: • Does AI-assisted learning-oriented assessment have any significant

impact on Iranian EFL learners' vocabulary learning?

2. Method

2.1. Participants

The participants were obtained from a population of 350 from a private institute in Tabriz, namely Goldis. The final sample consisted of 40 intermediate male EFL learners, aged 15-24 years, all native Persian speakers, recruited from an institute following identical curricula. This method guarantees that participants have been exposed to a comparable academic background. A convenience sampling method was employed to select participants due to practical constraints such as accessibility and willingness to participate. This method ensured that the selected participants were representative of the general population of lower-intermediate EFL learners at the institute, as they shared similar educational backgrounds, language proficiency levels, and learning environments. According to the placement criteria of the institute, they were intermediate students. Nonetheless, to guarantee the integrity and uniformity of the participants, a proficiency test was administered before the commencement of the primary research. From the proficiency test, selected candidates with scores in the range of 30-39 have been considered for study and classified as proficient. Participants were then non-randomly allocated to the experimental and control groups, each with 20 students. The same instructor taught both groups to eliminate teacher confounding variables. The participants have already studied Evolve 1-3. They remained on Evolve 4 for the duration of the study.

2.2. Instruments

To gather the data needed for the study, the researcher applied the following instruments at various stages of the study.

2.2.1. Oxford Placement Test (OPT)

The OPT developed was systemically applied to assess and verify if the proficiency levels of English language differed in any significant ways between the experimental and control groups investigated [12]. This test was chosen because it is a widely recognized and standardized assessment tool that accurately measures language proficiency across different CEFR levels. One of the key features of the OPT is its ability to function as a homogenizing tool by assessing a range of linguistic competencies, including grammar, vocabulary, and reading comprehension. The test is a well-structured formal evaluation diluted to six levels of proficiency on the CEFR scale and assigns test scores to well-defined value boundaries for each of the discrete levels: Basic (A1: 0-17), Elementary (A2: 18-29), lower intermediate (B1: 30-39), upper intermediate (B2: 40-47), advanced (C1: 48-54) and very advanced (C2: 54-60). These categorizations conform to established standards of language proficiency, making it possible to assess participants' skills. The OPT results collected at the onset of the study were vital in that they provided researchers with the ability to intentionally select individuals whose scores fell into the Lower Intermediate (B1: 30-39) range to maintain uniform language proficiency standards within the groups.

2.2.2. Vocabulary Knowledge Scale (VKS)

The VKS developed was employed to measure students' vocabulary

knowledge before intervention as pre-VKS and after intervention as post-VKS. This instrument, originally developed as a comprehensive word knowledge test, requires language learners to demonstrate their familiarity and usage of target words using a five-point scale that ranges from complete unfamiliarity ("I don't remember having seen this word before") to the ability to use the word in a sentence accurately and appropriately [13]. The VKS assesses two main constructs: vocabulary size, which is measured through four items that capture the continuum from total unfamiliarity to correct meaning identification, and vocabulary depth, which is evaluated by asking students to produce a grammatically and semantically correct sentence using the word.

This scale was chosen for its ability to provide verifiable evidence of both receptive and productive knowledge, making it an ideal tool for research focused on word identification and utilization in EFL contexts. To ensure cultural and linguistic relevance, the VKS instructions were translated into Persian and administered on a separate sheet, and its validity has been supported by previous research. In the pre-VKS, to verify that the students were unfamiliar with the vocabulary they were expected to learn during their EFL classes in the treatment period, an 80-item vocabulary questionnaire was administered before the experiment. After analyzing the questionnaire responses, 50 items that the students did not recognize were selected as the target words for treatment, while the 40 items that were familiar to the students were removed from further consideration.

For the post-VKS, these 50 unfamiliar words were employed to assess any vocabulary gains resulting from the treatment. In the current study, scoring was conducted independently by two raters to ensure inter-rater reliability, with responses scored as follows: a score of 0 for complete unfamiliarity, 1 for basic recognition without understanding, 2 for correctly providing a synonym or translation, and either 3 or 4 for using the word in context, with a 3 assigned for contextually correct but ungrammatical usage and a 4 for fully correct usage, resulting in a per-word score that ranges from 0 to 4. To maintain consistency, both raters engaged in a discussion to resolve any discrepancies in scoring. This approach ensured that differences in interpretation were addressed collaboratively, leading to a more reliable and standardized assessment process.

2.3. Procedure

This study utilized an only post-test quasi-experimental design to explore the impact of AI-assisted learning-oriented assessment on vocabulary learning among Iranian EFL learners. The procedure was grounded in rigorous, reliable, ethical principles, systematic steps implemented chronologically from November 2024 to January 2025 at Goldis Language Institute, Tabriz, Iran. The study was approved by the Ethical Committee of Goldis Language Institute, and they acted as the gatekeepers to ensure to follow ethical research matters were followed before the study. Recruitment was conducted during regular class sessions. The researcher described the purpose, procedures, and voluntary nature of the study and assured participants of anonymity and confidentiality as well as the right to withdraw from the study without consequence. Informed written consent was obtained from all subjects using printed forms that were signed, returned, and kept in a locked filing cabinet accessible to the researcher.

Prior to the main data collection, a pilot study was done to improve the clarity and reliability of the research instruments used. 20 learners who matched the proficiency level of the main sample completed the VKS. This pilot study was conducted to detect ambiguities in the test items, assess the time needed to complete the test items, and validate the Persian version of the VKS. The pilot study showed the instrument to have high reliability for internal consistency, with Cronbach's alpha coefficients of 0.79 for the VKS, supporting its use in the main study. To ensure homogeneity amongst participants, the OPT was first conducted in a 60-minute session where all participants took part. Answer sheets were collected manually and scored by the researcher using the official scoring key.

Only participants with 30–39 scores (lower-intermediate, B1) were considered to maintain consistency. Out of an original pool of 350 male students, only 40 were kept, with some ruled out because their scores were outside of this range. Results were entered manually in a spreadsheet, with a colleague double-checking the scores for accuracy. Participants were divided into two groups: Experimental (n=20) and Control (n=20). This non-random assignment ensured intra-institutional uniformity. To control for teacher-related variation, all groups were taught by the same instructor, who was fluent in Persian and English.

Prior to the treatment, the VKS, as a pre-test, was administered in an 80-minute session. Scores were recorded manually on a scoring sheet. This test was conducted to ensure learners' unfamiliarity with target words. The words that learners were familiar with were discarded from the treatment. It was distributed in print and recorded in the spreadsheet. Over eight weeks, each group received distinct instructional approaches tailored to their assigned assessment method. The experimental group engaged with customized Nearpod software designed to deliver dynamic scaffolding during vocabulary tasks. During weekly 50-minute sessions, students interacted with Evolve 4 reading activities containing target words, triggering four levels of computerized mediation upon errors: implicit prompts (e.g. contextual highlighting), contextual clues (synonyms/definitions), explicit explanations (grammatical rules), and direct answers. The software logged responses and mediation usage, while the instructor monitored progress without direct intervention. Meanwhile, the control group followed conventional instruction: target vocabulary was taught through textbook drills, rote memorization, and teacherled translations, with corrections limited to end-of-unit tests and no AI support. Both groups adhered to the same Evolve 4 curriculum and session duration, with the instructor trained to standardize delivery across conditions, ensuring methodological consistency while isolating the effects of AI-assisted interventions.

Identical to the pre-test, post-intervention data were collected after conducting treatment, and inter-rater reliability was calculated, with discrepancies resolved through discussion to ensure consistency in post-VKS. All data were stored and anonymized using participant codes (e.g. Ex-01). The same classroom conditions (e.g. lighting, seating) and timing (morning sessions) were maintained across institutes to minimize external variables. Data analysis was conducted in SPSS 27, with post-test scores compared using an independent sample t-test, ensuring statistical rigor.

2.4. Research Design

This quasi-experimental research, which had only a post-test-control group design, required the existence of two groups: an experimental group and a control group. Quasi-experimental research is research that includes experimentation but is not truly experimental. Instead, their effects are based on the manipulation of the independent variable (Cook & Campbell, 1979). Participants are not randomly assigned to conditions or sequences of conditions. The experimental group was given treatment by the innovative methodologies of the AI-assisted learning-oriented assessment, and the control group was instructed conventionally. Thus, the dependent variable in the scope of this study becomes vocabulary learning, while AI-assisted learning-oriented assessment becomes the independent variable.

2.5. Data Analysis

The collected data were entered into SPSS 27 for further statistical analysis. At the onset, the OPT scores checked the initial homogeneity between the two groups. Then, Cronbach's alpha was used to check the internal consistency of the VKS. Descriptive statistics, including mean and standard deviation (SD) and standard errors (SEs), were presented for VKS. The Pearson correlation coefficient was used to evaluate inter-rater reliability between the two raters. The researcher used a Normality test to check the normal distribution of data. In the case of normal data, an independent sample t-test was used to explore the effect of the independent variable on the dependent variable.

3. Results

In order to answer the posed research question, some calculations, statistical routines, and results were produced. The results from the analysis of the post-VKS administered to both groups are indicated below. The details about descriptive statistics of groups regarding the post-VKS are illustrated in Table 1.

Pos-VKS	Group	N	Mean Std. Deviation		Std. Error Mean	
Experimental group		20	130.05	2.372	.530	
	Control group	20	99.85	2.476	.553	

Table 1: Group Statistics

As Table 1 demonstrates, the mean score of the post-VKS for the experimental group is 130.05 (SD= 2.372, SE= .530), and the control group had a mean of 99.85 (SD= 2.476, SE= .553).

Additionally, the Pearson correlation coefficient was used to evaluate inter-rater reliability and compare the consistency between both raters. Table 2 outlines these analyses.

		Rater 1	Rater 2				
Post-VKS of Control Group (Rater 1)	Pearson Correlation	1	.945**				
	Sig.(2-tailed)		.000				
	Ν	20	20				
Post-VKS of Experimental Group (Rater 2)	Pearson Correlation	.945**	1				
	Sig.(2-tailed)	.000					
	Ν	20	20				
**. Correlation/is/significant at the 0.01 level (2-tailed).							

Table 2: Inter-Rater Correlation for the Post-VKS Scores

As Table 2 displays, for the post-VKS scores of the control group, the inter-rater correlation was almost perfect for the control group, as r = .945 (p < .001), i.e. excellent scoring consistency. In the same way, the post-VKS scores of the experimental group exhibited identical reliability (r = .945, p < .001), indicating that raters consistently applied the scoring criteria between groups after intervention.

Table 3 presents the results of normality tests conducted on the post-VKS scores for both the experimental and control groups. The Kolmogorov-Smirnov and Shapiro-Wilk tests were used to assess whether the data followed a normal distribution, a key assumption for parametric statistical analyses like the independent samples t-test.

Shapiro-Wilk	Group	Kolmogorov-Smirnova			Shapiro-Wilk			
		Statistic	df	Sig.	Statistic	df	Sig.	
	Experimental group	.192	20	.053	.949	20	.356	
	Control group	.176	20	.106	.963	20	.612	
a. Lilliefors Significance Correction								

Table 3: Tests of Normality

As Table 3 demonstrates, for both groups, the p-values for the Kolmogorov-Smirnov and Shapiro-Wilk tests were greater than the conventional alpha level of .05, indicating that the data did not significantly deviate from normality. This supports the use of parametric tests (e.g. t-tests) for further analysis.

Table 4 displays the results of the independent samples t-test comparing the post-VKS scores of the experimental and control groups.

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Cor of the Di	nfidence Interval fference	
									Lower	Upper	
Pos- VKS	Equal variances assumed	.062	.805	39.379	38	.000	30.200	.766	28.647	31.752	
	Equal variances not assumed			39.379	37.930	.000	30.200	.766	28.647	31.752	

Table 4: Independent Samples Test

As Table 4 displays, Levene's Test for Equality of Variances indicated no significant difference in variances between the groups (F = .062, p = .805), confirming the assumption of homogeneity of variance. The t-test revealed a highly significant difference in vocabulary knowledge scores between the experimental group (M = 130.05) and the control group (M = 99.85), with a t-value of 39.379 (df = 38) and a significance level of p < .001. The mean difference of 30.20 (SE = .766) was substantial, with the 95% confidence interval ranging from 28.65 to 31.75. These results strongly suggest that the AI-assisted learning-oriented assessment had a significant positive impact on vocabulary acquisition among Iranian EFL learners when compared to traditional methods.

4. Discussion

The findings of this study provide robust evidence supporting the effectiveness of AI-assisted LOA in enhancing vocabulary acquisition among Iranian EFL learners. The experimental group, which received instruction through the Nearpod platform integrated with scaffolded AI-driven assessment, significantly outperformed the control group, which followed conventional instructional methods. The statistically significant difference in post-VKS scores, with a large effect size, affirms the pedagogical advantage of integrating AI with formative assessment tools.

These results are consistent with previous studies that have highlighted the positive role of technology-enhanced learning environments in vocabulary development [5,11]. Nearpod's interactive and adaptive design likely contributed to improved engagement, deeper cognitive processing, and more personalized feedback—all of which are critical in supporting long-term vocabulary retention. The platform's scaffolding capabilities allowed learners to receive real-time corrections and hints, enabling immediate awareness and adjustment, which is often missing in traditional instruction [3,7].

Additionally, this study contributes to the broader body of literature on learning-oriented assessment by operationalizing AI not merely as a content delivery tool, but as an active participant in the feedback and assessment cycle. This supports earlier findings who emphasized that LOA can significantly improve vocabulary learning outcomes when learners are continuously exposed to personalized and recursive feedback mechanisms [9]. However, while the experimental group demonstrated superior performance, the study also raises important considerations about implementation contexts. Although Nearpod's effectiveness is clear, its utility may depend on adequate digital literacy, instructor training, and infrastructural support. Furthermore, concerns identified in earlier studies—such as diminished teacher presence or over-reliance on automation (Alharbi & Khalil, 2023)—must be addressed through balanced pedagogical integration. Finally, while the VKS offered a reliable and comprehensive measure of both receptive and productive vocabulary knowledge, it primarily captures short-term gains [10]. Future studies may benefit from incorporating delayed post-tests to evaluate long-term retention and lexical depth over extended periods.

5. Conclusion

This study investigated the impact of AI-assisted LOA on vocabulary acquisition among intermediate-level Iranian EFL learners. The significant improvement in the experimental group's post-VKS scores provides compelling evidence that AI-powered platforms like Nearpod can substantially enhance vocabulary learning when integrated with adaptive feedback and scaffolded instructional strategies. The results affirm that when augmented with AI, LOA boosts learner engagement and immediate performance and supports deeper, more contextualized word knowledge. These findings have important implications for EFL curriculum designers, educators, and policy-makers aiming to modernize language instruction through data-informed and learnercentered approaches. Nonetheless, successful implementation requires thoughtful integration that considers technical readiness, teacher training, and learner preferences. As digital technologies become increasingly central to education. future research should explore long-term effects, cross-cultural applications, and potential challenges to further refine AI-assisted LOA frameworks in EFL contexts.

References

 Tiansoodeenon, M., Meeporm, B., Kaewrattanapat, N., & Tarapond, S. (2023). Enhancing Vocabulary Acquisition through Progressive Word Increments in English Language Learning. Journal of Liberal Arts RMUTT, 4(2), 88-100.

- Munro, N., Baker, E., Masso, S., Carson, L., Lee, T., Wong, A. M. Y., & Stokes, S. F. (2021). Vocabulary acquisition and usage for late talkers treatment: Effect on expressive vocabulary and phonology. *Journal of Speech, Language, and Hearing Research, 64*(7), 2682-2697.
- Dujardin, E., Auphan, P., Bailloud, N., Ecalle, J., & Magnan, A. (2021). Tools and teaching strategies for vocabulary assessment and instruction: A review. *Social Education Research*, 34-66.
- Ismail Omar, L. (2021). The use and abuse of machine translation in vocabulary acquisition among L2 Arabic-Speaking Learners. AWEJ for Translation & Literary Studies, 5(1).
- Shamshiri, F., Esfahani, F. R., & Hosseini, S. E. (2023). Models of assessment in the classroom: a comparative research of CALL-based vs. traditional assessment on vocabulary learning among Iranian EFL learners. *Language Testing in Asia, 13*(1), 43.
- Inam, S., Jawaid, M., & Khan, R. A. (2023). Assessment Of Workplace Related Factors Affecting Tolerance Of Ambiguity Among Trainee Doctors. JPMA. *The Journal of the Pakistan Medical Association*, 73(9), 1827-1832.
- Balqis, N., & Zaki, L. B. (2025). Classroom Action Research: Improving Young Learner's Vocabulary Using Nearpod.

Esensi Pendidikan Inspiratif, 7(1).

- Sánchez, L. M., & Carballo, Y. A. (2025). The impact of implementing the Computer-Assisted Language Learning (CALL) approach using the Nearpod platform in improving phrasal verbs vocabulary among adult learners at the virtual institute Centro de Matemáticas e Idiomas Segura in San Isidro de Alaju. *Ciencia Latina Revista Científica Multidisciplinar*, 9(2), 1060-1078.
- 9. Seyed, F. S., & Tavassoli, K. (2023). The Impact of Learningoriented Assessment on EFL Learners' Vocabulary Learning and Retention in Online Classes. *JELT Journal* | *Farhangian University*, 2(1), 84-99.
- 10. Alharbi, K., & Khalil, L. (2023). Artificial intelligence (AI) in ESL vocabulary learning: An exploratory study on students and teachers' perspectives. *Migration Letters, 20*(S12), 1030-1045.
- 11. Wang, Y., Wu, J., Chen, F., Wang, Z., Li, J., & Wang, L. (2024). Empirical assessment of AI-powered tools for vocabulary acquisition in EFL instruction. IEEE Access.
- 12. Dave, A. (2004). Oxford placement test 2: Test pack. Oxford University Press.
- 13. Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.