Iranian Journal of Educational Sociology

Journal Website

Article history:
Received 01 February 2026
Revised 22 June 2026
Accepted 29 June 2026
Initial Publication 25 August 2026
Final Publication 01 November 2026

Iranian Journal of Educational Sociology

Volume 9, Issue 6, pp 1-17

Quarterly
Iranian Journal of Educational Sociology

E-ISSN: 2645-3460

Lexicon Mastery Using AI-Generated Materials in Massed, Spaced, and Hybrid Practice for Enhanced Vocabulary Retention and Faster Retrieval among Iranian EFL Learners

Mohammad. Zakizadeh[1], Mohammad Taghi. Hassani[2*], Mehrdad. Rezaee[1]

[1] Department of English Language, CT.C., Islamic Azad University, Tehran, Iran
[2] Department of Foreign Language Center, Imam Hossein University, Tehran, Iran

*** Corresponding author email address**: mthesni@ihu.ac.ir

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Zakizadeh, M., Hassani, M. T., & Rezaee, M. (2026). Lexicon Mastery Using AI-Generated Materials in Massed, Spaced, and Hybrid Practice for Enhanced Vocabulary Retention and Faster Retrieval among Iranian EFL Learners. *Iranian Journal of Educational Sociology*, *9*(6), 1-17.
https://doi.org/10.61838/kman.ijes.1494

A B S T R A C T

Purpose: The present study aimed to determine the internal reliability of the Comprehensive Talent Identification Model based on the SHAHAK Plan and the Renzulli Scale among male secondary school students in Tehran.

Methods and Materials: This applied descriptive-survey study was conducted using a field research design during the 2025–2026 academic year. The statistical population consisted of all male students enrolled in upper secondary schools in District 5 of Tehran. Using convenience sampling and based on Morgan's sample size recommendations, 385 students were selected for the final analysis. Data were collected using the Comprehensive Talent Identification Questionnaire derived from the Renzulli Scales for Rating the Behavioral Characteristics of Superior Students. The instrument consisted of 126 items distributed across fourteen dimensions, including learning, creativity, motivation, leadership, art, music, drama, communication accuracy, communication expression, planning, mathematics, reading, technology, and science. Following translation and cultural adaptation procedures, the questionnaire was administered through teacher ratings. Data were analyzed using SPSS version 25. Descriptive statistics were calculated, and inferential analyses included the Kolmogorov–Smirnov test, Kaiser–Meyer–Olkin (KMO) measure, Bartlett's Test of Sphericity, and Cronbach's alpha coefficient to evaluate internal reliability.

Findings: Cronbach's alpha coefficients for the fourteen dimensions ranged from 0.822 to 0.965, all exceeding the acceptable threshold of 0.70. The highest reliability coefficient was obtained for Motivational Characteristics ($\alpha = 0.965$), whereas the lowest was observed for Communication Expression ($\alpha = 0.822$). The overall questionnaire achieved an exceptionally high Cronbach's alpha coefficient of 0.989, indicating outstanding internal consistency.

Conclusion: The consistently high reliability coefficients across all dimensions demonstrate that the instrument is a dependable and psychometrically sound tool for identifying diverse domains of student talent and can be effectively utilized in educational assessment, talent development, and academic guidance programs.

Keywords: *Talent Identification, SHAHAK Plan, Renzulli Scale, Gifted Students, Psychometric Properties, Secondary School Students.*

1. Introduction

Vocabulary knowledge occupies a central position in English as a Foreign Language (EFL) learning because it directly shapes learners' ability to understand input, express meanings, participate in communication, and develop higher levels of linguistic competence. In EFL contexts, where learners have limited access to natural English-speaking environments, vocabulary development often depends on the quality, frequency, and structure of classroom exposure. Vocabulary is not merely a collection of isolated lexical items; rather, it functions as the semantic foundation through which listening, speaking, reading, and writing are gradually developed. Learners with limited vocabulary knowledge often experience difficulty in speaking and comprehension, and such limitations may reduce their willingness to communicate in English (Amalia & Anggraeni, 2021). For Iranian EFL learners, this problem is particularly important because English is usually learned in formal educational settings, while authentic exposure outside the classroom remains restricted. Therefore, designing effective vocabulary learning conditions is a major concern in applied linguistics, language pedagogy, and technology-enhanced language education.

Previous research has shown that learners use a wide range of vocabulary learning strategies, including memorization, note-taking, dictionary use, contextual guessing, repetition, multimedia exposure, and metacognitive regulation. Among Iranian EFL learners, strategy use is often influenced by learners' proficiency level, instructional background, motivation, and access to meaningful input (Amirian & Heshmatifar, 2013). Metacognitive vocabulary strategy training has also been shown to expand vocabulary breadth by helping learners monitor, regulate, and evaluate their learning processes (Askari, 2014). However, despite the usefulness of strategic learning, vocabulary knowledge is still vulnerable to forgetting when exposure is limited, mechanical, or disconnected from meaningful contexts. Vocabulary learning requires not only initial encoding but also repeated retrieval, contextual reinforcement, and consolidation across time. Thus, the central question is not only how learners encounter new words, but also how instructional schedules and materials can be arranged to support longer retention and faster retrieval.

Theoretical perspectives on learning provide a strong basis for examining vocabulary learning through different practice schedules. Ausubel's theory of meaningful learning emphasizes that new information becomes durable when it is meaningfully connected to prior cognitive structures rather than memorized mechanically (Bryce & Blown, 2023). This view is particularly relevant to vocabulary instruction because lexical items are more likely to be retained when learners encounter them in meaningful, contextualized, and personally interpretable input. Similarly, Vygotskian perspectives emphasize the social and mediated nature of learning, suggesting that cognitive development is supported through interaction, scaffolding, and culturally meaningful tools (Nardo, 2021; Ofori-Attah, 2021). In vocabulary learning, instructional tools such as video clips, transcripts, teacher mediation, AI-generated stories, and repeated contextual exposure can function as mediational resources that help learners internalize lexical knowledge. From this perspective, vocabulary mastery is not simply a private memorization process but a situated and mediated activity supported by materials, interaction, and structured practice.

Second language acquisition theory further clarifies why vocabulary input should be comprehensible, repeated, and pedagogically organized. Krashen's Monitor Model highlights the importance of meaningful input in language acquisition and argues that acquisition is facilitated when learners are exposed to language that is understandable and slightly beyond their current level (Hassan, 2022). In the same line, key concepts in second language acquisition indicate that vocabulary learning depends on input processing, noticing, frequency, output opportunities, and the gradual strengthening of form–meaning connections (VanPatten & Benati, 2015). Therefore, vocabulary instruction should provide learners with repeated opportunities to notice lexical forms, connect them to meaning, and retrieve them in different contexts. Bloom's taxonomy also provides a useful framework for assessing vocabulary outcomes because recall and recognition represent different levels of cognitive access to learned information (Krathwohl, 2022). Recognition tasks require learners to identify meanings from available options, whereas recall tasks require stronger productive access to lexical meaning. For this reason, both recall and recognition are necessary for evaluating vocabulary mastery.

A large body of research has demonstrated that vocabulary learning is strengthened when words are embedded in rich pedagogical contexts. Explicit and implicit instruction through reading can improve vocabulary development, especially when learners encounter target words in meaningful textual environments (Khamesipour,

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

2015). Vocabulary instruction beyond the elementary level is also most effective when it provides multiple sources of word learning and productive contexts that help learners use words across different communicative conditions (Ford-Connors & Paratore, 2015). Similarly, systematic vocabulary instruction can improve text comprehension when it promotes depth of word knowledge rather than superficial memorization (Wright & Cervetti, 2017). Genre-based listening instruction has also been reported to influence comprehension performance among Iranian EFL learners, suggesting that contextualized exposure to different types of discourse can support language understanding (Hassani & Sadeghi, 2014). In addition, vocabulary clusters can improve vocabulary achievement by helping learners organize lexical items in semantically meaningful patterns (Khalili, 2016a), while orthographic knowledge contributes to word identification and reading comprehension among Iranian learners (Khalili, 2016b). These findings collectively suggest that vocabulary learning improves when words are presented through meaningful, organized, and repeated input rather than isolated lists.

The problem of forgetting is particularly important in vocabulary learning. Ebbinghaus's work on memory and forgetting has long shown that newly learned information declines over time unless it is reinforced through review and retrieval (Roediger, 2015). Applications of the learning curve to educational planning suggest that distributed instruction can reduce forgetting by aligning review with the natural decline of memory traces (Nang, 2018). Spacing effects have therefore become highly relevant to EFL vocabulary instruction. Spaced learning allows repeated exposure to words across expanding intervals, increasing the likelihood that vocabulary knowledge will remain accessible after delay. Research with young EFL learners has shown that spacing can support vocabulary learning more effectively than immediate repetition alone (Lotfolahi & Salehi, 2017). Within-session spacing has also been found to improve delayed recall among children, confirming that temporal distribution can enhance memory durability (Zigterman et al., 2015). Moreover, presentation rate influences foreign-language vocabulary learning, indicating that the timing and rhythm of exposure affect how lexical items are encoded and retained (Zeelenberg et al., 2015). Therefore, the temporal organization of practice is a key instructional variable in vocabulary retention.

Practice schedule is also closely related to retrieval. Retrieval practice strengthens memory by requiring learners to actively access stored information, making later recall more efficient and durable. Foreign language learning research has shown that interleaving can produce benefits over blocking in pronunciation learning, implying that varied and distributed learning sequences may promote more flexible access to linguistic knowledge (Carpenter & Mueller, 2013). Although massed practice may create rapid short-term gains because learners repeatedly encounter material within a compressed time period, it may not provide sufficient temporal distance for durable consolidation. Spaced practice, in contrast, introduces desirable difficulty by requiring learners to retrieve information after some forgetting has occurred. Hybrid practice may combine the advantages of both approaches: initial massed exposure may help learners reach early familiarity, while later spaced review may stabilize memory and improve delayed retrieval. This distinction is important for EFL vocabulary instruction because learners often mistake immediate fluency after massed exposure for long-term mastery. A method that supports both fast learning and long-term accessibility is therefore pedagogically valuable.

Cognitive science also helps explain why multimedia and contextualized materials may enhance vocabulary learning. Predictive processing theories describe cognition as an active process in which learners continuously generate and update expectations based on incoming information (Clark, 2013). In vocabulary learning, video clips can support this process by combining verbal input, visual context, action, emotion, and situational cues. Embodied and interoceptive accounts of cognition further suggest that meaning is grounded in bodily, emotional, and contextual experience rather than abstract symbol manipulation alone (Seth, 2013). Thus, multimedia clips may strengthen vocabulary learning because they present words within perceptual and affective contexts that make meanings more concrete and memorable. Situational language teaching also supports vocabulary learning by placing lexical items in meaningful communicative contexts rather than presenting them as decontextualized forms (Huang & Liu, 2024). In EFL classrooms, short clips can provide compact but rich contexts in which learners see and hear vocabulary used in purposeful discourse.

The rise of artificial intelligence has introduced new possibilities for designing vocabulary learning materials. AI tools can generate level-appropriate stories, exercises, prompts, and contextualized examples that recycle target vocabulary across varied situations. In EFL writing instruction, AI has been used to support language development by providing feedback, modeling, and

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

opportunities for revision (Yang et al., 2025). AI tools can also increase EFL learners' engagement, enjoyment, and motivation, which are crucial affective conditions for sustained language learning (Yuan & Liu, 2025). In AI-based learning contexts, learners' motivation may be shaped by self-efficacy and resilience, suggesting that technology-enhanced instruction should not only provide materials but also support learners' confidence and persistence (Shi & Zhang, 2025). AI speech evaluation systems have also been examined for their perceived promise and limitations in EFL speaking practice, indicating that learners may benefit from automated, individualized, and repeated practice opportunities when such systems are pedagogically controlled (Zou et al., 2026). These developments show that AI can be used not as a replacement for instruction, but as a tool for generating repeated, contextualized, and adaptive vocabulary input.

The use of AI-generated materials is especially relevant to spaced and hybrid vocabulary learning because target words can be recycled in multiple contexts without excessive teacher preparation. For example, after learners encounter new vocabulary in short clips, AI can generate short stories that reuse the same words in different semantic and syntactic environments. This may deepen lexical knowledge by strengthening form–meaning connections and expanding contextual associations. At the same time, caution is necessary because AI-generated content must be reviewed for accuracy, appropriateness, and level suitability. Research on language models has raised questions about how such models memorize, retain, and reproduce knowledge, which highlights the need for careful human supervision when AI-generated materials are used in educational settings (Cao et al., 2023). Social and technological systems may also create unequal learning advantages when some learners benefit more from access, familiarity, or accumulated resources, a phenomenon that can be interpreted through Matthew-effect dynamics (Wang, 2023). Therefore, AI-based vocabulary instruction should be designed in a way that provides all learners with equal access, clear structure, and pedagogically meaningful repetition.

Learner variables must also be considered in vocabulary learning research. Gender differences in language learning have been examined in relation to strategy use, motivation, achievement, and classroom behavior (Kheder & Rouabhia, 2023). Learning styles, gender, and learning outcomes may also interact in ways that influence how learners respond to instructional materials and practice conditions (Marantika, 2022). However, gender differences should not be assumed without empirical evidence, particularly when multimedia and AI-supported learning materials are used. In addition, sleep and memory consolidation play an important role in learning new English words among Persian speakers, suggesting that vocabulary retention is influenced by post-learning consolidation processes as well as instructional exposure (Soleimannejad et al., 2022). Web-based vocabulary systems such as Leitner box applications integrated with social networks also demonstrate that technology can support repeated exposure and review, especially when learning is distributed and interactive (Gao et al., 2015). These findings reinforce the importance of examining both instructional method and learner-related conditions in EFL vocabulary retention.

Methodologically, the investigation of vocabulary learning through AI-supported massed, spaced, and hybrid practice requires a design that can first identify learners' real preferences and then test instructional effects quantitatively. Sequential exploratory mixed-methods designs are useful when qualitative findings are needed to inform the development of quantitative interventions and instruments (Creswell, 2018). In the present study, this logic is particularly appropriate because learners' perceptions of vocabulary learning materials were first explored through questionnaire and interview data, and the results were then used to select short clips and AI-generated supplementary materials for the intervention. This design allows the study to connect learner experience with experimental comparison, making the intervention more ecologically valid for Iranian EFL learners.

Despite extensive research on vocabulary instruction, memory, technology-enhanced learning, and AI-supported language education, several gaps remain. Many studies have examined vocabulary learning strategies, spacing effects, or digital tools separately, but fewer have compared massed, spaced, and hybrid practice using AI-generated contextual materials for delayed retention and retrieval. In addition, many vocabulary studies focus on immediate achievement, while long-term retention after extended delay and retrieval speed remain less frequently examined. This gap is important because true vocabulary mastery requires not only recognizing or recalling word meanings shortly after instruction, but also accessing them quickly and accurately after weeks or months. Therefore, a study that combines learner-selected multimedia input, AI-generated lexical recycling, delayed recall and recognition tests, and speed-based retrieval assessment can contribute meaningfully to EFL vocabulary pedagogy.

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

The aim of this study was to examine the comparative effectiveness of AI-supported massed, spaced, and hybrid practice on long-term vocabulary retention and retrieval speed among Iranian intermediate EFL learners.

2. Methods and Materials

2.1. Study Design and Participants

This study employed a sequential exploratory mixed-methods design consisting of a qualitative phase followed by a quantitative experimental phase. The rationale for adopting this design was to obtain an in-depth understanding of learners' experiences, preferences, and perceptions regarding vocabulary learning before developing and implementing the quantitative intervention. Findings derived from the qualitative stage informed the selection of instructional materials and the design of the treatment conditions used in the experimental phase. This approach enabled the integration of learners' authentic experiences into the development of pedagogically relevant interventions and provided a stronger foundation for examining the effectiveness of different vocabulary practice schedules.

The participants were 120 Iranian learners of English as a Foreign Language (EFL) studying at private language institutes in Qazvin, Iran. The sample consisted of 60 male and 60 female learners whose ages ranged from 18 to 30 years. All participants were native speakers of Persian and had studied English for at least three years. Prior to participation, learners were administered a standardized Vocabulary Levels Test (VLT) to ensure relative homogeneity in vocabulary proficiency. Only intermediate-level learners whose scores fell within the predetermined proficiency range were included in the study. Following the homogenization procedure, participants were randomly assigned to one of three experimental groups: Massed Practice (n = 40), Spaced Practice (n = 40), and Hybrid Practice (n = 40). Random assignment was used to minimize potential selection bias and to increase the internal validity of the study. Ethical approval was obtained from the relevant institutional review board, and all participants provided informed consent prior to participation. Participation was voluntary, and learners were informed that they could withdraw from the study at any time without penalty.

2.2. Data Collection Tools

Several instruments were employed to collect both qualitative and quantitative data. During the exploratory phase, a researcher-developed Vocabulary Learning Experience Questionnaire was administered to 50 EFL learners from the target population. The questionnaire consisted of 35 items measured on a five-point Likert scale ranging from strongly disagree to strongly agree. The instrument was designed to investigate learners' perceptions of effective vocabulary learning strategies, preferred learning materials, technology-assisted learning experiences, and attitudes toward multimedia resources. Content validity was established through expert review by three specialists in applied linguistics and language education. The reliability of the questionnaire was assessed through a pilot study, yielding a Cronbach's alpha coefficient of .89, indicating satisfactory internal consistency.

To complement the questionnaire findings, semi-structured demographic and educational interviews were conducted with the same participants. The interviews were carried out in Persian to allow participants to express their views freely and accurately. Interview questions focused on previous vocabulary learning experiences, preferred instructional materials, perceptions of technology-enhanced learning, and factors influencing vocabulary retention and retrieval. The qualitative data obtained from the interviews were analyzed thematically and used to guide the development of instructional materials for the intervention phase.

Vocabulary proficiency was measured using the Vocabulary Levels Test (VLT), a widely recognized instrument developed to assess learners' receptive vocabulary knowledge across different frequency levels. The VLT was employed as a placement and homogenization tool before the commencement of the intervention. The test has been extensively validated in second-language acquisition research and demonstrates strong psychometric properties across diverse EFL contexts.

To evaluate vocabulary retention, two researcher-developed achievement tests were designed based on the cognitive objectives of Bloom's Taxonomy. The first assessment measured vocabulary recall and required learners to produce the meaning of target lexical items in either English or Persian. The second assessment measured vocabulary recognition through multiple-choice items requiring participants to identify correct meanings, synonyms, or contextual uses of target words. Both tests were reviewed by a panel of experts in language assessment to establish content validity. Reliability estimates obtained through a pilot administration indicated acceptable levels of

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

internal consistency, with Cronbach's alpha coefficients exceeding .80 for both measures.

Vocabulary retrieval speed was assessed through a computerized Speed Retrieval Test developed using Microsoft PowerPoint. Each vocabulary item was displayed for a predetermined and strictly controlled time interval, preventing participants from allocating unused time from one item to another. The test measured the speed and accuracy with which learners accessed previously learned lexical information. Response times and accuracy scores were automatically recorded and subsequently analyzed to determine differences among the treatment groups.

2.3. Interventions

The intervention materials were developed based on findings obtained from the qualitative phase. Analysis of questionnaire and interview data revealed that learners considered short multimedia clips to be highly motivating and beneficial for vocabulary acquisition. Consequently, two authentic English-language video clips, each approximately 90 seconds in duration, were selected as the primary instructional materials. The clips contained contextualized vocabulary appropriate for intermediate EFL learners and were accompanied by complete transcripts. From these clips, 70 target lexical items were identified for instruction and assessment purposes.

Participants in the Massed Practice group received intensive vocabulary instruction through repeated exposure to the two video clips within closely spaced instructional sessions. Learners watched the clips several times during consecutive sessions and engaged in vocabulary-focused activities immediately afterward. These activities included matching exercises, contextual sentence completion tasks, translation activities, and recall exercises. Following completion of the intensive learning phase, participants completed immediate mastery assessments. No further exposure to the instructional materials occurred before the delayed post-test, which was administered approximately 100 days after the completion of instruction.

Participants in the Spaced Practice group received the same instructional content but according to a distributed learning schedule. Learners first viewed and studied the video clips during an initial instructional session. The materials were revisited one week later, followed by additional review sessions separated by progressively increasing intervals of two to three weeks. Beginning with the third repetition cycle, participants were exposed to four supplementary short stories generated through artificial intelligence technology. These stories incorporated the 70 target lexical items extracted from the original clips and provided learners with additional contextualized encounters with the vocabulary. The AI-generated stories were carefully reviewed by language experts to ensure linguistic accuracy, readability, and appropriateness for the participants' proficiency level. The final delayed assessment was administered approximately 100 days after the completion of the instructional period.

Participants in the Hybrid Practice group received an integrated intervention combining the principles of massed and spaced learning. Initially, learners engaged in intensive massed practice with the two instructional clips until they demonstrated satisfactory mastery of the target vocabulary. Mastery was operationally defined as achieving at least 85% accuracy on formative vocabulary assessments administered during the learning phase. Learners who did not reach the mastery criterion repeated the massed practice cycle until satisfactory performance was achieved. Following successful mastery, participants entered the spaced practice phase, which mirrored the schedule used in the Spaced Practice group, including repeated review sessions and exposure to the AI-generated stories. This combined approach was designed to capitalize on the immediate learning benefits of intensive practice while simultaneously promoting long-term retention through distributed review.

At the conclusion of the intervention period, all participants completed delayed recall tests, delayed recognition tests, and computerized retrieval speed assessments. Alternate versions of the tests were used across measurement occasions to minimize practice effects, memory contamination, and potential cheating. Test forms were counterbalanced to ensure that differences in performance reflected genuine learning outcomes rather than familiarity with specific assessment items.

2.4. Data Analysis

Qualitative data obtained from the questionnaires and interviews were analyzed using thematic analysis. Interview transcripts were coded and categorized to identify recurring themes related to vocabulary learning preferences, effective learning materials, and technology-assisted language learning experiences. The resulting themes informed the development of the intervention materials and instructional procedures used in the experimental phase.

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

Quantitative data were analyzed using the Statistical Package for the Social Sciences (SPSS). Descriptive statistics, including means and standard deviations, were calculated for all dependent variables. Prior to inferential analyses, assumptions of normality, homogeneity of variance, and multivariate normality were examined. To evaluate the effects of the three instructional approaches on vocabulary learning outcomes, a one-way multivariate analysis of variance (MANOVA) was conducted. The independent variable was instructional condition with three levels (Massed, Spaced, and Hybrid), while the dependent variables consisted of vocabulary retention scores and vocabulary retrieval speed scores. When significant multivariate effects were identified, follow-up univariate analyses and post hoc comparisons using Bonferroni adjustments were conducted to determine the specific differences among groups. Effect sizes were calculated and interpreted to assess the practical significance of the observed differences. Statistical significance was established at the .05 level for all analyses.

3. Findings and Results

The findings are presented in accordance with the sequential exploratory structure of the study. First, the psychometric quality of the vocabulary learning experience questionnaire and the exploratory findings obtained from the initial 50 learners are reported. Then, the pilot study results are presented to justify the use of English clips as instructional input. Finally, the delayed vocabulary retention and retrieval outcomes of the Massed, Spaced, and Hybrid practice groups are reported using one-way ANOVA and one-way MANOVA.

The internal consistency of the 35-item questionnaire was examined using Cronbach's alpha. As shown in Table 1, the obtained alpha coefficient was .813, and the standardized alpha coefficient was .809. Since both values were higher than the acceptable threshold of .70, the questionnaire demonstrated satisfactory reliability. This result indicates that the items had adequate internal consistency and that learners interpreted the questionnaire items in a reasonably consistent manner.

Table 1

Reliability and Association Indices for the Vocabulary Learning Experience Questionnaire

Analysis	Index	Value	df	Sig.	N / Items
Questionnaire reliability	Cronbach's alpha	.813	—	—	35 items
Questionnaire reliability	Cronbach's alpha based on standardized items	.809	—	—	35 items
Gender × watching clips/movies	Pearson chi-square	3.535	3	.316	50 cases
Gender × watching clips/movies	Likelihood ratio	4.695	3	.196	50 cases
Gender × watching clips/movies	Linear-by-linear association	2.914	1	.088	50 cases
Symmetric measure	Phi	.266	—	.316	50 cases
Symmetric measure	Cramer's V	.266	—	.316	50 cases

The chi-square analysis was used to examine whether gender was significantly associated with using English clips and movies for vocabulary learning. The result was not statistically significant, $\chi^2(3) = 3.535$, $p = .316$. Therefore, the null hypothesis was retained, indicating that no statistically significant association existed between gender and learners' tendency to use English clips or movies for learning new vocabulary. The Phi and Cramer's V values were both .266 with $p = .316$, confirming the same conclusion. Although descriptive differences were observed between male and female learners, these differences were not strong enough to indicate a statistically significant relationship.

The descriptive results of the exploratory questionnaire showed that watching English clips, movies, and television programs was the most frequently endorsed vocabulary learning strategy. Among the 50 respondents, 36 learners, equivalent to 72%, reported that they "always" used clips or movies to learn new English words, while 11 learners, equivalent to 22%, reported using this strategy "often." Only one participant reported using this strategy "rarely," and two participants reported using it "sometimes." Therefore, 94% of the respondents selected either "often" or "always," indicating that multimedia exposure was perceived as a highly effective and preferred method of vocabulary learning. In contrast, interaction with native speakers was reported as the least accessible strategy. Forty-one learners, equivalent to 82%, stated that they "never" interacted with native speakers for vocabulary learning. This finding appears to reflect the limited availability of native English speakers in the learners' local context. Taking notes in class was the second most frequently endorsed strategy, with 29

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

participants, equivalent to 58%, selecting "always," and 13 participants, equivalent to 26%, selecting "often." Thus, note-taking was also perceived as a useful vocabulary learning strategy, although it was less strongly endorsed than watching clips and movies.

Table 2

Frequency Distribution of Major Vocabulary Learning Strategies in the Exploratory Phase

Strategy / Group	Response Category	Frequency	Percent	Valid Percent	Cumulative Percent
Watching clips/movies/TV	Rarely	1	2.0	2.0	2.0
Watching clips/movies/TV	Sometimes	2	4.0	4.0	6.0
Watching clips/movies/TV	Often	11	22.0	22.0	28.0
Watching clips/movies/TV	Always	36	72.0	72.0	100.0
Interaction with native speakers	Never	41	82.0	82.0	82.0
Interaction with native speakers	Rarely	3	6.0	6.0	88.0
Interaction with native speakers	Sometimes	2	4.0	4.0	92.0
Interaction with native speakers	Often	4	8.0	8.0	100.0
Taking notes in class	Never	1	2.0	2.0	2.0
Taking notes in class	Rarely	4	8.0	8.0	10.0
Taking notes in class	Sometimes	3	6.0	6.0	16.0
Taking notes in class	Often	13	26.0	26.0	42.0
Taking notes in class	Always	29	58.0	58.0	100.0
Male learners: watching clips/movies	Often	5	20.0	20.0	20.0
Male learners: watching clips/movies	Always	20	80.0	80.0	100.0
Female learners: watching clips/movies	Rarely	1	4.0	4.0	4.0
Female learners: watching clips/movies	Sometimes	2	8.0	8.0	12.0
Female learners: watching clips/movies	Often	6	24.0	24.0	36.0
Female learners: watching clips/movies	Always	16	64.0	64.0	100.0

The gender-based frequency results showed that male learners reported slightly stronger use of English clips and movies than female learners. Among male participants, 80% selected "always," while 20% selected "often." Among female participants, 64% selected "always," 24% selected "often," 8% selected "sometimes," and 4% selected "rarely." Although this descriptive pattern suggests that male learners reported more consistent use of clips and movies for vocabulary learning, the chi-square result indicated that this difference was not statistically significant. Therefore, watching English clips and movies was considered a broadly preferred strategy among both male and female learners.

Based on the exploratory results, watching English clips and movies was selected as the main instructional input for the intervention. Before the main treatment phase, a pilot study was conducted to examine whether short English clips used through a spaced learning procedure could support delayed vocabulary learning. Thirteen learners participated in the pilot phase and completed delayed recall and recognition tests 30 days after their final exposure to the clip. As presented in Table 3, the recall scores ranged from 14 to 20, with a mean of 17.962 and a standard deviation of 2.1454. The recognition scores ranged from 11 to 20, with a mean of 18.308 and a standard deviation of 2.5621. Since the recognition test was administered in multiple-choice format, its slightly higher mean was expected. The reliability coefficient for the two pilot test scores was also satisfactory, with Cronbach's alpha = .794 and standardized alpha = .802.

Table 3

Pilot Study Delayed Recall and Recognition Scores

Code Name / Gender	Recall Score	Recognition Score
01 / Female	20.0	19
02 / Female	20.0	20
03 / Male	15.0	11
04 / Female	18.5	20
05 / Female	14.5	19
06 / Male	19.0	20

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

07 / Male	20.0	20
08 / Female	19.0	18
09 / Female	19.5	19
10 / Female	18.5	19
11 / Male	17.0	19
12 / Female	14.0	15
13 / Male	18.5	19
Recall descriptive statistics	M = 17.962	SD = 2.1454
Recall score range	Min = 14.0	Max = 20.0
Recognition descriptive statistics	M = 18.308	SD = 2.5621
Recognition score range	Min = 11.0	Max = 20.0
Reliability of pilot scores	Cronbach's alpha = .794	Standardized alpha = .802

The pilot findings provided sufficient preliminary evidence for selecting English clips as the instructional medium in the main study. Accordingly, two short English clips were selected. The first clip was approximately 90 seconds long and focused on harvesting a type of fruit in a pond. The second clip was a motivational video encouraging viewers to value every second of life. The transcripts of both clips were provided to the learners so that they could read the dialogue and underline new lexical items. The same clips and target vocabulary were then used in the Massed, Spaced, and Hybrid treatment conditions.

Before the final delayed analyses, preliminary score distributions were examined for the three instructional conditions. In the Massed group, the displayed pre-test records included 19 learners, and the recall and recognition means were 16.632 and 16.474, respectively. The reliability coefficient for the Massed group score set was satisfactory, with Cronbach's alpha = .793 and standardized alpha = .795. In the Hybrid group, the displayed pre-test records included 21 learners, and the mean recall and recognition scores were 17.429 and 18.238, respectively. In the Spaced group, the displayed score records included 16 learners, and the mean recall and recognition scores were 14.938 and 17.375, respectively. Across these preliminary data, recognition scores were generally higher than recall scores, which was expected because recognition tasks required learners to identify the correct answer, whereas recall tasks required direct production of meaning.

Table 4

Preliminary Recall and Recognition Statistics across the Three Practice Conditions

Group / Phase	Measure	Valid N	Mean	Median	Mode	SD	Minimum	Maximum
Massed	Recall	19	16.632	18.000	20.0	3.5895	—	—
Massed	Recognition	19	16.474	18.000	20.0	3.9492	—	—
Hybrid pre-test	Recall	21	17.429	17.000	17.0	2.2377	12.0	20.0
Hybrid pre-test	Recognition	21	18.238	19.000	20.0	2.3001	13.0	20.0
Spaced	Recall	16	14.938	—	—	3.8204	6.5	20.0
Spaced	Recognition	16	17.375	—	—	2.6300	13.0	20.0
Massed score reliability	Recall and recognition combined	—	—	—	—	Cronbach's alpha = .793	Standardized alpha = .795	2 ite

The final delayed vocabulary tests were administered after a delay period of approximately 75 to 100 days. The delayed testing procedure was designed as an unexpected vocabulary assessment in order to reduce the possibility that participants would prepare through short-term cramming. Because the final delayed vocabulary outcome was calculated as a retention score derived from recall and recognition performance, the first inferential analysis was conducted using one-way ANOVA. The independent variable was instructional method with three levels, namely Massed, Spaced, and Hybrid practice, and the dependent variable was delayed vocabulary retention.

Before conducting the one-way ANOVA, normality and homogeneity assumptions were examined. The case processing summary showed that 38 learners from the Massed group and 38 learners from the Spaced group completed the delayed retention test, while two learners were missing from each group. In the Hybrid group, 32 learners completed the delayed retention test, and eight learners were missing. Therefore, the final retention analysis

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

included 108 valid cases. The Kolmogorov–Smirnov results did not indicate significant deviations from normality in the three groups. However, the Shapiro–Wilk result for the Spaced group was significant, p = .040, suggesting some deviation from normality for that group. Levene's test was significant, p < .001, indicating that the assumption of homogeneity of variances was not met. For this reason, Games–Howell post hoc comparisons were used because this procedure is appropriate when group variances are unequal.

Table 5

Assumption Testing for Final Delayed Retention Scores

Assumption / Group	Valid N	Missing N	Kolmogorov–Smirnov Statistic	K–S Sig.	Shapiro–Wilk Statistic	S–W Sig.
Massed	38	2	.061	.200	.983	.809
Spaced	38	2	.133	.088	.939	.040
Hybrid	32	8	.106	.200	.976	.678
Levene's test based on mean	—	—	11.548	.000	df1 = 2	df2 = 105
Levene's test based on median	—	—	10.235	.000	df1 = 2	df2 = 105
Levene's test based on median with adjusted df	—	—	10.235	.000	df1 = 2	df2 = 85.562
Levene's test based on trimmed mean	—	—	11.321	.000	df1 = 2	df2 = 105

The one-way ANOVA revealed a statistically significant difference in delayed vocabulary retention across the three instructional groups, F(2, 105) = 18.854, p < .001. The descriptive results showed that the Massed group obtained the lowest retention mean, M = 11.8355, SD = 3.72459. The Spaced group obtained a higher mean, M = 15.0789, SD = 3.10302, and the Hybrid group obtained the highest mean, M = 15.9219, SD = 1.48030. The total mean across all groups was 14.1875, with a standard deviation of 3.45078. These findings indicate that delayed vocabulary retention differed significantly according to practice condition, with Spaced and Hybrid practice producing stronger delayed retention than Massed practice.

Table 6

One-Way ANOVA and Descriptive Statistics for Final Delayed Retention

Analysis	Source / Group	Sum of Squares	df	Mean Square	F	Sig.	Mean	SD
ANOVA	Between groups	336.663	2	168.332	18.854	.000	—	—
ANOVA	Within groups	937.477	105	8.928	—	—	—	—
ANOVA	Total	1274.141	107	—	—	—	—	—
Descriptive statistics	Massed	—	—	—	—	—	11.8355	3.72459
Descriptive statistics	Spaced	—	—	—	—	—	15.0789	3.10302
Descriptive statistics	Hybrid	—	—	—	—	—	15.9219	1.48030
Descriptive statistics	Total	—	—	—	—	—	14.1875	3.45078

The Games–Howell post hoc comparisons showed that the Massed group differed significantly from both the Spaced and Hybrid groups. The mean difference between Massed and Spaced practice was -3.24342, p < .001, indicating that the Spaced group outperformed the Massed group in delayed retention. The mean difference between Massed and Hybrid practice was -4.08635, p < .001, indicating that the Hybrid group also outperformed the Massed group. However, the difference between Spaced and Hybrid practice was not statistically significant, mean difference = -.84293, p = .306. Therefore, although the Hybrid group obtained the highest mean, its performance was statistically comparable to that of the Spaced group. A non-parametric Kruskal–Wallis pairwise comparison supported the same conclusion. The Massed–Spaced and Massed–Hybrid comparisons were significant after adjustment, whereas the Spaced–Hybrid comparison was not significant.

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

Table 7

Post Hoc and Non-Parametric Pairwise Comparisons for Final Delayed Retention

Test	Comparison	Mean Difference / Test Statistic	Std. Error	Sig.	Adjusted Sig.	95% CI Lower	95% CI Upper
Games–Howell	Massed – Spaced	-3.24342	.78642	.000	—	-5.1256	-1.3612
Games–Howell	Massed – Hybrid	-4.08635	.65844	.000	—	-5.6767	-2.4960
Games–Howell	Spaced – Massed	3.24342	.78642	.000	—	1.3612	5.1256
Games–Howell	Spaced – Hybrid	-.84293	.56733	.306	—	-2.2095	.5237
Games–Howell	Hybrid – Massed	4.08635	.65844	.000	—	2.4960	5.6767
Games–Howell	Hybrid – Spaced	.84293	.56733	.306	—	-.5237	2.2095
Kruskal–Wallis pairwise	Massed – Spaced	-27.974	7.181	.000	.000	—	—
Kruskal–Wallis pairwise	Massed – Hybrid	-33.437	7.510	.000	.000	—	—
Kruskal–Wallis pairwise	Spaced – Hybrid	-5.464	7.510	.467	1.000	—	—

The second inferential analysis examined both delayed retention and retrieval performance simultaneously. Since the study included two dependent variables, namely retention score and retrieval score, a one-way MANOVA was conducted with instructional group as the independent variable. Retrieval was measured through two speed-test formats, written and oral, and the average retrieval score was entered into the analysis. Before conducting MANOVA, normality and multicollinearity were examined. For retrieval scores, the Shapiro–Wilk test was non-significant for the Massed group, $p = .573$, and the Hybrid group, $p = .148$, but significant for the Spaced group, $p = .020$. For retention scores, the same pattern was observed, with the Spaced group showing a significant Shapiro–Wilk result, $p = .040$. Therefore, because at least one group showed deviation from normality, Spearman-based interpretation was considered appropriate for examining the relationship between variables. The correlation between retention and retrieval was positive and statistically significant, $r = .556$, $p < .001$. This correlation was moderate and below the .70 threshold usually considered problematic for multicollinearity. Therefore, the two dependent variables were sufficiently related to justify MANOVA but not so strongly related as to represent the same construct.

Table 8

MANOVA Assumptions, Descriptive Statistics, and Multivariate Effects for Retention and Retrieval

Analysis	Variable / Effect / Group	Statistic / Mean	df / N	Sig. / SD	Partial Eta Squared
Retrieval normality	Massed, Kolmogorov–Smirnov	.110	38	.200	—
Retrieval normality	Massed, Shapiro–Wilk	.976	38	.573	—
Retrieval normality	Spaced, Kolmogorov–Smirnov	.158	38	.018	—
Retrieval normality	Spaced, Shapiro–Wilk	.930	38	.020	—
Retrieval normality	Hybrid, Kolmogorov–Smirnov	.132	32	.169	—
Retrieval normality	Hybrid, Shapiro–Wilk	.950	32	.148	—
Retention–retrieval association	Pearson correlation	.556	108	.000	—
Descriptive statistics	Retention, Massed	11.8355	38	3.72459	—
Descriptive statistics	Retention, Spaced	15.0789	38	3.10302	—
Descriptive statistics	Retention, Hybrid	15.9219	32	1.48030	—
Descriptive statistics	Retention, Total	14.1875	108	3.45078	—
Descriptive statistics	Retrieval, Massed	11.1250	38	2.21539	—
Descriptive statistics	Retrieval, Spaced	12.0921	38	2.53451	—
Descriptive statistics	Retrieval, Hybrid	12.4609	32	1.02215	—
Descriptive statistics	Retrieval, Total	11.8611	108	2.13075	—
Multivariate effect	Group, Pillai's Trace	.265	F = 8.034; df = 4, 210	.000	.133
Multivariate effect	Group, Wilks' Lambda	.735	F = 8.666; df = 4, 208	.000	.143
Multivariate effect	Group, Hotelling's Trace	.361	F = 9.293; df = 4, 206	.000	.153
Multivariate effect	Group, Roy's Largest Root	.360	F = 18.918; df = 2, 105	.000	.265

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

Multivariate intercept	Pillai's Trace	.975	F = 1996.481; df = 2, 104	.000	.975
Multivariate intercept	Wilks' Lambda	.025	F = 1996.481; df = 2, 104	.000	.975
Multivariate intercept	Hotelling's Trace	38.394	F = 1996.481; df = 2, 104	.000	.975
Multivariate intercept	Roy's Largest Root	38.394	F = 1996.481; df = 2, 104	.000	.975

The MANOVA results indicated a statistically significant multivariate effect of instructional method on the combined dependent variables of retention and retrieval. Pillai's Trace was .265, $F(4, 210) = 8.034$, $p < .001$, partial $\eta^2 = .133$. Wilks' Lambda also confirmed the same result, $\Lambda = .735$, $F(4, 208) = 8.666$, $p < .001$, partial $\eta^2 = .143$. These findings show that the three instructional methods differed significantly when delayed vocabulary retention and retrieval speed were considered together.

The descriptive statistics showed that, for retention, the Massed group had the lowest mean score, M = 11.1250 for retrieval and M = 11.8355 for retention. The Spaced group obtained higher scores, with M = 12.0921 for retrieval and M = 15.0789 for retention. The Hybrid group obtained the highest scores, with M = 12.4609 for retrieval and M = 15.9219 for retention. Therefore, the general descriptive pattern showed that Spaced and Hybrid practice produced stronger outcomes than Massed practice, particularly for delayed vocabulary retention. Although the Hybrid group had the highest mean in both retention and retrieval, the difference between Hybrid and Spaced practice was small.

Games–Howell post hoc results for the MANOVA-related group comparisons showed that, for retention, Massed practice differed significantly from Spaced practice, mean difference = -3.2434, $p < .001$, and from Hybrid practice, mean difference = -4.0863, $p < .001$. However, Spaced and Hybrid practice did not differ significantly from each other, mean difference = -.8429, $p = .306$. For retrieval, the Massed and Hybrid groups differed significantly, mean difference = -1.3359, $p = .005$, indicating better retrieval performance in the Hybrid group. However, the difference between Massed and Spaced practice was not significant, mean difference = -.9671, $p = .187$, and the difference between Spaced and Hybrid practice was also not significant, mean difference = -.3688, $p = .692$. Therefore, the strongest and most consistent group differences were observed for delayed retention, while retrieval differences were more limited and mainly favored Hybrid practice over Massed practice.

Overall, the findings demonstrated that practice schedule had a significant effect on vocabulary learning outcomes after a long delay. Massed practice was associated with the weakest delayed retention and retrieval performance. Spaced practice substantially improved long-term retention compared with Massed practice, and Hybrid practice produced the highest descriptive means across both retention and retrieval. Nevertheless, the statistical comparisons showed that Hybrid and Spaced practice were not significantly different in delayed retention, suggesting that distributed exposure was the central factor supporting long-term vocabulary maintenance. The significant difference between Hybrid and Massed practice in retrieval also suggests that combining initial mastery-based exposure with later spaced repetition may contribute to faster access to lexical knowledge after delay.

4. Discussion and Conclusion

The present study investigated the effectiveness of AI-supported massed, spaced, and hybrid practice for enhancing long-term vocabulary retention and retrieval among Iranian EFL learners. The findings revealed several important patterns. First, the exploratory phase demonstrated that learners overwhelmingly perceived watching English clips and movies as one of the most effective approaches for learning new vocabulary. Second, the pilot study showed that vocabulary learned through clip-based instruction could be retained successfully after a delayed period. Third, the main experimental findings demonstrated significant differences among the three instructional conditions in terms of delayed vocabulary retention. Both the spaced and hybrid groups significantly outperformed the massed group, while no statistically significant difference was found between the spaced and hybrid groups. Finally, the multivariate analyses indicated that instructional method significantly influenced the combined outcomes of retention and retrieval, with hybrid practice producing the highest descriptive means and massed practice producing the lowest scores.

The first notable finding concerns learners' strong preference for English clips and movies as a vocabulary learning tool. More than two-thirds of participants reported that they always relied on audiovisual materials for learning new lexical items. This result is consistent with contemporary views of vocabulary acquisition that emphasize meaningful contextual exposure rather than isolated memorization. Vocabulary learning becomes more efficient when learners encounter words within authentic communicative situations that provide semantic, visual, and

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

situational support (Ford-Connors & Paratore, 2015; Wright & Cervetti, 2017). The popularity of clips among participants may be explained by the fact that audiovisual materials provide multiple channels of information simultaneously, allowing learners to connect words with actions, emotions, and contextual cues. Such contextualized learning environments align with situational language teaching principles, which argue that language is learned more effectively when presented in meaningful contexts rather than through decontextualized instruction (Huang & Liu, 2024). Furthermore, the findings support previous evidence indicating that Iranian learners frequently adopt media-based vocabulary learning strategies when opportunities for authentic communication are limited (Amirian & Heshmatifar, 2013).

The pilot study findings further reinforced the value of clip-based vocabulary learning. Learners achieved relatively high delayed recall and recognition scores after exposure to short English clips. These outcomes suggest that multimedia input can create durable memory traces capable of surviving substantial periods without review. The result can be interpreted through theories of meaningful learning and cognitive integration. According to meaningful learning theory, information that is connected to existing cognitive structures is retained more effectively than information learned through rote memorization (Bryce & Blown, 2023). Video clips naturally facilitate this process because they provide visual narratives and contextual frameworks that help learners establish meaningful links between lexical forms and semantic representations. Moreover, predictive processing theories suggest that learners actively generate expectations and interpretations while processing incoming information, which may strengthen memory formation when vocabulary is embedded in dynamic audiovisual contexts (Clark, 2013).

The most important finding of the study was the superiority of spaced and hybrid practice over massed practice in long-term vocabulary retention. The ANOVA results clearly demonstrated that both spaced and hybrid groups obtained significantly higher delayed retention scores than the massed group. This finding strongly supports the extensive literature on the spacing effect. Since the pioneering work of Ebbinghaus, memory research has consistently shown that distributed practice produces stronger long-term retention than concentrated practice sessions (Roediger, 2015). When learners revisit information after intervals of forgetting, retrieval becomes more effortful, and this increased retrieval effort strengthens memory consolidation. The present findings are therefore consistent with studies showing positive effects of spaced learning on vocabulary acquisition among EFL learners (Lotfolahi & Salehi, 2017). They also align with research demonstrating that distributed review schedules enhance delayed recall in educational settings (Zigterman et al., 2015).

The superiority of spaced practice can also be interpreted through memory consolidation theories. Learning does not end when information is initially encoded. Instead, newly acquired knowledge must undergo stabilization and consolidation processes before it becomes durable. Research on vocabulary learning and memory consolidation suggests that repeated exposure across time allows lexical representations to become progressively stronger and more resistant to forgetting (Soleimannejad et al., 2022). The delayed retention advantage observed in the spaced group may therefore reflect more effective consolidation processes than those achieved through massed exposure. Similarly, educational applications of Ebbinghaus's learning curve suggest that strategically distributed review sessions can counteract the natural forgetting process and promote long-term retention (Nang, 2018). The present findings provide further support for this principle in the context of AI-supported vocabulary instruction.

Although the hybrid group achieved the highest descriptive retention mean, the difference between hybrid and spaced practice was not statistically significant. This finding is theoretically meaningful because it suggests that distributed exposure may represent the primary mechanism responsible for durable vocabulary retention. While hybrid practice incorporated an initial mastery phase through massed exposure, the subsequent spaced component appears to have contributed most substantially to long-term memory performance. In other words, once learners reached a sufficient level of initial familiarity, continued spaced retrieval and contextual recycling likely became the dominant factor influencing retention. This interpretation aligns with retrieval-based accounts of learning, which emphasize that repeated successful retrieval across intervals strengthens memory more effectively than repeated exposure alone (Carpenter & Mueller, 2013). The finding also suggests that the additional benefits of massed practice may diminish over time when delayed retention is the primary outcome.

The use of AI-generated stories within the spaced and hybrid treatments may have contributed substantially to the positive outcomes observed in both groups. Rather than

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

repeatedly encountering vocabulary in identical contexts, learners experienced target words in multiple AI-generated narratives. This repeated contextual variation likely strengthened lexical networks and expanded semantic associations. Contemporary research on AI-assisted language learning suggests that artificial intelligence can enhance learner engagement, motivation, and interaction with instructional materials (Yang et al., 2025; Yuan & Liu, 2025). Furthermore, AI-based educational environments have been associated with increased learner self-efficacy and persistence, both of which contribute to successful learning outcomes (Shi & Zhang, 2025). The present findings extend this literature by suggesting that AI-generated materials may also facilitate long-term vocabulary retention when integrated into structured spaced-learning schedules.

The retrieval findings provide additional insights into vocabulary learning processes. Although the MANOVA results showed stronger effects for retention than retrieval, hybrid practice demonstrated a significant advantage over massed practice in retrieval performance. Retrieval speed represents a particularly demanding aspect of lexical mastery because learners must access word meanings rapidly and accurately under time constraints. Faster retrieval indicates that lexical information has become more accessible within memory networks. The advantage observed for hybrid practice may be explained by the combined influence of intensive initial exposure and subsequent distributed retrieval opportunities. Initial massed practice may strengthen familiarity and recognition, while later spaced encounters may improve accessibility and automaticity. This combination could explain why hybrid learners achieved the highest retrieval means despite the absence of significant differences between hybrid and spaced practice in overall retention.

The moderate positive relationship between retention and retrieval also deserves attention. Although the correlation was statistically significant, it was not sufficiently strong to indicate that both variables measured exactly the same construct. This suggests that learners who retained vocabulary effectively were generally able to retrieve it more quickly, but retention and retrieval remained partially independent outcomes. Such a finding is theoretically reasonable because knowing a word and accessing it rapidly are related yet distinct aspects of lexical competence. Bloom's revised taxonomy differentiates between various cognitive levels, and successful vocabulary learning requires not only remembering information but also accessing and applying it efficiently (Krathwohl, 2022). Consequently, instructional approaches should aim to improve both retention and retrieval rather than focusing exclusively on one outcome.

The findings may also be interpreted through broader theories of language acquisition. Krashen's perspective emphasizes the importance of comprehensible and meaningful input in language learning (Hassan, 2022). The video clips and AI-generated stories used in the present study provided learners with repeated comprehensible input that was contextualized and engaging. Similarly, Vygotskian approaches highlight the role of mediated learning and instructional tools in cognitive development (Nardo, 2021; Ofori-Attah, 2021). In this study, video clips, transcripts, AI-generated stories, and structured practice schedules functioned as mediational tools that facilitated vocabulary acquisition. These tools supported learners as they progressed from initial exposure to long-term mastery, demonstrating the value of carefully designed instructional scaffolding.

Another noteworthy finding concerns the absence of a meaningful relationship between gender and preference for clip-based learning. Although descriptive differences existed, statistical analyses indicated that both male and female learners viewed clips and movies as beneficial vocabulary-learning resources. This finding corresponds with studies suggesting that gender differences in language learning are often smaller than commonly assumed and may depend heavily on contextual factors (Kheder & Rouabhia, 2023). Similarly, investigations of learning styles, gender, and educational outcomes have shown that instructional effectiveness is frequently influenced more by pedagogical design than by gender itself (Marantika, 2022). Therefore, the effectiveness of clip-based and AI-supported vocabulary instruction appears to be broadly applicable across learners regardless of gender.

The findings also have implications for technology-enhanced language learning. Previous studies have shown that digital systems such as web-based vocabulary platforms can facilitate vocabulary acquisition by supporting repeated review and learner engagement (Gao et al., 2015). The present study demonstrates that AI-generated content may further extend these benefits by enabling the creation of customized materials that recycle target vocabulary across diverse contexts. At the same time, the findings support recent discussions emphasizing that AI should be used as a pedagogical tool rather than an autonomous instructional agent (Zou et al., 2026). The success of the intervention likely depended not only on AI-generated materials

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

themselves but also on the systematic organization of those materials within evidence-based learning schedules.

Overall, the results suggest that vocabulary learning is most successful when meaningful multimedia input, repeated contextual exposure, retrieval practice, and AI-supported lexical recycling are combined within distributed instructional schedules. Massed practice may provide rapid short-term familiarity, but its benefits appear less durable over extended periods. In contrast, spaced and hybrid approaches provide learners with repeated opportunities to reconstruct and strengthen lexical knowledge, resulting in superior long-term retention and more efficient retrieval. Consequently, the present study contributes to the growing body of evidence demonstrating that effective vocabulary instruction requires attention not only to what learners study but also to how and when learning experiences are distributed across time.

Several limitations should be acknowledged. First, the study was conducted with intermediate Iranian EFL learners from language institutes in a single geographical region, which may limit the generalizability of the findings to learners with different proficiency levels, educational backgrounds, or cultural contexts. Second, although the delayed testing period was considerably long, the study did not include multiple delayed assessments that could provide a more detailed picture of forgetting and retention patterns over time. Third, the intervention focused primarily on receptive and productive vocabulary knowledge and retrieval speed, while other aspects of lexical competence, such as collocational knowledge, pragmatic usage, and lexical fluency in spontaneous communication, were not examined. Finally, the AI-generated stories were reviewed for instructional purposes, but the study did not systematically evaluate the linguistic characteristics of the generated texts, which may have influenced learning outcomes.

Future studies should examine the effectiveness of AI-supported spaced and hybrid vocabulary learning across different proficiency levels, age groups, and educational settings. Researchers may also compare different forms of artificial intelligence, such as conversational agents, adaptive tutoring systems, and generative language models, to determine which technologies best support vocabulary retention and retrieval. Longitudinal studies involving multiple delayed testing intervals would provide valuable insights into the durability of vocabulary learning over extended periods. Additionally, future investigations could explore how learner variables such as motivation, self-efficacy, learning styles, and digital literacy interact with AI-supported instructional approaches. Examining the effects of these methods on broader aspects of lexical competence, including collocations, phraseology, speaking fluency, and academic vocabulary development, would further enrich understanding of vocabulary acquisition processes.

Language teachers should consider incorporating short authentic video clips into vocabulary instruction because such materials appear highly engaging and facilitate meaningful lexical learning. Rather than relying exclusively on intensive review sessions, educators should distribute vocabulary practice across multiple learning occasions to promote long-term retention. AI-generated stories and supplementary materials can be used to recycle target vocabulary in varied contexts, allowing learners to encounter words repeatedly without excessive repetition of the same content. Curriculum designers should integrate retrieval-based activities, delayed review sessions, and contextualized multimedia resources into vocabulary programs. Language institutes may also benefit from adopting technology-enhanced learning environments that combine human instruction with carefully supervised AI-generated materials. Finally, teachers should emphasize long-term mastery rather than immediate test performance by providing learners with structured opportunities to revisit, retrieve, and apply vocabulary across extended periods of time.

Authors' Contributions

Authors equally contributed to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We hereby thank all participants for agreeing to record the interview and participate in the research.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in studies involving human participants were under the ethical standards of the institutional and, or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

Amalia, R. D., & Anggraeni, A. (2021). Limitation Vocabulary Is a Factor That Students Do Not Want to Speak. *Professional Journal of English Education*, *4*(1). https://journal.ikipsiliwangi.ac.id/index.php/project/article/view/7328

Amirian, S. M. R., & Heshmatifar, Z. (2013). A Survey on Vocabulary Learning Strategies: A Case of Iranian EFL University Students. *Journal of Language Teaching and Research*, *4*(3), 636-641. https://doi.org/10.4304/jltr.4.3.636-641

Askari, M. I. (2014). The Effect of Metacognitive Vocabulary Strategy Training on the Breadth of Vocabulary Knowledge. *Journal of Academic and Applied Studies*, *4*(5), 1-20. https://www.academia.edu/download/35901365/V4I5-025.pdf

Bryce, T. G. K., & Blown, E. J. (2023). Ausubel's Meaningful Learning Re-Visited. *Current Psychology*. https://doi.org/10.1007/s12144-023-04440-4

Cao, B., Tang, Q., Lin, H., Han, X., Chen, I., Wang, T., & Sun, L. (2023). Retentive or Forgetful? Diving into the Knowledge Memorizing Mechanism of Language Models. https://arxiv.org/abs/2305.09144

Carpenter, S. K., & Mueller, F. E. (2013). The Effects of Interleaving versus Blocking on Foreign Language Pronunciation Learning. *Memory & Cognition*, *41*, 671-682. https://doi.org/10.3758/s13421-012-0291-4

Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and brain sciences*, *36*(3), 181-204. https://doi.org/10.1017/S0140525X12000477

Creswell, J. W. (2018). *Research Design: Qualitative and Quantitative Approach* (2nd ed.). SAGE Publications. https://doi.org/10.4135/9781849208956

Ford-Connors, E., & Paratore, J. R. (2015). Vocabulary Instruction in Fifth Grade and Beyond: Sources of Word Learning and Productive Contexts for Development. *Review of Educational Research*, *85*, 50-91. https://doi.org/10.3102/0034654314540943

Gao, Y., Tasir, Z., Harun, J., & Jumaat, N. F. (2015). Learning English Vocabulary Using Web-Based Leitner Box with Social Network. *Jurnal Teknologi*. https://jurnalteknologi.utm.my/index.php/jurnalteknologi/article/view/6764

Hassan, M. (2022). Krashen's Monitor Model in L2 Acquisition: A Critical Review. *Journal of Ultimate Research and Trends in Education*. https://journal.unesa.ac.id/index.php/jure

Hassani, M. T., & Sadeghi, B. (2014). The Effect of Teaching Different Genres on Listening Comprehension Performance of Iranian EFL Students. *Journal of Language Teaching and Research*, *5*(3), 517-523. https://doi.org/10.4304/jltr.5.3.517-523

Huang, X., & Liu, Y. (2024). Empirical Analysis of Vocabulary Teaching in Middle School Based on Situational Language Teaching. *Saudi Journal of Humanities and Social Sciences*, 45-54. https://saudijournals.com/journal/sjhss/home

Khalili, S. M. (2016a). The Effect of Orthographic Knowledge on Word Identification and Reading Comprehension of Iranian EFL Learners. *International Journal of Applied Linguistics & English Literature*, *5*(3), 152-160. https://doi.org/10.7575/aiac.ijalel.v.5n.3p.152

Khalili, S. M. (2016b). The Effect of Vocabulary Cluster on Iranian Intermediate EFL Learners' Vocabulary Achievement. *International Journal of Applied Linguistics & English Literature*, *4*(1), 210-215. https://doi.org/10.7575/aiac.ijalel.v.4n.1p.2099

Khamesipour, M. (2015). The Effects of Explicit and Implicit Instruction of Vocabulary through Reading on EFL Learners' Vocabulary Development. *Theory and Practice in Language Studies*, *5*(8), 1620-1627. https://doi.org/10.17507/tpls.0508.11

Kheder, K., & Rouabhia, R. (2023). Gender Differences in Learning Languages. *European Journal of Applied Linguistics Studies*, *6*(2), 97-110. https://oapub.org/lit/index.php/EJALS/article/view/581

Krathwohl, D. (2022). A Version of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 217. https://doi.org/10.1207/s15430421tip4104_2

Lotfolahi, A. R., & Salehi, H. (2017). Spacing Effects in Vocabulary Learning: Young EFL Learners in Focus. *Cogent Education*, *4*(1), 1-10. https://doi.org/10.1080/2331186X.2017.1287391

Marantika, J. E. R. (2022). The Relationship between Learning Styles, Gender and Learning Outcomes. *Cypriot Journal of Educational Sciences*, *17*(1), 56-67. https://doi.org/10.18844/cjes.v17i.6681

Nang, N. H. (2018). Application of Learning Curve of Hermann Ebbinghaus and The First Twenty Hours Theory of Josh Kaufman to Distribute Teaching Targets in Some Subjects at Mien Tay Construction University. *International Journal of Advanced Research*, *6*(10), 1265-1272. https://doi.org/10.21474/IJAR01/7933

Nardo, A. (2021). Exploring a Vygotskian Theory of Education and Its Evolutionary Foundations. *Educational Theory*, *71*(3), 331-352. https://doi.org/10.1111/edth.12485

Ofori-Attah, K. (2021). Early Childhood Education and Instructional Ecology: A Vygotskian Approach in Teaching Early Childhood Social Studies. *Athens Journal of Education*, *8*(2), 139-160. https://doi.org/10.30958/aje.8-2-2

Roediger, H. L. (2015). Remembering Ebbinghaus. *Contemporary Psychology: A Journal of Reviews*, *30*(7), 719-725. https://doi.org/10.1037/023895

Seth, A. K. (2013). Interoceptive Inference, Emotion, and the Embodied Self. *Trends in Cognitive Sciences*, *17*(11), 565-573. https://doi.org/10.1016/j.tics.2013.09.007

Shi, S., & Zhang, H. (2025). EFL students' motivation predicted by their self-efficacy and resilience in artificial intelligence (AI)-based context: From a self-determination theory perspective. *Learning and Motivation*, *91*, 102151. https://doi.org/10.1016/j.lmot.2025.102151

Soleimannejad, M., Mirzazadeh, M., & Radmanesh, N. (2022). Positive Effect of Sleep on Memory Consolidation and Learning New English Words in Persian Language Speakers.

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

Original Article, 148-152. https://doi.org/10.1724/smr02022.01375

VanPatten, B., & Benati, A. G. (2015). *Key Terms in Second Language Acquisition*. Bloomsbury Academic. https://www.bloomsbury.com/uk/key-terms-in-second-language-acquisition-9781474251205/

Wang, C. (2023). The Role of Intra- and Inter-Group Matthew Effect in the Social Dilemma of Public Goods Games. https://arxiv.org/abs/2305.09144

Wright, T. S., & Cervetti, G. N. (2017). A Systematic Review of the Research on Vocabulary Instruction That Impacts Text Comprehension. *Reading Research Quarterly*, *52*, 203-226. https://doi.org/10.1002/rrq.163

Yang, N., Gao, N., & Zhang, T. (2025). Artificial Intelligence in EFL Writing Enhancement. *International Journal of Web-Based Learning and Teaching Technologies*, *20*(1), 1-40. https://doi.org/10.4018/ijwltt.390442

Yuan, L., & Liu, X. (2025). The Effect of Artificial Intelligence Tools on EFL Learners' Engagement, Enjoyment, and Motivation. *Computers in human Behavior*, *162*, 108474. https://doi.org/10.1016/j.chb.2024.108474

Zeelenberg, R., De Jonge, M., Tabbers, H. K., & Pecher, D. (2015). The Effect of Presentation Rate on Foreign-Language Vocabulary Learning. *The Quarterly Journal of Experimental Psychology*, *68*, 1101-1115. https://doi.org/10.1080/17470218.2014.975730

Zigterman, J. R., Simone, P. M., & Bell, M. C. (2015). Within-Session Spacing Improves Delayed Recall in Children. *Memory*, *23*, 625-632. https://doi.org/10.1080/09658211.2014.915975

Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2026). Exploring EFL Learners' Perceived Promise and Limitations of Using an Artificial Intelligence Speech Evaluation System for Speaking Practice. https://doi.org/10.31219/osf.io/ec68h

E-ISSN: 2645-3460