Journal of English Language Teaching and Learning

University of Tabriz

2026, 18, 37

Gender and Response Category in Iraqi EFL Learners' Associative Responses to AI-Generated Word Stimuli

Marzieh Mehri (Corresponding Author)

Assistant professor of Foreign Language Education Department, Salman Farsi University of Kazerun, Kazerun, Iran.

Marzieh.mehri1371@yahoo.com

Salman Hintaw Abdulhussein

Ph.D. Candidate of Foreign Language Education Department Salman Farsi University of Kazerun, Kazerun, Iran.

hintaws@gmail.com

Pouya shakerian

MA at Data and Discourse Studies, History and Social Sciences, Technical University of Darmstadt, Darmstadt, Germany.

Spouya035@gmail.com

ARTICLE INFO:

Received date:
2026.02.02
Accepted date:
2026.04.04

Print ISSN: 2251-7995
Online ISSN: 2676-6876

Keywords:

AI Generated Stimuli, Gender, Response Category, Vocabulary Learning, Word Association.

Abstract

This study investigated the word associative behavior of Iraqi EFL leaners to expand the current understanding of different types and classifications of the responses to word association tests (WATs) held with the help of AI. The participants' response behavior was investigated based on the response forms (nouns, verbs, adjectives, adverbs, and prepositions), categories (syntagmatic, paradigmatic, encyclopedic and clang), and quality (concrete versus abstract), regarding the possible role of gender. Besides, the differences in responses considering language proficiency were investigated. Eighty-one 19- to 26-year-old adults (36 males and 45 females), studying in different branches of medicine with two levels of English proficiency (upper intermediate and advanced) were chosen randomly. In this correlational study, after checking their proficiency with the Quick Oxford placement test (QOPT), a word association test, inspired by Wolter's (2002) and designed by AI, was administered. For both male and female learners, a consistent pattern was observed across response forms and categories with no significant variation in the distribution of responses. Besides, for both genders, the frequency of the abstract responses was more than the concrete ones. Finally, both advanced and upper-intermediate participants exhibited a stable pattern across response categories, with syntagmatic responses being the most frequent. The result revealed that both gender and proficiency may not play any major role in influencing the participants' word associative behavior in response to AI-generated stimuli. The findings have implications for the sequence of teaching vocabulary items.

Citation: Mehri, M.; Hintaw Abdulhussein, S. & Shakerian, P. (2026). Gender and Response Category in Iraqi EFL Learners' Associative Responses to AI-Generated Word Stimuli. *Journal of English Language Teaching and Learning*, 18 (37), 357-376. DOI: 10.22034/elt.2026.71388.2847

Introduction

Research on word association, which originated in the modern era in the late nineteenth century, has evolved into a substantial and varied field that incorporates both psycholinguistic and applied linguistic approaches (Fitzpatrick & Thwaites, 2020). Word association test is a usual method in psychology employed to elicit the private world of a person. In its easiest and simplest form, a group of lexical stimuli are presented orally or written to the subjects, who are instructed to respond with the first lexical item that occurs to them. It reveals the participants' understanding of "which word goes with what (Rapp, 2002). Despite variations in the specific words elicited, Word Association Tests (WATs) consistently reveal uniform patterns in lexical selection and response organization (McCarthy, 1990). Regarding the use of WATs, Wimhurst (2022) states that although they are popular in research for assessing L2 English vocabulary, they are not commonly used by language teachers for proficiency testing. According to Zhang and Koda (2017), WATs measure the vocabulary depth of L2 learners by focusing on their network knowledge. However, they argue that test results are not absolute and can fluctuate based on the test's design, how it is given and scored, and the profile of the test-takers.

Language learners are more likely to understand and retain concepts effectively when they connect them to related words. Building associations is key to learning, as a greater number of links makes a concept easier to grasp (Bahar, 2006; Kaya & Taşdere, 2016). As established in psycholinguistics, words are encoded in memory as part of interconnected webs of related concepts. This associative structure is fundamental since a word's meaning is defined through its connections within this semantic network (Aitchison, 1994). Brown (2006) suggests a more modern metaphor for understanding the mental lexicon by comparing it with a computer. He emphasizes its ongoing modifications, the integration of new words, the creation of new linkages and the attrition of vocabulary that is not actively maintained. Tryana and Mahmud (2023) state that words association improves vocabulary by helping students memorize words through context, which ultimately increases their vocabulary skills.

Research has shown that the type of stimulus significantly influences the semantic associations it triggers. According to the study done by Saffran et al. (2003) subjects produce a markedly higher proportion of verbs (action words) when responding to pictures compared to words, and to manipulable and inanimate objects compared to non-manipulable and animate ones. The strongest tendency to elicit action words was observed for pictures of manipulable, inanimate objects. Furthermore, responses to words were more likely to be conventional dictionary-style associates than the more diverse associations generated by pictures. They conclude that pictures and words access conceptual knowledge through distinct pathways. This finding supports a model of semantic organization where information is distributed across brain networks based on how the knowledge was originally acquired.

With the expansion of the use of AI in language education, different efforts have been made to investigate their efficiency in teaching L2 vocabulary items (Jomaa, 2025). According to Attygalle et al. (2025), supplementing the keyword method with AI-generated images from systems like DALL-E provides powerful visual stimuli that enhance the memorization process, leading to greater vocabulary retention than using keywords alone. Likewise, Hsu et al. (2023) found that AI image recognition technologies are useful tools to foster vocabulary acquisition.

Besides, investigating the role of gender in language learning, researchers of psychology and applied linguistics have found there are meaningful differences in cognitive performance of male and female learners. Studies which explore gender differences pertaining to different aspects connected with vocabulary learning are abundant. The results are questionable and arguable depending on which specific aspect is analyzed. Moreover, a review by Jimenez (2010) notes different results on gender tendencies or differences. As Sunderland (2010) maintains, a careful analysis of this collection of studies permits us to judge the relationships between gender and vocabulary, but may be context and test type-specific with other factors like first language, age or second language proficiency, influencing them. Thus, gender is reported as a complicated and subtle issue. In addition, Jiménez (2003) reports that female learners employed a wider range and more effective vocabulary learning strategies than males.

Literature Review

The word association test (WAT), originally developed by Sir Francis Galton and subsequently refined by Wilhelm Wundt in the late nineteenth century (Galton, 1879), began as a psychological instrument for examining the subconscious mind before being adopted by psycholinguists to investigate the mental lexicon. According to Cremer et al. (2010), the associative patterns produced by this method are believed to reveal how words and concepts are structured and connected within the mind. They further noted that free WATs provide valuable insight into a language user's current cognitive state by indicating the relative strength of particular word relationships.

This theoretical foundation has been supported by empirical research across different languages. For instance, Kwong (2013) demonstrated that although word association methods have certain limitations, they provide valuable direct psycholinguistic insight into how our mental lexicon is organized, particularly by showing how extensively words are connected and the various types of semantic relationships that exist between concepts.

Building on this cross-linguistic perspective, recent studies have examined how these organizational patterns may differ across learner populations. Faeh (2025) investigated how English and Japanese learners organize words in their minds by analyzing the results of 720-word associations from 18 participants. The study found that younger native English speakers tended to make meaning-based associations in both their first language and limited second language, while younger native Japanese speakers with little English experience made more position-based associations in both languages. Older, experienced L2 learners showed more balanced association patterns but remained predominantly position-based. These findings challenge previous work association research limited to English speakers and student participants, suggesting that both L2 learning and different methods of L1 vocabulary instruction across educational systems may influence how people connect words in their mental lexicons.

While Faeh focused on proficiency and language background, other researchers have examined how the grammatical properties of words themselves shape associative patterns. Li and Zhang (2025) investigated how word class affects second-language Chinese word associations using 108 carefully controlled stimulus words including nouns, verbs, and adjectives selected from the Corpus of Modern Chinese. Their findings revealed that meaning-

based associations dominated across all three-word classes, with adjectives demonstrating the most developed semantic networks. Phonological associations also played an important role for all word classes, though adjectives triggered proportionally fewer of these. All three-word classes elicited substantial encyclopedic associations, with nouns showing the highest triggering rate. When examining specific association subtypes, the researchers found that nouns were primarily characterized by synonymous, determinative, and hierarchical associations; verbs by governing and synonymous associations; and adjectives by determinative, synonymous, and antonymous associations.

Thus far, studies on word association in L2 vocabulary learning have been limited to traditional methods using oral or written stimuli. To our knowledge, no study has investigated potential differences in how male and female students respond to AI-generated stimuli. Understanding such differences is significant, as it would inform teachers in designing more effective and tailored instructional sequences. To address this gap, the present study is designed to answer the following questions:

1. Do male and female Iraqi EFL learners produce word association responses with significantly different distributions of nouns, verbs, adjectives, adverbs, and prepositions in response to AI-generated verbs?
2. Is there any significant difference between the categories (syntagmatic, paradigmatic, encyclopedic and clang) of word associative responses to AI-generated verb stimuli of male and female Iraqi EFL learners?
3. Is there any significant difference between male and female Iraqi EFL learners' associative responses to concrete versus abstract stimuli generated by AI?
4. Is there any significant difference between the category of the associative responses of male and female EFL learners at different levels of proficiency?

Method

Design of The Study

This correlational study examined the relationship between two independent variables, gender (male vs. female) and language proficiency (upper-intermediate vs. advanced), and three dependent variables related to word association behavior: response form, response category, and response quality using sets of quantitative data was investigated.

Participants

The sample included 81, nineteen- to twenty-six-year-old Arab native speaker adults (36 males and 45 females), at two levels of English language proficiency, upper intermediate and advanced, studying dentistry, medicine, pharmacy and nursing randomly selected from University of Al-Ameed, Karbala, Iraq. They were required to answer the QOPT and based on the rating scale of the QOPT, they were classified as two groups of “upper intermediate”, “advanced” learners. Table 1. shows the details of number and gender of the participants.

Table 1. *Details related to the number and gender of the participants*

	Males	Females	Total
First administered	50	50	100
Outliers	14	5	19
Selected Participants	36	45	81

Materials and Instruments

Instrument 1

Proficiency Test

QOPT Version 2 was administered to select two homogeneous groups (upper intermediate and advanced learners) based on language proficiency. Scores on the test ranged from 0 to 60. The participants whose score was between 40 and 49 were classified as upper-intermediate and between 50 and 60 were classified as advanced. The participants with scores lower than 40 were considered as the outliers of the study.

Instrument 2

WAT

The main instrument was a WAT that was designed using AI-generated pictures to represent targeted verbs as visual stimuli**.** This methodology could help researchers move beyond traditional written-word prompts and provided the opportunity to investigate the role of concreteness in lexical access. To choose the targeted verbs, a list of verbs from the Edinburgh Associative Thesaurus (Kiss et al., 1973), as utilized by Wolter (2001) was chosen. This provided a validated starting point with known associative properties. Then, each selected verb was given as a prompt to a generative AI image model named DALL-E 3 with specific instructions to make an image. The examples of the target verbs are draw, jump, care, bring, move, keep, visit, fall. The final test included 20 verbs.

Procedure

The study was conducted in two different sessions to prevent participants' fatigue and ensure data quality. The QOPT was administered to all potential students passing English courses at AI-ameed university. The purpose of this test was to select participants at the upper-intermediate and advanced proficiency levels. Participants whose scores fell outside the target proficiency band were excluded from the subsequent WAT. Upper intermediate and advanced students were invited to participate in a session held online for the WAT. The online session started with introducing the test and providing necessary instructions to each student, informing them that they would be shown a series of AI-generated images depicting some verbs. They were instructed to respond to each image with the first word that came to their mind. It was emphasized that there were no right or wrong answers. In order to encode the data, the elicited responses to the WAT were classified using traditional WAT frameworks into four associative types: paradigmatic, syntagmatic, clang-other, and encyclopedic. Paradigmatic responses are in the same word class as the stimulus word, such as antonyms or synonyms, e.g. black - white, thin - skinny. Syntagmatic responses have a sequential or collocational relationship with the stimulus word, e.g. dig - hole. In clang responses, the associations are usually made based on

the physical appearance (orthographic), e.g. “bow” of a ship - “bow” and arrow, or the form of words (phonetic), e.g. bare-bear rather than their meaning. Finally, encyclopedic responses are related to the individual’s personal knowledge about the words. To keep the consistency in the classification, an obtained response was considered as paradigmatic if its lexical class was from the same as the prompt word (such as jump: increase) and as syntagmatic if its lexical class was different from the prompt word (such as jump: high), a response was also classified as syntagmatic when it could make a syntactic association with the prompt word (such as keep - up with).

Data Analysis

At first, the scores of the QPT were analyzed by Kolmogorov-Smirnov test to ensure the assumptions of normality. Then, the elicited responses of the WAT were classified and the frequency of each association type was calculated. Chi-square and Goodness of Fit analysis were administered to determine whether the observed differences between the associative response types are statistically significant.

Results

The data from the WAT were rated by the researcher. Table 4.1 shows them in detail.

Table 2. *Examples of the responses to the WAT*

		Male	**Female**
Categories	Syntagmatic	Jump: rope - Enjoy: life	Break: glass - Send: email
	Paradigmatic	Visit: meet - Argue: discuss	Show: demonstrate - Jump: leap
	Encyclopedic	Kill: Bill - Travel: North	Argue: Maryam - Kill: jail
	Clang	Make: take - Fall: doll	Write: right - Make: fake
Forms	Noun	Cut: tree - Show: TV	Fall: leaves - Visit: parents
	Verb	Visit: meet - Bring: take	Move: go - Apply: use
	Adjective	Make: sure - Keep: calm	Bring: water - Keep: mind
	Adverb	Write: soon - Move: fast	Travel: abroad - Keep: away
	Preposition	Replace: with - Send: back	Replace: with - Cut: off
Quality	Abstract	Enjoy: holiday - Show: time	Make: sure - Fall: love
	Concrete	Write: pen - Break: glass	Cut: finger - Visit: parents

Besides, Table 2 represents the most common associative responses to the WAT:

Table 3. *Highest-Frequency Responses to WAT*

Stimuli	The Most Frequent Responses				
Jump	high	rope	child	Up	fall
Care	take	baby	careless	Careful	doctor
Bring	water	food	take	Fast	back
Move	go	fast	it	Around	keep
Keep	safe	calm	mind	away	going
Visit	doctor	relatives	parents	meet	see
Fall	down	love	leaves	ground	winter
Break	glass	window	law	heart	down
Travel	trip	agency	enjoy	North	abroad
Cut	tree	finger	hair	short	bloody
Enjoy	life	holiday	happy	pleasure	time
Kill	murder	crime	enemy	hard	people
Argue	fight	discuss	angry	discussion	friend
Write	letter	down	homework	book	pen
Send	email	message	forward	back	letter
Replace	with	substitute	position	that	alternative
Apply	university	change	job	use	visa
Make	sure	mistake	friend	decision	sentence
Show	play	TV	demonstrate	film	time

Table 3. shows the classification of the responses based on the response form (noun, verb, adjective, adverb, and preposition).

Table 4. *Response Categorization Based on Form*

Stimuli	Noun	Verb	Adjective	Adverb	Preposition	Total
Draw	32	36	8	1	4	81
Jump	27	11	33	2	8	81
Care	26	11	35	4	5	81
Bring	34	26	17	0	4	81
Move	11	29	22	3	16	81
Keep	18	20	35	1	7	81
Visit	23	35	16	1	6	81
Fall	45	6	21	1	8	81
Break	41	4	19	0	17	81
Travel	17	11	42	4	7	81
Cut	32	12	26	0	11	81
Enjoy	44	16	21	0	0	81
Kill	39	23	17	2	0	81

Argue	10	33	30	1	7	81
Write	41	3	19	3	15	81
Send	32	18	10	0	21	81
Replace	49	15	3	0	14	81
Apply	39	28	4	0	10	81
Make	36	9	28	0	8	81
Show	30	34	5	1	11	81
Total	626	380	411	24	179	1620

As Table 4 demonstrates, noun responses were the highest frequent responses and adverbs the lowest. Table 4.6 shows the classification of the responses based on the response categories (syntagmatic, paradigmatic, encyclopedic, clang).

Table 5. *Classification by Response Category*

Stimuli	Syntagmatic	Paradigmatic	Encyclopedic	Clang	Total
Draw	20	56	1	4	81
Jump	57	21	2	1	81
Care	54	21	4	2	81
Bring	41	36	3	1	81
Move	55	19	4	3	81
Keep	53	22	3	3	81
Visit	24	51	4	2	81
Fall	56	20	3	2	81
Break	53	23	4	1	81
Travel	55	21	3	2	81
Cut	59	19	3	0	81
Enjoy	51	22	6	2	81
Kill	27	50	3	1	81
Argue	25	52	4	0	81
Write	55	20	4	2	81
Send	53	23	3	2	81
Replace	55	19	7	0	81
Apply	36	40	5	0	81
Make	59	12	2	8	81
Show	13	58	9	1	81
Total	901	605	77	37	1620

As shown in Table 5, syntagmatic responses were the most frequent and clang responses the least. Table 4.7. shows the classification of the responses based on the response quality (concrete and abstract).

Table 6. *Classification by Response Quality*

Stimuli	Abstract	Concrete	Total
Draw	26	55	81
Jump	28	53	81
Care	39	42	81
Bring	27	54	81
Move	55	26	81
Keep	54	27	81
Visit	24	57	81
Fall	44	37	81
Break	38	43	81
Travel	56	25	81
Cut	27	54	81
Enjoy	59	22	81
Kill	61	20	81
Argue	65	16	81
Write	21	60	81
Send	39	42	81
Replace	59	22	81
Apply	65	16	81
Make	64	17	81
Show	62	19	81
Total	913	707	1620

As Table above demonstrates, the frequency of concrete responses was higher than the abstract ones.

4.1. The results of the analysis of the response forms across gender

According to the first research question, the purpose was to find whether there was any significant difference between the forms of word associative responses of male and female EFL learners to verb stimuli. Table 7. shows the results of the frequency of associative response forms across gender.

Table 7. *Frequency of associative response forms across gender*

	Forms					
	Noun (%)	**Verb (%)**	**Adj. (%)**	**Adv. (%)**	**Prep. (%)**	**Total (%)**
Female	337(37.4)	211 (23.4)	229 (25.5)	16 (1.8)	107 (11.8)	900 (100)
Male	289 (40.2)	169 (23.4)	182 (25.3)	8 (1.1)	72 (10)	720 (100)
Total	626 (38.7)	380 (23.5)	411 (25.4)	24 (1.4)	179 (11)	1620 (100)

As Table 7. reveals, for both male and female learners, a consistent pattern was detected for all the four associative response forms. In both gender groups, noun responses were the most frequent (male: 289, female: 337), adjective responses were the second most frequent (male: 182, female: 229), and verb responses had the third lowest frequency of occurrence (male: 169,

female: 211). Preposition responses had the fourth lowest frequency (male: 72, female: 107) and finally adverb responses had the lowest frequency of occurrence (male: 8, female: 16). Totally, 626 nouns, 380 verbs, 411 adjectives, 179 prepositions, and 24 adverbs were obtained, showing that the participants had more tendencies towards answering the word association test with nouns rather than other categories. Table 4.9. shows the results of the Chi-Square tests of associative response forms across gender.

Table 8. *Chi-Square tests of response forms across gender*

	Value	**Df**	**Asymp. Sig. (2-sided)**
Pearson Chi-Square	3.248[a]	4	.517
Likelihood Ratio	3.287	4	.511
Linear-by-Linear Association	2.368	1	.124
N of Valid Cases	1620		

As shown in Table 8, the Chi-square test indicated no statistically significant difference in the frequency of response forms between male and female learners (P = .517). Moreover, the Goodness of Fit test was used to investigate the significance of differences between the observed frequency and expected frequency of the response forms, for female, male, and the total number of learners separately.

Table 9. *Goodness of Fit test for the female group's response forms*

	Female
Chi-Square	334.644[a]
Df	4
Asymp. Sig.	.000

As indicated in Table 9., the Goodness of Fit test results showed that the significance for P value was less than .05 (P = .000 < .05), showing that the differences in the frequency of response forms of female group were statistically significant. In other words, the differences between the observed and expected frequency of the response forms for female group were significant. Table 10. shows the results of the Goodness of Fit test for the male group.

Table 10. *Goodness of Fit test for the male group's response forms*

	Male
Chi-Square	324.819[a]
Df	4
Asymp. Sig.	.000

As shown in Table 10, the test results revealed that (P = .000 < .05), indicating that the differences in the frequency of response forms of male group were statistically significant. Table 11. shows the results of the Goodness of Fit test for the total number of participants.

Table 11. *Goodness of Fit test for the total number of participants' response forms*

	Total
Chi-Square	657.204[a]
Df	4
Asymp. Sig.	.000

As indicated in Table 11, the difference in the frequency of response forms among the participants was statistically significant (p = .000).

4.2. The results of the analysis of the response categories across gender

The second research question aimed to find whether there was any significant difference between the categories of word associative responses of male and female EFL learners. Table 12. demonstrates the frequency of associative response categories across gender.

Table 12. *Frequency of response categories across gender*

	Categories				**Total (%)**
	Clang (%)	**Encyclopedic (%)**	**Paradigmatic (%)**	**Syntagmatic (%)**	
Female	24 (2.6)	48 (5.4)	334 (37.2)	494 (54.8)	900 (100)
Male	13 (1.8)	29 (4)	271 (37.6)	407 (56.6)	720 (100)
Total	37 (2)	77 (4.5)	605 (37.1)	901 (56.4)	1620 (100)

As Table 12 indicates, for both male and female learners, there is an almost consistent pattern for all the four associative response categories. In both gender groups, syntagmatic responses were the most frequent (male: 407, female: 494), paradigmatic responses were the second highest frequent (male: 271, female: 334), encyclopedic responses were the third highest frequent (male: 29, female: 48), and clang responses were the least frequent (male: 13, female: 24). The results of the total of the responses show that there was a tendency towards making syntagmatic associations between the participants, rather than the other three categories of association. Table 13. indicates the Chi-Square tests of associative response categories across gender.

Table 13. *Chi-Square tests of response categories across of gender*

	Value	**Df**	**Asymp. Sig. (2-sided)**
Pearson Chi-Square	2.956[a]	3	.398
Likelihood Ratio	3.002	3	.391
Linear-by-Linear Association	1.811	1	.178
N of Valid Cases	1620		

As shown in table 13., the results of the Chi-square test revealed that the significance for P value was more than .05 (P = .398 > .05), showing that the differences in the frequency of response categories of the male and female learners were not statistically significant.The Goodness of Fit test was also used for female, male, and the total number of learners separately.

Table 14. *Goodness of Fit test for the female group's response categories*

	Female
Chi-Square	693.209[a]
Df	3
Asymp. Sig.	.000

As indicated in Table 14, there was a statistically significant difference in the frequency of response categories for the female group (Goodness-of-Fit test, $p < .001$). Table 4.16. shows the results of the Goodness of Fit test for the male group.

Table 15. *Goodness of Fit test for the male group's response categories*

	Male
Chi-Square	613.889[a]
Df	3
Asymp. Sig.	.000

As shown in Table 4.16, the results showed a statistically significant difference in the frequency of response categories for the male group ($p < .001$). This indicates that the differences between the expected and observed frequencies were significant. Table 15. shows the results of the Goodness of Fit test for the total number of participants.

Table 16. *Goodness of Fit test for the total number of participants' response categories*

	Total
Chi-Square	788.643[a]
Df	3
Asymp. Sig.	.000

As indicated in table 16., the test results demonstrated that the significance for P value was less than .05 ($P = .000 < .05$), showing that the differences in the frequency of response categories of the total number of participants were statistically significant. In other words, the differences between the observed frequency and expected frequency of the response categories for all participants were significant.

4.3. The results of the analysis of the response quality across gender

In the third research question, the purpose was to find whether there was any significant difference between male and female EFL learners' associative responses in terms of the quality of the responses (Abstract vs. Concrete). Table 4.18. demonstrates the frequency of associative abstract vs. concrete responses across gender.

Table 17. *Frequency and percentage of Abstract vs. Concrete responses across gender*

		Quality		**Total (%)**
		Abstract (%)	Concrete (%)	
	Female	506 (56.2)	394 (43.8)	900 (100)
	Male	407 (56.5)	313 (43.5)	720 (100)
Total		913 (56.3)	707 (43.7)	1620 (100)

As it is shown in Table 4.18. there is a similar pattern in the responses provided by male and female learners, the frequency of the abstract responses (male: 407, female 506) is more than the concrete ones (male: 313, female: 394). The results of the total of the responses show that the participants had a penchant for using abstract responses more than the concrete ones. Table 4.19. demonstrates Chi-Square tests of associative abstract vs. concrete responses across gender.

Table 18. *Chi-Square tests of Abstract vs. Concrete responses across gender*

	Value	**df**	**Asymp. Sig. (2-sided)**	**Exact Sig. (2-sided)**	**Exact Sig. (1-sided)**
Pearson Chi-Square	.015[a]	1	.902		
Continuity Correction[b]	.005	1	.942		
Likelihood Ratio	.015	1	.902		
Fisher's Exact Test				.920	.471
Linear-by-Linear Association	.015	1	.902		
N of Valid Cases	1620				

There was no statistically significant difference in the frequency of associative abstract and concrete responses between male and female learners, as shown in Table 18 (**P** = .902). In order to investigate the significance of differences between the observed frequency and expected frequency of the abstract and concrete responses, for female, male, and the total number of learners, the Goodness of Fit test was used (Table 19).

Table 19. *Goodness of Fit test for the female group's response quality*

	Female
Chi-Square	13.938[a]
Df	1
Asymp. Sig.	.000

As indicated in Table 19., the results showed that the significance for P value was less than .05 (P = .000 < .05), demonstrating that the differences in the frequency of response quality of female group were statistically significant. Table 20. shows the results of the Goodness of Fit test for the male group.

Table 20. *Goodness of Fit test for the male group's response quality*

	Male
Chi-Square	12.272[a]
Df	1
Asymp. Sig.	.000

Table 20 shows a statistically significant difference between the observed and expected frequencies of response qualities for the male group ($P < .001$). Table 4.22. shows the results of the Goodness of Fit test for the total number of participants.

Table 21. *Goodness of Fit test for the total number of participants' response quality*

	Total
Chi-Square	26.195[a]
Df	1
Asymp. Sig.	.000

The Goodness-of-Fit test results in Table 21 indicate a statistically significant discrepancy between the observed and expected frequencies of response qualities for the overall participant group ($P < .001$).

4.4. Analysis of the response categories across language proficiency

The fourth research question was to find whether there was any significant difference between the associative response categories of the EFL learners who have different levels of language proficiency. Table 22. shows the frequency of associative response categories across proficiency.

Table 22. *Frequency and percentage of response categories across proficiency*

	Categories				Total (%)
	Clang (%)	Encyclopedic (%)	Paradigmatic (%)	Syntagmatic (%)	
Adv.	23 (2.6)	43 (5)	320 (37.2)	474 (55.2)	860 (100)
Upper	14 (1.8)	34 (4.5)	285 (37.5)	427 (56.2)	760 (100)
Total	37 (2.3)	77 (4.7)	605 (37.3)	901 (55.7)	1620 (100)

As it is demonstrated in Table 22, for both advanced and upper-intermediate level participants, there is an almost stable pattern for all the categories. In both groups, syntagmatic responses were the most frequent (advanced: 474, upper-intermediate: 427), paradigmatic responses had the second highest frequency level (advanced: 320, upper-intermediate: 285), encyclopedic responses had the third highest frequency of level (advanced: 43, upper-intermediate: 14), and clang responses had the lowest level of frequency (advanced: 23, upper-intermediate: 14). The participants showed their tendency towards making syntagmatic associations rather other three categories of the associations. Table 23. indicates the Chi-Square tests of associative response categories across proficiency.

Table 23. *Chi-Square tests of associative response categories across proficiency*

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.551[a]	3	.671
Likelihood Ratio	1.568	3	.667
Linear-by-Linear Association	.891	1	.345
N of Valid Cases	1620		

The Chi-square test indicated that the difference in associative response categories between proficiency levels (advanced and upper-intermediate) was not statistically significant ($p = .671$) (see Table 23). To investigate the significance of differences between the observed frequency and expected frequency of the response categories, for advanced, upper-intermediate, and the total number of learners, the Goodness of Fit test was used (see Table 24).

Table 24. *Goodness of Fit test for the advanced group's response categories*

	Advanced
Chi-Square	672.344[a]
Df	3
Asymp. Sig.	.000

As indicated in table 4.25. (P = .000 < .05) which means that the differences in the frequency of response categories of advanced group were statistically significant. Table 4.26. shows the results of the Goodness of Fit test for the upper-intermediate group.

Table 25. *Goodness of Fit test for the upper-intermediate group's response categories*

	Upper
Chi-Square	634.242[a]
Df	3
Asymp. Sig.	.000

As shown in table 25. (P = .000 < .05), indicating that the differences in the frequency of response categories of upper-intermediate group were statistically significant. Table 4.27. demonstrates the results of the Goodness of Fit test for the total number of participants.

Table 26. *Goodness of Fit test for the total number of participants' response categories*

	Total
Chi-Square	1306.232[a]
Df	3
Asymp. Sig.	.000

As shown in Table 26, there was a statistically significant discrepancy between the observed and expected frequencies of response categories for the entire sample ($P < .001$).

Discussion and Conclusion

This study investigated the word association behavior of Iraqi university students to AI-generated stimuli to expand the current understanding of different types and classifications of responses. Considering the response forms, for both male and female learners, a consistent pattern was detected, noun responses had the highest frequency of occurrence for both gender groups and the differences in the frequency of associative responses of the male and female learners were not statistically significant. It was found that both gender groups were likely to chase a consistent pattern in their associative behavior for all the five associative response forms i.e., noun responses > adjective responses > verb responses > preposition responses > adverb responses. The finding that verbs are associated less frequently than nouns align with the order of acquisition for children learning English (Gentner, 1982). This finding can be explained by Gentner's (1982) cognitive account, stating that the greater cognitive complexity of verbs compared to nouns accounts for their later acquisition across diverse languages. In addition, the general tendency toward adjective and noun responses for males and females might be due to the collocational nature of adjectives and nouns, as it was mentioned in the study done by Rahimi and Haghighi (2009).

In response to the second research question, in both gender groups, syntagmatic responses were the most frequent, paradigmatic responses had the second highest frequency of occurrence, encyclopedic responses had the third highest frequency of, and clang responses had the lowest frequency. The statistics revealed that the differences in the frequency of associative responses of the males and females were not statistically significant and gender had no effect. High frequency of the syntagmatic responses is in accordance with the view that non-native English speakers produce associative responses which are mostly syntagmatic rather than paradigmatic (Deignan et al, 1996; Rahimi & Haghighi, 2009). These findings also support those which found by Wolter (2001, 61), showed a syntagmatically-dominant mental lexicon for non-native speakers. By contrast, these results, which show a preference for syntagmatic responses, are not in line with previous studies by Dóczi (2006) and Khazaeenezhad and Alibabaee (2013). In general, the literature implicates that there is a general consensus (Coulthard et al., 2000) concerning word association of a tendency for native speakers to provide paradigmatic responses for word association tests and for non-native speakers and children to provide syntagmatic ones. On the other hand, clang responses were the least frequently used type of association. It is in line with Rahimi and Haghighi's (2009) study in which they stated that clang associations are not very frequent.

As for the third question, for the responses provided by males and females, the frequency of the concrete responses was significantly more than the abstract ones but the statistics show that the differences in the frequency of associative abstract and concrete responses of the males and females were not statistically significant, therefore, gender does not play any key role here. This finding supports Bauer and Altarriba's (2008) finding that gender differences in word association are not universal but are instead dependent on the category of the cue word. This finding can be explained by the nature of the stimulus provided. In this study, participants were shown an AI-generated image instead of a written or oral word. A visual stimulus may trigger a more universal perceptual and emotional response. Consequently, the stimulus type could have elicited similar reactions from both sexes which led to the non-significant result.

Regarding the role of language proficiency in different types of associative responses in this study, it was found that language proficiency does not play a significant role. It was in line with Yolcu and Demirel's findings (2018), suggesting that proficiency level is not a determining factor in word associations. Instead, other variables including personality and personal background play a more important part in this regard. This was in contrast with Smadi (2020) who found that proficiency level has a moderate effect on word associations. This contrast might be explained by differences in the language proficiency levels which were targeted. Unlike Smadi (2020), who compared low- and high-proficiency learners), this study focused on intermediate and advanced learners, where the difference in exposure to words is less noticeable.

Overall, the results of this study can have useful implications for vocabulary teaching through helping learners in making stronger links between the different words. The associations which EFL learners make on these tests demonstrate the key role these associations may play in teaching and/or learning of new words (Schmitt, 2000). It is useful to focus on features of the entire lexicon, which might be a better alternative to retaining new words by just exposure

or employing a mnemonic technique which concentrates purely on individual characteristics of words (Meara & Wolter, 2006). Therefore, it seems to a great extent logical to help L2 learners organize their own mental lexicon by strengthening links between words. As this study was limited to Iraqi university students, it can be replicated by adults and children of different age groups with other nationalities. Besides, as the number of prompt pictures were used in this test was limited to 20, it can be increased or changed based on different criteria. The participants were also asked to provide the first word coming to their minds. They could have been asked to provide two or three associative responses alternatively, and this new design would have been much more differentiating between participants at different levels of language proficiency.

Acknowledgments (Required)

The authors have no financial or non-financial interests to declare.

References

Aitchison, J. (1994). *Words in the mind: An introduction to the mental lexicon* (2nd ed.). Blackwell.

Attygalle, N. T., Kljun, M., Quigley, A., Pucihar, K. Č., Grubert, J., Biener, V., Leiva, L. A., Yoneyama, J., Toniolo, A., Miguel, A., Kato, H., & Weerasinghe, M. (2025). Text-to-image generation for vocabulary learning using the keyword method. In T. Li, F. Paternò, K. Väänänen, L. Leiva, L. D. Spano, & K. Verbert (Eds.), *Proceedings of the 30th International Conference on Intelligent User Interfaces* (pp. 1381–1397). ACM. https://doi.org/10.1145/3643865.3657521

Bahar, M. (2006). *Geleneksel-alternatif ölçme ve değerlendirme*. Pegem A Yayıncılık.

Coulthard, M., Knowles, M., Moon, R., & Deignan A. (2000). *Lexis.* Birmingham: Centre for English Language Studies.

Bauer, L. M., & Altarriba, J. (2008). An investigation of sex differences in word ratings across concrete, abstract, and emotion words. *The Psychological Record*, *58*(3), 465–474. https://doi.org/10.1007/BF03395629

Brown, P. S. (2006). *A small-scale exploration into the relationship between word-association and learners' lexical development* [Unpublished master's thesis]. University of Birmingham.

Cremer, M., Dingshoff, D., de Beer, M., & Schoonen, R. (2010). Do word associations assess word knowledge? A comparison of L1 and L2, child and adult word associations. *International Journal of Bilingualism, 15*(2), 187–204. https://doi.org/10.1177/1367006910381189

Deignan, A., Knowles, M. Sinclair, J., Willis, D. (1998). *Lexis.* Birmingham: The University of Birmingham.

Dóczi, B. (2006). Mapping the mental lexicon of pre-intermediate learners: Word associations in a depth of word knowledge elicitation task. In M. Nikolov, & J. Horváth (Eds.), *UPRT 2006: Empirical studies in English applied linguistics* (pp. 117-137). Pécs: Lingua Franca Csoport.

Faeh, A. G. (2025). Word associations of inexperienced and experienced learners of English and Japanese: Toward a fuller understanding of lexical storage and retrieval in second language learning and teaching. *The Language Teacher*, *49*(4), 9. https://doi.org/10.37546/JALTTLT49.4-2

Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, *53*(3), 237–274. https://doi.org/10.1017/S0261444820000105

Galton, F. (1879). Psychometric experiments. *Brain, 2*(2), 149–162.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought, and culture* (pp. 38-62). Springer-Verlag.

Hsu, T. C., Chang, C., & Jen, T. (2023). Artificial intelligence image recognition using self-regulation learning strategies: Effects on vocabulary acquisition, learning anxiety, and learning behaviours of English language learners. *Interactive Learning Environments. Advance online publication*. https://doi.org/10.1080/10494820.2023.2165508

Jiménez, R. M. (2003). Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics, 13*(1), 54-77.

Jiménez, R. M. (Ed.). (2010). *Gender perspectives on vocabulary in foreign and second language.* Basingstoke: Palgrave MacMillan.

Jomaa, N. J., Attamimi, R., & Al Mahri, M. (2025). The use of artificial intelligence (AI) in teaching English vocabulary in Oman: Perspectives, teaching practices, and challenges. *World Journal of English Language*, *15*(3), 1–13. https://doi.org/10.5430/wjel.v15n3p1

Kaya, M., & Taşdere, A. (2016). İlkokul türkçe eğitimi için alternatif bir ölçme değerlendirme tekniği: kelime ilişkilendirme testi (KİT). *Turkish Studies International Periodical for The Languages, Literature and History of Turkish or Turkic*, *11*(9), 803-820. https://doi.org/10.7827/TurkishStudies.9499

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 153–165). Edinburgh University Press.

Khazaeenezhad, B., & Alibabaee, A. (2013). Investigating the role of L2 language proficiency in word association behavior of L2 learners: A case of Iranian EFL learners. *Theory and Practice in Language Studies, 3*(1), 108-115.

Kwong, O. Y. (2013). Exploring the Chinese mental lexicon with word association norms. In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)* (pp. 153–162). Department of English, National Chengchi University. https://aclanthology.org/Y13-1013

Li, M., & Zhang, L. (2025). Word class effects on L2 Chinese word associations. *SAGE Open*, *15*(1), 1. https://journals.sagepub.com/home/sgo

McCarthy, M. (1990). *Vocabulary*. Oxford University Press.

Meara, P. & Wolter, B. (2004). V_links: Beyond vocabulary depth. *Angles on the English-speaking World, 4*, 85-96.

Saffran, E. M., Coslett, H. B., & Keener, M. T. (2003). Differences in word associations to pictures and words. *Neuropsychologia*, *41*(11), 1541–1546. https://doi.org/10.1016/S0028-3932(03)00080-0

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Smadi, A. M. (2020). The role of language proficiency in word association behavior among Jordanian EFL learners. *Journal of Literature, Languages and Linguistics, 67,* 39, 39-47. http://www.iiste.org/Journals/index.php/JLLL/index

Sunderland, J. (2010). Theorizing gender perspectives in foreign and second language learning. In R. M. Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 1-22). Basingstoke: Palgrave MacMillan.

Tryana, T., & Mahmud, L. H. (2023). Vocabulary Teaching by Words Association. *JIIP (Jurnal Ilmiah Ilmu Pendidikan)*, *6*(9), 7246-7253. https://jiip.stkipyapisdompu.ac.id/jiip/index.php/JIIP/issue/view/20

Rapp, R. (2002). The computation of word associations: Comparing syntagmatic and paradigmatic approaches. In *COLING 2002: The 19th International Conference on Computational Linguistics* (pp. 1-7). https://aclanthology.org/C02-1007/

Rahimi, A., & Haghighi, H. (2009). How learners make mental links with words. *California Linguistic Notes, 34*(2), 1-26.

Wimhurst, A. (2022). Using word association tests to assess L2 vocabulary proficiency. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-6). Wiley. Https://doi.org/10.1002/9781118784235.eelt1020

Wolter, B. (2001). A depth of individual word knowledge model. *Studies in Second Language Acquisition, 23*, 41-69.

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics, 27*(4), 741-747.

Yolcu, G., & Demirel, S. (2018). *Word association analysis in different proficiency levels* [Unpublished manuscript]. Department of Foreign Language Education, Middle East Technical University.

Zhang, D., & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: Issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education*, 2*(1)*, 1–10. https://doi.org/10.1186/s40862-017-0024-0