International Journal of Linguistics, Literature and Translation
ISSN: 2617-0299 (Online); ISSN: 2708-0099 (Print)
DOI: 10.32996/ijllt
Journal Homepage: www.al-kindipublisher.com/index.php/ijllt

IJLLT
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Effects of Assisted Tools and Learning Conditions on L2 Vocabulary Learning: A Study based on Large Language Model

Long Fei[1] and Hu Yinyan[2] ✉
[1]*Associate Professor, School of Foreign Languages, Harbin University of Commerce, Harbin, China*
[2]*Graduate Student, School of Foreign Languages, Jinan University, Guangzhou, China*
Corresponding Author: Hu Yinyan, **E-mail**: hyy4991126@163.com

| ABSTRACT
In this study, we investigated the effects of assisted input, namely gloss and dictionary, on L2 incidental vocabulary learning. Furthermore, intentional and incidental learning conditions were also compared while merely utilizing a dictionary. Additionally, the gloss material was provided by a large language model (LLM) ChatGPT through prompt engineering. Besides, learning gains were measured not only solely from knowledge breadth (form-meaning connection) but also from more dimensions regarding knowledge depth (synonym discrimination, derivation production, collocation production). Sixty-four English learners of grade 2 from a senior high school were divided into three treatment groups and one control group. Those two kinds of comparisons were made respectively between every three groups. Results indicated that the gloss provided by LLM showed efficiency in collocation retention while the dictionary brought better effects in derivations and synonym discrimination. Furthermore, intentional learning may exert a good role in the long-term retention of knowledge depth and enhanced synonym discrimination effectively. The results are discussed along with students' feedback from the questionnaire.

| KEYWORDS
L2 vocabulary learning, Learning conditions, Gloss, Dictionary, LLM

| ARTICLE INFORMATION
ACCEPTED: 02 October 2024 **PUBLISHED:** 28 October 2024 **DOI:** 10.32996/ijllt.2024.7.11.6

1. Introduction

Learning vocabulary incidentally with larger amounts of L2 input may provide the greatest potential to promote L2 vocabulary growth. Gloss is one type of input modification that facilitates vocabulary learning by providing additional information and more sufficient contextual cues (Webb & Nation, 2017; Ko, 2012). Generally, gloss refers to a brief explanation of a word or phrase in a text, which can be provided in different forms (textual, visual, or auditory), languages (L1, L2, or bilingual), locations (in-text, marginal, bottom, glossary, or pop-up) and media (paper or hypermedia) (Zhang & Ma, 2021; Webb & Nation, 2017).

Dictionary, as another type of input modification, is widely acknowledged as an important vocabulary learning strategy with explicit and deep elaboration other than inferred meanings, which are more instable (Hulstijn et al., 1996; Webb & Nation, 2017). Therefore, they are both word-focused inputs that make the word more salient to help learners convert vocabulary input to intake (Ko, 2012; Hill & Laufer, 2003; Hulstijn, 1996). Abundant studies have provided evidence that gloss and dictionary use would facilitate vocabulary learning in terms of form and meaning retention (Hulstijn et al., 1996; Hill & Laufer, 2003; Zandieh & Jafarigohar, 2012; Malone, 2018).

Many studies lay emphasis on the consequences of different forms, types, or modes of them separately (Zandieh & Jafarigohar, 2012; Knight, 1994; Jung, 2016; Zhang & Ma, 2021; Huang & Eslami, 2013; Ko, 2012; Zhang et al., 2020; Yanagisawa et al., 2020).

Copyright: © 2024 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

Only a few studies highlight the differences in effect between them on meaning retention (Hulstijn et al., 1996; Levine et al., 2004). Additionally, in those studies, dictionary use is optional, while gloss is completely given. In that case, the real results of the differences they brought are untenable to some extent.

When it comes to measuring word knowledge regarding vocabulary growth, there are mainly two different divisions. First, receptive knowledge (words that are recognized when heard or read) and productive knowledge (words that can be called to mind and used in speech or writing), with the former being greater. Second, breadth of word knowledge and depth of word knowledge. Vocabulary breadth primarily involves form-meaning mapping knowledge. Vocabulary depth denotes how well a word is known as it might involve knowledge of paradigmatic relations (antonymy, synonymy, hyponymy, gradation) and syntagmatic relations (collocational restrictions) (Milton, 2009; Haastrup & Henriksen, 2000; Zhang et al., 2020). Most researchers are concerned about vocabulary breadth with form recognition and form-meaning connection measures (Hill & Laufer, 2003; Jung, 2016; Malone, 2018; Zhang & Ma, 2021). There is a scarcity of focus on vocabulary learning in terms of its depth.

Besides, vocabulary learning conditions have long shifted in emphasis from intentional to incidental in most studies. It remains unclear whether intentional learning would produce better long-term vocabulary learning gains. The effect of learning conditions needs to be further explored (Webb & Nation, 2017; Zhang et al., 2020).

Furthermore, recently significant progress has been made in the development of large language models (LLMS) such as GPT-3, PaLM, Flan-UL2, LLaMA, and ChatGPT. These large models are pre-trained on large amounts of textual data and employ a variety of training techniques, including instruction tuning, reinforcement learning based on human feedback (RLHF), and others. Among them, ChatGPT, launched by OpenAI, shows a very amazing language understanding, generation, and knowledge reasoning ability; it can understand user intentions extremely well and truly achieve multiple rounds of communication (Zhang et al., 2023). Employing instruction tuning, this study adopted ChatGPT's data as our gloss material to explore not only the difference between gloss and dictionary but also the potential of smart technology-empowered language learning and teaching.

The current study first explores incidental vocabulary acquisition while utilizing a gloss or dictionary to determine whether there are differences between them. Then, the distinctions between intentional and incidental effects were also examined while utilizing a dictionary solely.

As for the measurements of learning effects, four aspects of vocabulary knowledge (form-meaning connection, synonym discrimination, derivation production, and collocation production) were measured to provide a rigorous assessment of retention, especially the depth of it from the last three measures.

2. Literature review

2.1 Intentional and incidental vocabulary learning

Incidental and intentional vocabulary acquisition are two widely researched topics in instructed L2 vocabulary learning and are centered on the role of learner intention (Bruton et al., 2011; Hulstijn, 2003; Schmitt, 2008).

The former refers to activities that are purposefully geared to commit word knowledge into memory (Hulstijn, 2001、2003), while the latter refers to lexical knowledge that occurs as a by-product of activities not designed with the explicit purpose of vocabulary learning (Zhang et al., 2020; Loewen, 2015).

The distinction between incidental and intentional vocabulary learning takes its origin from experimental psychology (Hulstijn, 2001). They are distinguished by different pre-learning instructions. It is generally agreed that intentional vocabulary learning, in which student attention is directly engaged and focused on vocabulary, offers a greater chance for vocabulary learning and is responsible for the majority of L2 vocabulary learning (Laufer, 2011; Hulstijn, 2001; Nation, 2001). However, Bruton et al. (2011) argued that it is much easier to ensure that learning is a by-product of another activity. Empirical evidence suggested that intentional vocabulary learning would lead to better short-term learning outcomes than incidental vocabulary learning (Laufer, 2003).

According to a recent meta-analysis from Zhang et al. (2020), intentional learning conditions consistently generate larger effect sizes. Learning conditions turn out to be a significant moderator in within-group contrasts, while for between-group contrasts, the difference between the two subgroups is not significant. The argument continues for which condition is best for learning outcomes

(Malone, 2018). Incorporating learning conditions as a moderator variable would help to demonstrate the validity of the controversial incidental/intentional vocabulary learning distinction.

2.2 Incidental vocabulary learning with gloss and dictionary

Under incidental and intentional learning conditions, further studies on dictionary-assisted and gloss-assisted vocabulary learning were widely conducted.

Dictionary use is widely acknowledged by L2 vocabulary researchers as an important vocabulary-learning strategy (Nation, 2001; Webb & Nation, 2017; Zhang et al., 2020). A vocabulary gloss is a short definition or an explanation of a word (Nation, 2001) in different forms, languages, locations, and media. (Zhang & Ma, 2021; Jung, 2016; Ko, 2012; Yanagisawa et al., 2020). Dictionary intends to provide more sufficient and complete contextual cues of one word with instances of its word families, collocations, phrases, and illustrative sentences. Accordingly, some scholars reckoned that it is usually more effective (Hulstijn et al., 1996; Hill & Laufer, 2003).

Some empirical studies hold mixed results. There are many studies exploring vocabulary learning outcomes solely with a gloss or a dictionary. For instance, Jung (2016) studied gloss, and Knight (1994) focused on dictionary use. Furthermore, most studies investigated them separately to determine which type, language, or form is better. Zandieh & Jafarigohar (2012) studied hypertext gloss. Huang & Eslami (2013) are concerned about the language type of dictionaries, while Ko (2012) is concerned about glosses. Yanagisawa et al. (2020) and Zhang & Ma (2021) conducted a meta-analysis to explore the overall effects of different glosses, while Zhang et al. (2020) conducted a meta-analysis to find out the different dictionaries. Only a few studies showed concerns about distinctions between them (Hulstijn et al., 1996; Levine et al., 2004). This is to ignore the fact that they all took dictionary use as optional.

With regards to the outcome brought by gloss or dictionary-assisted reading, fruitful studies provide evidence that they both boost language acquisition and reading comprehension (Knight, 1994; Hulstijn et al., 1996; Zandieh & Jafarigohar, 2012; Jung, 2016; Webb & Chang, 2015). The overall outcome of vocabulary learning would be strengthened by frequency of occurrence (Feng & Webb, 2020; Webb & Chang, 2015; Hulstijn et al., 1996). Furthermore, studies such as Malone (2018), Hulstijn et al. (1996), and Hill & Laufer (2003) projected form-recognition effects. Other research illustrated form-meaning outcomes (Zandieh & Jafarigohar, 2012; Malone, 2018; Jung, 2016). Though lexical development is defined both in terms of breadth (or size) and depth (or level of knowledge), only one study (Dai et al., 2019) investigated whether dictionary use would help learners acquire vocabulary depth knowledge (collocational knowledge in their case). Zhang et al. (2020) suggested that future studies may adopt a dimensions approach to examine the effects in improving other aspects of vocabulary knowledge beyond form-meaning mappings.

3. Research questions

1. To what extent do gloss and dictionaries increase vocabulary knowledge?
2. To what extent do incidental learning and intentional learning increase vocabulary knowledge?
3. Do gloss and dictionary contribute to L2 incidental vocabulary learning gains differently?
4. Do learning conditions contribute to learning gains differently?

Note that in Q1 & 2, no interactions were considered in this study as it only focused on effects within groups under incidental conditions for gloss or dictionary use and with dictionary use for incidental or intentional conditions. Therefore, they would be answered together within each group. No interactions would be contemplated in Q3&4 either with the same controlment.

4. Method

4.1 Participants

A total of 70 students of grade 2 from a senior high school in Chengdu, Sichuan Province, took part in this study with similar ages of 17 or 18 (M = 17.46, SD = .50) and nearly averaged gender (24 females and 26 males). Among them, exact 64 participants were assigned to four groups for this study according to a recent municipal English diagnostic test (maximum score: 150) and further statistics indicated that there was no significant difference between four groups $F(3, 63) = .01$, $p = .99$. The participants achieved moderate scores on the test (M = 101.62, SD = 14.59) indicating that they were all at an immediate level of English proficiency given that threshold score for first-tier universities is 97. Kolmogorov-Smirnov test of normality indicated that the participants were normally distributed among groups ($p > .05$) based on their test scores.

Prior to starting the study, informed consents were signed stating that involvement in the study was voluntary by participants.

4.2 Research Instruments

4.2.1 Target Words

Due to the setting and purpose of the test, target words were selected first to be in line with the following conditions. First, it is without doubt that they should be unknown to participants, which increased the internal validity of the study (Nation & Webb, 2011).

Secondly, at least two words that were previously known to participants should be in paradigmatic relation (i.e., synonyms) with them. This serves for the 3 options of our synonym discrimination test with two known synonyms plus an unknown target word. Therefore, we searched for our target words and the test questions from two materials: Modern Collegiate English Synonym Discrimination and TEM-4 tests. We primarily obtained test questions with two words that are within the scope of the syllabus in grade 1 or 2 to ensure that they are known to our participants and one word in the syllabus of grade 3 to ensure that they are unknown to our participants. After further tests and assurance from teachers, we selected the target words and questions successfully.

Finally, their derivation and collocation patterns should be in accordance with the participants' cognitive stage and the examination mode of the college entrance examination, which espouses rational learning and feedback. For example, changes in the part of the speech for derivation and collocation of phrases and prepositions are predominantly tested.

Ultimately, we pinned down 10 target words, including 5 verbs, 4 nouns, and one adjective (attain, utilize, prolong, abolish, retort, upbringing, personnel, aftermath, illusion, artificial).

As Grains put it, retention in short term memory is not effective if the number of chunks of information exceeds seven and productive knowledge is even harder (Milton, 2009); 6 words and only 3 words out of them were adopted in derivation and collocation production test respectively.

4.2.2 Materials

By the same token, for study purposes, the reading material was coined with the help of two native English speakers and a group of six English teachers based on target words.

Several conditions were ensured for coining it. First, it needs to be age appropriate and smooth, containing all target words. Second, 95% of the vocabulary items are within participants' prior knowledge, and the comprehension of it largely relies on the study of target words to foster optimum learning (Webb & Nation, 2017) as well as to motivate enthusiasm under incidental conditions. Third, target words only occur once in their original form, irrespective of derivations and collocations. The whole text, coined by the target words, which were in bold in the text, was fluent with great logic, narrating a thorough topic. Finally, another two native speakers and six students read the text to ensure the readability and fluency of the text. If it didn't read well by them, the text would be refined again until the best feedback was acquired.

As for gloss and dictionary materials, dictionary materials were printed papers provided with primitive information about the target word and its derivational words and collocations, if they exist, excluding other information in Oxford Intermediate Learner's English Chinese Dictionary (8th Edition). Gloss materials were mainly coined first by ChatGPT.

Large language models, primarily ChatGPT, currently demonstrate significant capabilities in simulating human language behavior, generating comprehensible text, and providing appropriate responses based on contextual cues. Moreover, these models continuously enhance their generative capabilities by incorporating human feedback, showcasing strong natural language generation proficiency (Feng & Zhang, 2023). Specifically, these large models excel in: (1) Language generation: they can generate sentences by completing prompts based on provided ones. For instance, it can fill the blank as in the sentence "The word is good because___"; (2) In-Context Learning: they can generate solutions for new test cases by following examples of given tasks. For instance, if we give it a task and an example of the accomplished task by us, it will perform new tasks like we did in the example. Otherwise, there is only a given task without an an exampland e, and it would perform the task directly. If the results are not satisfying, we can keep modifying our prompt by instruction tuning through conversation turns. In these turns, ChatGPT can learn through human feedback and learn from the context, thus giving a more perfect answer and hitting our purposes more precisely;

(3) World knowledge acquisition: encompassing factual knowledge and common sense (Zhao et al., 2023). Therefore, abilities of ChatGPT, such as prompt learning, instruction tuning as well as a chain of thought, not only make basic language analysis possible but also show the potential of future knowledge processing in specialized fields. Under such circumstances, we first gained our gloss materials through prompt engineering by asking, "To acquire the most gains of incidental vocabulary learning for Chinese high school students, can you provide some assisted input for the word "attain" and "utilize" as an example." to meet our study needs. ChatGPT then provided examples in 6 aspects:

1. Provide Contextual Examples
2. Visual Representations
3. Synonyms and Related Words
4. Sentence Completion
5. Opposites or Contrasting Words
6. Collocations or Phrases

Gloss materials were then refined by removing the following parts compared with dictionary ones: 1. Additional meanings (i.e., other than the meaning in the reading text). 2. All further explanation of derivational words and collocations. On top of that, both are provided in the same way in terms of language (bilingual), pattern (parts in bold), mode (textual), and manner of provision (separate printed papers).

To clarify the findings, a questionnaire was also designed utilizing five-point Likert Scale for participants to choose the extent of material's role exerted in synonym discrimination, derivation and collocation learning gains. Point settings were 0-4 expressing ideas from no use to a great use. Before the three choices, there was a primary question asking if they had viewed the materials given wholly. Data with a negative answer would be excluded.

4.3 Dependent Measures

1. Form-meaning connection

Section A of the test examined form-meaning connection of the single-word items with 10 multiple-choice items. The set of options was drawn by Pavia et al. (2019). Each item contained four meaning options, including a. the correct answer in L1; b. a distracter in word families with the target word; c. a distracter of the other target word; and d. I don't know. All multiple-choice items contained such a pattern in this study (one correct answer, two distracters, and I don't know). An example question is shown in (A):

(A) What is the meaning of "attain"?
*A.获得 B.利用 C.**关联** D.我不知道*

2. Synonym discrimination

Section B explored synonym discrimination between learned words and prior knowledge with 30 multiple-choice items in the form of a sentence context. Three words (one target word and two known words) were tested in every three different items with the same three words options. Items with target words were put at the rear firstly. Sentence questions were subtly selected or refined from TEM-4 and Modern Collegiate English Synonym Discrimination to prevent coincidence in all given materials. An example question is provided in (B):

(B) We all know that Mary has had a strict____. ( )
A. growth
B. upbringing
C. development
D. I don't know

3. Derivation production

Section C required participants to give a derivational word (change of part of speech) that occurred in the two assisted materials, which served as the depth of knowledge of the target word as it indicated how well the word was known by participants. Such

derivation is also a key point of study and texts in high school. Six original words were given, and they wrote behind each of them. An example question is projected in (C):

(C) Please write the part of speech derivations of the following words according to the material.
(e.g., excited-exciting/excitement)
Artificial:
Attain:

4. Collocation production
Section D, another key point in such stage, contributed to the retention of collocations in materials. Three words in primary form were given, and they were written behind each of them. An example question is illustrated in (D):

(D) Please write the collocations of the following words according to the material provided.
(e.g., habit-eating habit)
Personnel:
Artificial:

4.4 Procedure

The whole study was conducted over two weeks. First, the participants were divided into four groups with the same level according to their diagnostic test scores. Groups consisted of a control group (C) (n = 14) and three experimental groups: ETG (with text and gloss) (n = 18), ETD (with text and dictionary) (n = 16), and ED (with dictionary only) (n = 16). In the first week, all participants completed the pretest of four sections first. Then, they read the materials and accomplished posttest immediately. Concretely speaking, ETG and ETD groups read the text first, and then they were required to read the material of printed gloss or dictionary, learning each target word. They learned each word only once without time limitation. After finishing reading the two materials, they handed them in and did the test. ED group was given only the material of a printed dictionary to learn each word once as well. What is more, whichever group they were in, participants should learn every word provided compulsorily, not optionally. A week later, participants completed the delayed posttest with a different order in each section and a questionnaire attached to it.

It rates a mention that the test sections, with section A referring to language breadth and sections B-D referring to language depth, were completed separately in order to avoid reciprocal factors.

4.5 Analysis

R language was used to analyze the data, removing all the missing data.
In four test sections, all data were scored dichotomously, with 1 for a correct response and 0 for an incorrect response. Scores in section B were given only in 10 items with target words. Scores were given in sections C and D as long as their answers met with testing aim. To answer the first and second research questions, repeated measures ANOVA were used to compare scores within each group at different times of testing (pretest, immediate posttest, and delayed posttest). To answer the third and fourth research questions, One-Way ANOVA was used to determine significant between-group differences in their vocabulary learning gains across the four sections: once for group C, ETG, and ETD and once for group C, ETD, and ED. Before using that, analysis using One-Way ANOVA indicated that for pretests, there is no significant difference between four groups in section A $F(3, 63) = 1.64$; $p = .19$ and section B $F(3, 63) = 2.22$; $p = .10$. All groups scarcely gave productive results in the last two sections. No violation of the assumptions of normality and homoscedasticity had occurred in our data. Statistics in their overall pretest scores also indicated that we had successfully selected target words with which they had not been familiar ($M = .95$, $SD = 1.21$).

5. Results

The descriptive statistics of vocabulary test scores on each section of the test are presented first in Table 1. To answer the first and second research questions together, repeated measures ANOVA were used to compare scores within each group at different times of tests. To answer the third and fourth research questions, One-Way ANOVA was adopted between groups to determine whether they differed significantly in their vocabulary learning gains.

Table 1. Descriptive statistics of vocabulary test scores.

Participant subgroups	Form-meaning connection			Synonym discrimination			Derivation production			Collocation production		
	Pretest	Immediate posttest	Delayed posttest	Pretest	Immediate posttest	Delayed posttest	Pretest	Immediate posttest	Delayed posttest	Pretest	Immediate posttest	Delayed posttest
Control (n = 14)	.71 (.27)	.21 (.43)	.00 (.00)	.00 (.00)	.21 (1.43)	.07 (.27)	.00	.00 (.00)	.00 (.00)	.00	.00 (.00)	.00 (.00)
ETG (n = 18)	.89 (1.08)	6.00 (1.49)	5.00 (1.75)	.33 (.59)	4.50 (2.50)	5.44 (3.01)	.00	1.28 (1.22)	.67 (.91)	.00	.78 (.94)	.44 (.78)
ETD (n = 16)	.56 (.89)	6.06 (2.17)	5.19 (1.97)	.50 (.89)	5.38 (2.30)	8.13 (1.59)	.00	2.69 (.48)	1.06 (1.06)	.00	.25 (.45)	.19 (.40)
ED (n = 16)	.69 (.87)	9.13 (1.54)	8.25 (1.44)	.50 (.52)	7.50 (2.85)	7.69 (2.15)	.00	2.13 (1.67)	1.75 (1.77)	.00	.81 (1.22)	.56 (1.09)
Total (N = 64)	.61 (.88)	5.50 (3.50)	4.76 (3.23)	.34 (.62)	4.53 (3.39)	5.50 (3.69)	.00	1.56 (1.45)	.89 (1.27)	.00	.48 (.87)	.31 (.73)

Note. Maximum score on form-meaning connection and synonym discrimination was 10, the maximum score on derivation production was 6 and on collocation production was 3.

5.1 Comparisons within groups

On the form-meaning connection section of the test, all the experimental groups embodied significant difference between scores from the pretest to the immediate posttest and from pretest to the delayed posttest with strong effect size (ETG: F (2, 34) = 84.43; $p < .001$; partial $\eta^2 = .832$, ETD: F (2, 30) = 65.68, $p < .001$; partial $\eta^2 = .814$, ED: F (2, 30) = 242.87, $p < .001$ partial $\eta^2 = .942$). The analysis for the control group indicated that there was no significant difference between them F (2, 26) = 1.86; $p = .191$; partial $\eta^2 = .125$.

On the synonym discrimination section, all experimental groups, again, held significant differences between scores from the pretest to the immediate posttest and from the pretest to the delayed posttest with a large effect size (ETG: F (2, 34) = 42.27; $p < .001$; partial $\eta^2 = .713$, ETD: F (2, 30) = 91.57; $p < .001$; partial $\eta^2 = .859$, ED: F (2, 30) = 86.04; $p < .001$; partial $\eta^2 = .852$.) For the control group there was no significant difference F (2, 26) = 2.60; $p = .09$; partial $\eta^2 = .167$.

On the derivation production section, experimental groups continued to vary significantly from the pretest to the immediate posttest and from pretest to the delayed posttest (ETG: F (2, 34) = 14.45; $p < .001$; partial $\eta^2 = .459$, ETD: F (2, 30) = 77.31; $p < .001$; partial $\eta^2 = .838$, ED: F (2, 30) = 86.04; $p < .001$; partial $\eta^2 = .852$). ETG declined in large effect size obtained previously.

Control group could barely give answers for productive knowledge so their scores for this section and the next were zero across all the tests.

On the collocation production section, experimental groups all embodied relatively slighter differences between the pretest to the immediate posttest and between the pretest to the delayed posttest with a smaller effective size as well (ETG: F (2, 34) = 7.86; $p = .002$; partial $\eta^2 = .316$, ETD: F (2, 30) = 3.82; $p = .033$; partial $\eta^2 = .203$, ED: F (2, 30) = 4.85; $p = .015$; partial $\eta^2 = .244$).

The pairwise comparisons between pretest and immediate posttest scores, as well as pretest and delayed posttest scores indicated that for the sections A、B and C, there was a significant increase for all experimental groups with Cohen's d over 1 (strong effect size). In addition, there was a slight difference between pretest and delayed posttest for ETD with moderate effect ($p = .83$, Cohen's $d = .95$) and ED with strong effect size ($p = .57$, Cohen's $d = 1.03$) in terms of section D. For the control group, the pairwise comparison between pretest and immediate posttest scores indicated no difference in section A with modest effect size (Cohen's $d = .32$) and in section B with moderate effect size (Cohen's $d = .97$). The pairwise comparison between pretest and delayed posttest scores still indicated no difference in section A with a moderate effect size (Cohen's $d = .97$) and in section B with moderate effect size (Cohen's $d = .53$).

The pairwise comparisons between pretest and immediate posttest scores and between pretest to delayed posttest scores are shown in Table 2.

Table 2: Pairwise comparison for different sections for the test.

Time of testing (*i*)	Time of testing (*j*)	Word knowledge dimension	Difference between means (*i-j*)	Std error	*p*	Cohen's *d*
ETG		Form-meaning				
1	2	connection	-5.111***	1.711	.000	3.961
	3		-4.111***	2.054	.000	2.904
ETG		Synonym				
1	2	discrimination	-4.166***	2.382	.000	2.699
	3		-5.111***	3.007	.000	2.817
ETG		Derivation				
1	2	production	-1.278***	1.227	.000	2.081
	3		-.667**	.907	.006	1.489
ETG		Collocation				
	2	production	-.778**	.943	.003	1.659
1	3		-.444*	.784	.028	1.128
ETD		Form-meaning				
1	2	connection	-5.500***	2.529	.000	3.594

	3		-4.625***	2.305	.000	3.231
ETD		Synonym				
1	2	discrimination	-4.875***	2.446	.000	3.059
	3		-7.625***	2.062	.000	6.153
ETD		Derivation				
1	2	production	-2.688***	.478	.000	11.167
	3		1.063**	1.063	.001	2.000
ETD		Collocation				
1	2	production	-.250*	.447	.041	1.111
	3		-.188	.403	.083	.950
ED		Form-meaning				
1	2	connection	-8.437***	1.787	.000	6.996
	3		-7.562***	1.931	.000	6.545
ED		Synonym				
1	2	discrimination	-7.000***	3.011	.000	4.167
	3		-7.187***	2.344	.000	5.406
ED		Derivation				
1	2	production	-2.125***	1.668	.000	2.551
	3		1.750**	1.770	.001	1.977
ED		Collocation				
1	2	production	-.813*	1.223	.018	1.328
	3		-.563	1.094	.057	1.027

Note. Test time 1 = Pretest, 2 = Immediate posttest, 3 = Delayed posttest.
*p < .05, **p < .01, ***p < .001

5.2 Comparisons between groups

To answer the third and fourth research questions, One-Way ANOVA was adopted between every relevant three groups to determine whether they differed significantly in their vocabulary learning gains under two tools using and two learning conditions across the four sections of the test. Post-hoc comparisons using Scheffe's test are illustrated in Tables 3 and 4.

Table 3: Pairwise comparison for different groups on the immediate posttest.

	Comparison group	Word knowledge dimension	*p*	95% confidence interval for the difference	
				Upper bound	Lower bound
Input	ETG-ETD	Section A	.933	-1.421	1.296
	ETG-ETD	Section B	.467	-2.655	.906
	ETG-ETD	Section C	.000***	-2.109	-.711
	ETG-ETD	Section D	.064	-.024	1.079
	ETG-ETD	Overall depth	.065	-3.601	.087
Condition	ETD-ED	Section A	.000***	-4.480	-1.644
	ETD-ED	Section B	.030*	-4.078	-.172
	ETD-ED	Section C	.310	-.356	1.482
	ETD-ED	Section D	.130	-1.252	.127
	ETD-ED	Overall depth	.076	-4.428	.178

Table 4: Pairwise comparison for different groups on the delayed posttest.

	Comparison group	Word knowledge dimension	*p*	95% confidence interval for the difference	
				Upper bound	Lower bound
Input	ETG-ETD	Section A	.941	-1.550	1.175
	ETG-ETD	Section B	.002**	3.504	7.242
	ETG-ETD	Section C	.389	-1.117	.325
	ETG-ETD	Section D	.384	-.208	.722
	ETG-ETD	Overall depth	.003**	-4.774	-.865
Condition	ETD-ED	Section A	.000***	-4.435	-1.769
	ETD-ED	Section B	.739	-.984	-1.859
	ETD-ED	Section C	.291	-1.781	.406
	ETD-ED	Section D	.315	-.992	.242
	ETD-ED	Overall depth	.036*	-5.469	-.156

5.2.1 Gloss and dictionary use

Between groups of EDG, ETD, and C, the results indicated an overall statistically significant difference ($F (2, 45) = 84.75$; $p < .001$; partial $\eta^2 = .790$) on the immediate posttest. When the results for different sections of the test were considered separately, there was a significant difference found for form-meaning connection ($F (2, 45) = 72.09$; $p < .001$; partial $\eta^2 = .762$), for synonym discrimination ($F (2, 45) = 26.88$; $p < .001$; partial $\eta^2 = .544$), for derivation production ($F (2, 45) = 41.98$; $p < .001$; partial $\eta^2 = .651$) with strong effect sizes and for collocation production ($F (2, 45) = 6.38$; $p = .004$; partial $\eta^2 = .221$) with a modest effect size.

Further results were elaborated between EDG and ETD on the immediate posttest here. Post-hoc comparisons using the Scheffe test for the form-meaning connection section, however, indicated no significant difference with a weak effect size (Cohen's $d = .03$). Between them, there was also no significant difference in synonym discrimination with a moderate effect size (Cohen's $d = .36$) and in collocation production ($p = .64$, 95% CI [-.02, 1.08]). In terms of derivation production, there was a significant difference with a large effect size (Cohen's $d = 1.34$). On the immediate posttest overall, no significant difference existed between ETG and ETD regarding the breadth of knowledge gains (section A), and it was the opposite concerning the depth of knowledge (section B+C+D) with large effect size (Cohen's $d = .71$).

The results presented a statistically significant difference ($F (2, 45) = 99.28$; $p < .001$; partial $\eta2 = .815$) between three groups on the delayed posttest. Then the results for different sections of the test were illustrated individually, there was a significant difference found for form-meaning connection ($F (2, 45) = 52.37$; $p < .001$; partial $\eta^2 = .699$), for synonym discrimination ($F (2, 45) = 57.95$; $p < .001$; partial $\eta^2 = .720$) with large effect sizes and for derivation production ($F (2, 45) = 6.21$; $p = .004$; partial $\eta^2 = .216$) with a modest effect size. No significant difference was found in collocation production ($F (2, 45) = 2.79$; $p = .072$; partial $\eta^2 = .110$) with a weak effect size.

Between EDG and ETD on the delayed posttest, post-hoc comparisons indicated no significant difference in form-meaning connection with weak effect size ($p = .94$, Cohen's $d = .01$), in derivation production ($p = .39$, Cohen's $d = .40$) and in collocation production ($p = .38$, Cohen's $d = .44$) with moderate effect sizes. Significant difference was only found in synonym discrimination with large effect size ($p = .002$, Cohen's $d = 1.17$). On the delayed posttest overall, significant difference did not lie between ETG and ETD regarding the breadth of knowledge gains (section A) and lied concerning the depth of knowledge (section B+C+D) with a large effect size ($p = .003$, Cohen's $d = 1.13$).

5.2.2 Intentional and incidental learning

Between groups of ED, ETD, and C, the results presented a statistically significant difference ($F (2, 43) = 103.03$; $p < .001$; partial $\eta^2 = .827$) between three groups on the immediate posttest. Then the results for different sections of the test were illustrated separately; there were significant differences with large effect sizes form-meaning connection ($F (2, 43) = 52.37$; $p < .001$; partial $\eta^2 = .853$), for synonym discrimination ($F (2, 43) = 43.54$; $p < .001$; partial $\eta2 = .669$) and for derivation production ($F (2, 43) = 29.46$;

$p < .001$; partial $\eta^2 = .566$). No significant difference was found in collocation production ($F (2, 43) = 2.79$; $p = .17$; partial $\eta^2 = .172$) with a weak effect size.

Between ED and ETD, post-hoc comparisons indicated significant differences in form-meaning connection with a large effect size ($p < .001$, Cohen's $d = 1.65$) and synonym discrimination ($p = .03$, Cohen's $d = .82$) with a moderate effect size. No significant differences could be observed in derivation production ($p = .31$, Cohen's $d = .52$) and in collocation production ($p = .13$, Cohen's $d = .67$) with moderate effect sizes. On the immediate test overall, a significant difference lied between ED and ETD regarding the breadth of knowledge gains (section A). Slight difference could be found in depth of knowledge (section B+C+D) with a moderate effect size ($p < .076$, Cohen's $d = .86$).

Between those three groups, the results indicated an overall statistically significant difference ($F (2, 43) = 141.33$; $p < .001$; partial $\eta^2 = .868$) on the delayed posttest. When the results for different sections of the test were considered separately, there was a significant difference found for form-meaning connection ($F (2, 43) = 123.72$; $p < .001$; partial $\eta^2 = .852$), for synonym discrimination ($F (2, 43) = 119.19$; $p < .001$; partial $\eta^2 = .847$), for derivation production ($F (2, 43) = 41.98$; $p = .001$; partial $\eta^2 = .265$) with a modest effect size while there was no significant difference for collocation production ($F (2, 43) = 6.38$; $p = .083$; partial $\eta^2 = .109$) with a slight effect size.

Further results were discovered between ED and ETD then. Post-hoc comparisons test for the form-meaning connection section indicated a significant difference with a strong effect size ($p < .001$, Cohen's $d = 1.80$). However, there was no significant difference in synonym discrimination with a modest effect size ($p = .739$, Cohen's $d = 0.23$), in derivation production ($p = .291$, Cohen's $d = 1.72$) and in collocation production ($p = .315$, Cohen's $d = 0.50$) with moderate effect sizes. On the delayed posttest overall, a significant difference existed between ED and ETD regarding both the breadth of knowledge gains (section A) and the depth of knowledge (section B+C+D) with a large effect size ($p = .03$, Cohen's $d = 1.09$).

6. Discussion

This study first expanded on earlier research by providing empirical evidence if gloss and dictionary use can facilitate incidental vocabulary knowledge of form-meaning connection, synonym discrimination, derivation production and collocation recognition. On the other hand, whether dictionary use under incidental and intentional conditions boost those learning gains was also explored. Moreover, the present study provided a more in-depth investigation of the contrasts between assisted tools (gloss and dictionary) and between learning conditions (intentional and incidental), respectively, by measuring learning gains in breadth and depth. Additionally, these two kinds of contrast were observed under controlled variables between every three groups respectively ignoring the interactions. The results were discussed, along with the feedback from the questionnaire.

6.1 How do assisted tools and learning conditions increase learning gains

In answer to the first and second research questions, the results indicated an overall gain in vocabulary knowledge of 40.34% in mean scores from pretest to immediate posttest for ETG, of 45.89% for ETD and of 63.34% for ED. The results are consistent with earlier research that gloss and dictionary have a positive effect on L2 vocabulary leaning (Knight, 1994; Hulstijn et al., 1996; Ko, 2012; Yanagisawa et al., 2020). It is without doubt that incidental leaning as a by-product activity and intentional leaning are both imperative for L2 vocabulary learning growth (Pavia et al., 2019).

In contrast the mean scores for the control group regressed by 33% from the pretest to the immediate posttest. The results from pretest to delayed posttest of the experimental groups revealed a gain of 35.62% for ETG, of 46.55% for ETD and 58.82% for ED. The results of the control group from pretest to delayed posttest revealed a regression of 66.67%.

In separate vocabulary knowledge sections compared with their maximum scores, ETG acquired more gains in form-meaning connection and synonym discrimination followed by collocation production and derivation production from pretest to immediate posttest and from pretest to delayed posttest. ETD obtained most gains for form-meaning connection and synonym discrimination followed by derivation production and collocation production from pretest to immediate posttest. ED acquired high gains for form-meaning connection and synonym discrimination followed by derivation production and collocation production from pretest to immediate posttest and from pretest to delayed posttest.

It is apparent that overall form-meaning connection is easier to acquired which is in line with previous studies that form-oriented tasks resulted in higher gains (Hill & Laufer, 2003 ; Zhang et al., 2020). Such tasks have lied in the measurement of abundant studies in terms of language breadth. ETG provided similar tendency of results from immediate posttest and delayed posttest in form-meaning connection with that of Yanagisawa et al. (2020).

This study contributes to furnishing data adopting ampler dimensions of tests to investigate gains in more aspects under various conditions. As Zhang et al. (2020) pointed out, only one study investigated whether dictionary use would help learners acquire vocabulary depth knowledge (collocational knowledge in their case) and appealed to studies in more aspects considering the inconclusive results.

In terms of the depth of knowledge, the three groups all obtained more gains in synonym discrimination which accorded to higher scores of feedbacks in the questionnaire. The main reason is that section C and D are productive knowledge, which is inherently more difficult (Milton, 2009; Zhang & Ma, 2021). It is intriguing that ETG intended to bring better effects in collocation production than derivation production, which is also in accordance with the feedback of the questionnaire with M = 1.82, SD = .21 for this section and with M = 1.52, SD = .17 for derivation production. Both ETD and ED, instead, acquired higher gains in derivation production along with almost the same feedback in the questionnaire. It goes along with Huang & Eslami's (2013) study that students paid more attention to derivations than its usage when using the dictionary.

Apart from supporting the beneficial effects of the more widely dimensions of incidental vocabulary learning with gloss and dictionary, the results also provided more evidence in support of learning gains under the intentional condition. With dictionary use, both learning conditions (ETD & ED) had a similar decrease in form-meaning connection on the delayed posttest. Incidental one (ETD) embodied an apparent drop in derivation production and an increase in synonym discrimination. Although the former phenomenon goes against Zandieh & Jafarigohar's (2012) finding that the incidental group gained new vocabulary only moderately but also did only moderately lose from it, that study focused on breadth knowledge mainly with hypertext gloss instead of paper dictionary. In contrast, intentional gains tend to keep abreast with immediate tests relatively, which provides novel evidence that intentional learning may produce better long-term vocabulary learning gains answering to Webb & Nation (2017) and Zhang et al. (2020) and is contrary to Zhang's (2020) results with meta-analysis. But again, they mainly contemplated gains of breadth while it was depth here. Therefore, more studies are needed comparing incidental and intentional conditions in long-term effects in terms of more language types.

6.2 The comparison of gloss and dictionary use in incidental learning

In answer to the third research question, the results indicated that gloss and dictionary use exerted a similar role in form-meaning connection but brought different effects in the depth of knowledge for derivation production on the immediate posttest and for synonym discrimination on the delayed posttest both with better effects in ETD. As aforementioned, form-oriented tasks are always easier to accomplish. Dictionary, on the contrary, exerts a great role in strengthening knowledge of language depth. Many reasons contribute to this phenomenon. The primary one would be that by using a dictionary, we might attempt to learn how the word is used with other words and pay attention to its grammatical functions and collocations or compare it with words or word parts we already know (Webb & Nation, 2017). Therefore, when it comes to language depth, especially the authentic use of one word, dictionary may be more effective, which is also in line with Hill & Laufer (2003). According to their questionnaire, ETD gave more affirmation in all sections than ETG, which was higher in derivation production and synonym discrimination. Their ideas acknowledge the empirical evidence accordingly. Pedagogically, dictionary is recommended as the material in the relevant learning process.

It also rates a mention that according to the gloss provided by ChatGPT primarily, it seems that in its "cognition", incidental learning gains of a word includes paradigmatic relations and syntagmatic relations by default. It highlights the importance of language depth again.

It differs to a small extent that meaning retention was better using a dictionary in Hulstijn's (1996) research while there was no big difference in our study the fact that there are no more comparison data, and the effect sizes are weak on immediate posttest and delayed posttest. Besides, in our study, words were learned compulsorily even for the dictionary group, which was the opposite of Hulstijn's.

Since there are scarce studies comparing gloss and dictionary and barely few in terms of depth knowledge resulting from them, this study contributes to some novel empirical findings in this case.

Dai et al. (2019) found that dictionary use would help learners acquire collocational knowledge. Apart from that, gloss could do the same with collocation examples only excluding additional information in our case. It indicates that words collocation may be easier to obtain even if there is not too much information. Therefore, in learning, teachers can choose a briefer and a more straightforward way instead of too much information. Moreover, dictionary use tends to cultivate better depth knowledge in derivation production and synonym discrimination. It is consistent with prior studies of students' preference consulting a dictionary. Huang & Eslami (2013) found that students learned derivations and additional meanings more than the usage of a word. Synonym discrimination, individually, kept its effects in the delayed posttest. It lends credibility to Haastrup & Henriksen (2000), who put forward that learners build up semantic fields by adding terms to them as they elaborate their vocabularies and creating links between words they already know and new L2 words. Moreover, central to depth of knowledge is the process of network building. So, concretely speaking, affluent information in a dictionary contributes a lot to strengthening the authentic use of a word, which ends up enhancing derivation production and synonym discrimination. In reverse, such two aspects require enough information and time for students to acquire.

In summary, we can conclude that gloss and dictionary use vary differently in gains of depth knowledge, with the dictionary exerting a better role, especially in derivation production and synonym discrimination, partly supported by previous studies (Hulstijn et al., 1996; Huang & Eslami, 2013).

6.3 The comparison of learning conditions with dictionary

In answer to the fourth research question, the results indicated that with dictionary use incidental and intentional learning vary significantly in breadth knowledge, with intentional one nearly 30% higher in an increase on both immediate posttest and delayed posttest. This is in line with previous research on immediate posttests but not on delayed posttests. Zandieh & Jafarigohar (2012) found that the incidental group gained new vocabulary only moderately but also did only moderately lose from it, while the intentional group gained more vocabulary items immediately but also forgot more of them by the time of the delayed retention task. The reason for this can be explained by Schmitt & Schmitt (1995), who state that attention, consciousness, and awareness are necessary for learning. The results for the immediate posttest are also in accordance with Laufer (2003), Webb & Chang (2002), and Zhang et al. (2020). With such abundant proof, intentional learning can largely strengthen the breadth of knowledge.

Apart from previous research, two new findings of learning conditions in depth in the present study are illustrated below. First, on the immediate posttest, intentional learning obtained more gains of 21.30% in terms of synonym discrimination. It is partly in line with Haastrup's (2000) contention that vocabulary acquisition is enhanced by directing learners' attention to lexical relations and to the analysis of words, thus supporting network building and depth of knowledge. Intentional learning, in that case, could be credited to direct attention to a large extent. To acquire knowledge in terms of synonym discrimination, intentional learning could help students obtain effective results.

Second, intentional learning performs better in the depth of knowledge on delayed posttest, with more gains of 14.76%. It can be explained by Schmitt & Schmitt (1995) that newly acquired words and expressions will slip out of memory more easily if they are acquired with less effort. Although, it is not words but more complex uses of words, we can speculate that it follows the principle by the same token. To retain learning fruit in long run, teachers should attach great importance to intentional learning conditions.

Broadly speaking, this study contributes to filling the gap of intentional learning gains in the long term and of differences between intentional learning and incidental learning with regards to depth knowledge. Intentional learning produces better effects generally with superior gains in synonym discrimination on the immediate posttest and in overall depth knowledge on the delayed posttest. Opinions from questionnaire show a similar result between the two conditions, which lends no more credibility. Nevertheless, more empirical research comparing these is warranted.

6.4 Pedagogical implications

The findings of this study suggest that there is educational value in incorporating both glosses and dictionaries in language learning. In the classroom, teachers can introduce collocations to students using glosses in an incidental learning setting, which can be effective in helping students learn and retain collocations. On the other hand, derivations and synonyms of words are better

learned incidentally with the use of dictionaries. This is because a substantial amount of input is required for network building and memory consolidation when learning these aspects.

This study also provides empirical evidence supporting the positive impact of intentional learning with dictionaries on strengthening synonym discrimination. Additionally, it suggests that intentional learning may lead to long-term retention, particularly in terms of knowledge depth. This finding responds to the arguments put forth by Webb & Nation (2017) and Zhang et al. (2020).

Regarding Large Language Models, the study highlights their significant potential in language teaching and learning during the process of prompt engineering. LLMs can be a valuable tool in facilitating language instruction and can contribute to the improvement of language learning outcomes. For instance, if we intend to get more concrete output data of language depth, we can give ChatGPT prompt like: "To acquire the most gains of form-meaning connection, synonym discrimination, derivation production and collocation production under intentional vocabulary learning for Chinese high school students, can you provide some assisted input for the word "personnel" and "artificial" as an example." And it just gives a fine-grained answer which is different from the primary gloss:

> *For the word "personnel":*
> *Personnel (noun): Individuals employed in an organization or engaged in an organized undertaking such as military service.*
> *Synonyms: Staff, employees, workforce.*
> *Derivation: Personnel department (noun phrase): The department of an organization dealing with the hiring, administration, and training of staff.*
> *Collocations: Personnel management, personnel files, personnel training.*
> *For the word "artificial":*
> *Artificial (adjective): Something that is made or produced by human beings rather than occurring naturally, typically as a copy of something natural.*
> *Synonyms: Synthetic, man-made, non-natural.*
> *Derivation: Artificiality (noun): The quality of being artificial or produced by human beings.*
> *Collocations: Artificial intelligence, artificial light, artificial turf.*

This time, ChatGPT focused on our requirements to provide the answer with four layers of language depth which was in line with our purposes as well. Such potential is of great value to be further explored for a better selection of teaching materials and techniques.

7. Conclusion

This study has provided in-depth empirical support for the differences in incidental vocabulary learning when using glosses by LLMs and dictionaries. It has also highlighted the importance of measuring learning gains in dimensions such as knowledge depth. The findings suggest that gloss by LLMs facilitates the acquisition of collocations in a short time, while dictionaries enhance the retention of derivations and synonym discrimination in the long run. Moreover, this study has provided in-depth empirical support for the potential of dictionary use under various learning conditions. Tentatively, intentional learning appears to perform better in the short-term regarding synonym discrimination and in the long term regarding overall knowledge depth. These findings, along with the adoption of LLMs, hold significant curricular implications. However, it is important to acknowledge several limitations that should be considered for future research.

Firstly, the inclusion of two factors, assisted tools and learning conditions, was limited to only three treatment groups with controlled conditions. To obtain a more comprehensive understanding and determine genuine differences without additional controlments, further research with separate groups is recommended.

Secondly, the sample size of participants in this study was limited. To enhance the validity and reliability of the results, future studies should aim to increase the sample volume as much as possible.

Additionally, the measurement of gains in this study focused on breadth and depth of vocabulary knowledge. However, it is worth noting that the measurement of depth, particularly in the synonym section, may subtly influence the form-meaning connection in breadth as additional input. Future investigations should explore the implications of these measurement choices more thoroughly.

Furthermore, this study has highlighted the innovative contribution of large language models (LLMs) in language learning and teaching. There is great potential to explore further applications of LLMs in assisting language learning, and more empirical studies are needed to uncover their full potential.

Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers.

The research is supported by the Ministry of Education Humanities and Social Sciences Youth Foundation for Research Funds: A Study on Discourse Semantic Representation System and Resource Construction for Language (No. 23YJCZH150) and 2024 Heilongjiang Province Postdoctoral Research Startup Fund Project "Research on the Construction of Discourse Semantic Networks for Generative Artificial Intelligence".

References

[1] Bruton, A., Garc íaLo ṕez,M.,& Esquiliche Mesa, R. (2011). Incidental vocabulary learning: An impracticable term? *TESOL Quarterly*, 45, 759–768.
[2] Dai, Y., Wu, Z., & Xu, H. (2019). The effect of types of dictionary presentation on. The retention of metaphorical collocations: Involvement load hypothesis vs. cognitive load theory. *International Journal of Lexicography*, 32(4), 411-431.
[3] Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 42(3), 499-523.
[4] Feng, Z. W., & Zhang, D. K. (2023). GPT and language research. *Technology Enhanced Foreign Language Education*, (02),3-11+105.
[5] Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics*, 10(2), 221-240.
[6] Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition.
[7] Huang, S., & Eslami, Z. (2013). The use of dictionary and contextual guessing strategies for vocabulary learning by advanced English-language learners. *English Language and Literature Studies*, 3(3), 1.
[8] Hulstijn, J. (2001). Intentional and Incidental Second Language Learning: A Reappraisal of Elaboration, Rehearsal and Automaticity. In Cognition and second language instruction. *Cambridge University Press*. Ch. 10.
[9] Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), The handbook of second language acquisition (pp. 349–381). *Oxford, UK: Blackwell.*
[10] Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80, 327–339.
[11] Jung, J. (2016). Effects of glosses on learning of L2 grammar and vocabulary. *Language Teaching Research*, 20(1), 92-112.
[12] Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The modern language journal*, 78(3), 285-299.
[13] Ko, M. H. (2012). Glossing and second language vocabulary learning. *Tesol Quarterly*, 46(1), 56-79.
[14] Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review/Revue Canadienne des Langues Vivantes*, 59, 567–587.
[15] Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, 24(1), 29-49.
[16] Levine, A., Bejarano, Y. A. E. L., Carrell, P., & Vered, L. I. L. Y. (2004). Comparing. dictionary definitions and glosses in hypertext foreign language reading: Facilitating foreign language reading comprehension and vocabulary acquisition. *CATESOL Journal*, 16(1), 59-68.
[17] Loewen, S. (2015). Introduction to instructed second language acquisition. New York, *NY: Routledge/Taylor & Francis*.
[18] Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, 40(3), 651-675.
[19] Milton, J. (2009). Measuring second language vocabulary acquisition (Vol. 45). *Multilingual Matters*.
[20] Nation, I.S.P. (2001). Learning vocabulary in another language. Cambridge: *Cambridge*.
[21] Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition*, 41(4), 745-768.
[22] Schmitt, N. (2008). 'Instructed Second Language Vocabulary Learning.' *Language Teaching Research* 12.3: 329–363.
[23] Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: Theoretical underpinnings and practical suggestions. *ELT journal*, 49(2), 133-143.
[24] Webb, S., & Chang, A. C. S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37(4), 651-675.
[25] Webb, S., & Nation, P. (2017). How vocabulary is learned. *Oxford, UK: Oxford University Press*.
[26] Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading? A meta-regression analysis. *Studies in Second Language Acquisition*, 42(2), 411-438.
[27] Zandieh, Z., & Jafarigohar, M. (2012). The effects of hypertext gloss on. comprehension and vocabulary retention under incidental and intentional learning conditions. *English Language Teaching*, 5(6), 60-71.
[28] Zhang, C., & Ma, R. (2021). The effect of textual glosses on L2 vocabulary acquisition: A meta-analysis. *Language Teaching Research*.
[29] Zhang, S., Xu, H., & Zhang, X. (2020). The effects of dictionary use on second language vocabulary acquisition: A meta-analysis. *International Journal of Lexicography*, 34(1), 1-38.
[30] Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2023). *Sentiment Analysis in the Era of Large Language Models: A Reality Check*. arXiv preprint arXiv:2305.15005.
[31] Zhao, C. Y., Zhu, G B., & Wang J., Q. (2023). Insights and New Development Ideas for Language Models Brought by ChatGPT and Multimodal Models. *Data Analysis and Knowledge Discovery*, (03), 26-35.