# Effectiveness of High School Students' DDL of Verb-noun Collocations in an EFL Context

**Leshi Li, Xinyi Qi, Xuanye Tang, Jinghan Wang**

*School of International Education, Shanghai International Studies University, Shanghai, China*

**Abstract:** *This study explores the impact of data-driven learning (DDL) activities on the enhancement of collocation proficiency among Chinese high school students studying English as a foreign language (EFL). Additionally, participants' attitudes toward learning through the indirect DDL approach are examined. The empirical research spanned three weeks and involved two comparative groups, employing a combination of quantitative and qualitative methods with three instruments: a pre/post/delayed post-test, an attitude questionnaire, and interviews. The participants comprised 10 Chinese high school students. The control group followed traditional teaching methods, while the experimental group utilized DDL methods using the Corpus of Contemporary American English (COCA). Through nonparametric test analysis in SPSS, the results reveal minimal significant differences in corpus-assisted teaching of high school verb-noun collocation in the pre-test, post-test, and delayed post-test outcomes. However, a notable distinction emerges between the two groups in the writing segment of the delayed post-test, suggesting potentially deeper acquisition in the experimental group. Further empirical studies are expected to excavate its developmental potential and appropriate application for promoting high school student-centered learning in vocabulary.*

*Keywords: Data-driven learning; verb-noun collocations; high school students; EFL context*

## 1. Introduction

The foundation of English acquisition, vocabulary, especially the collocation of words, plays a vital role in representing people's linguistic proficiency (McCarthy, 1990). However, it is noticeable that collocations are still underestimated in the English as a Foreign Language (EFL) context, and several researchers have discovered that second language (L2) learners, find it difficult to learn collocations. According to Zou (2019), the proficiency of Chinese English learners in writing has been identified as lacking when it comes to understanding collocations. Jin and Webb (2020) found that L2 learners in the Chinese context encountered collocation problems learners, mainly for four reasons: limited processing resources available to attend to the collocating words, negative attendance to the unknown target collocations, too much reliance on the first language (L1) pattern, and overgeneralization of vocabulary. In this case, enhancing the collocation skills of L2 students through teaching methods is a significant research inquiry at the group level in the realm of foreign languages.

The blooming of technology advances the upcoming information era, bringing out a digitalized reform in China's compulsory education. As a crucial part of the connection of higher education, integrating information technology in high school English teaching has gained increasing attention regarding its instructional forms and methods. Recently, data-driven learning has surfaced as a novel approach, functioning to provide abundant authentic examples illustrating how words are employed in genuine contexts, offering a fresh perspective on various teaching methods associated with foreign language education. Many studies have proven the benefits of L2 learning when adopting the DDL strategy in pedagogy (Cotos et al., 2017; Crosthwaite, 2017; Lee & Lin, 2019). However, data-driven learning or Corpus research in English education lacking in the phases of secondary education. Thus, in order to gain an insight into the teaching model of English vocabulary analysis in high school, an evaluation study has been undertaken to observe whether the teaching methods assisted with corpora provide more appropriateness and individuality for Chinese high school students to learn English language than traditional teaching methods.

## 2. Literature review

### 2.1 Studies on Data-driven Learning (DDL)

#### 2.1.1 Definition of DDL

Data-driven learning (DDL), proposed by Johns (1991)[1], is a constructivist learning mode featured with the assistance of corpus in linguistic pedagogy. Students synthesize and form their schema based on their observation of abundant authentic language data from corpora, and therefore improve the learning effect. Two forms of linguistic education are applied in DDL—direct DDL and indirect DDL. Direct DDL needs learners to query the corpus by themselves on the computer, while indirect DDL requires teachers to collect and sort language data from the corpus beforehand for students to use (Yoon & Jo, 2014). Typical DDL performs in 3 phases: identification, classification, and generalization.

#### 2.1.2 Features of DDL

Many researchers have discussed the merits and insufficiencies of DDL. On one hand, the advantage of DDL is that it transforms the relationship between teachers and students in traditional language lessons. DDL makes students active researchers rather than passive listeners, and teachers collaborators instead of dominators. This provides students with the opportunity to solve problems, which facilitates students' abilities of self-learning, researching and problem-solving with their accumulating awareness in exposure to lexical patterns (O'Keeffe, 2021)[2].

On the other hand, DDL elicits problems too. In a review study, Boulton (2009) concluded from previous studies that the complexity of the data and concordance lines in the corpus can bring cognitive load to learners. One solution is to use mini-text, which is to select a few language data from a large-scale corpus to meet the teaching needs. Additionally, DDL provides only selected information instead of complete context, which also challenges many beginners. Regard of this, is implemented into language teaching in the integration of inductive approach and discovery learning (Flowerdew, 2015) in constructivism, scaffolding with "More Knowledgeable Other" in sociocultural theory (O'Keeffe, 2021) for students to reach their ultimate autonomy and proficiency in application[3].

All in all, DDL is capable of coping with some of the problems in traditional language teaching. Boulton and Vyatkina (2021) proposed its pedagogical potential to undergo experimentation and exploration in the future, especially in the field outside tertiary levels.

### 2.2 Verb-noun Collocation

Collocation is defined as the phenomenon where words are often used together in context (Hoey, 2013). As a key part of vocabulary discipline, it also can be divided into receptive knowledge and productive knowledge (Tran-Hoang-Thu & Alliant, 2009). The former refers to the knowledge used to comprehend words in listening and reading, and the latter stands for the knowledge used to produce words in speaking and writing (Zhou, 2010)[4]. Productive knowledge can be further subdivided into controlled productive knowledge and free productive knowledge. Controlled productive knowledge represents the ability to produce words with given cues and free productive knowledge is the ability to use words without prompts (Laufer, 1998). The development from receptive to productive knowledge reflects a higher learning level (Zhong, 2011). In the field of ELT, a growing awareness has been paid to the use of correct and appropriate collocations (Demir, 2017). Lewis et al. (1997) explored ways to introduce collocation in the classroom. Verb–noun collocations are particularly popular as targets for collocation learning (Boers et al., 2014) [5]because of their far-reaching effects on learners' language expression and writing skills. Many scholars have explored effective teaching methods for verb-noun collocation: Falahi and Moinzadeh (2012) reveal the positive effect of receptive and productive treatments on learning verb-noun collocations[6]. Tsai (2020) proves that the concept-based instruction and form-focused instruction approaches each had a positive effect on establishing initial form–meaning relationships for verb–noun collocations. Chan and Liou (2005) found that learners made significant collocation improvement immediately after the online practice using web-based Chinese-English bilingual concordance (keyword retrieval program) but regressed later. Contemporarily, exploring the DDL approach in teaching verb-noun collocation is gaining accumulating attention in Asian EFL classrooms (Kartal & Yangineksi, 2018)[7].

### 2.3 DDL in Vocabulary Learning

Much research on data-driven vocabulary learning has already been done, and a meta-analysis conducted by Lee et al. (2019)[8] shows that data-driven vocabulary learning is generally effective, especially for learners with high L2 proficiency. Also, in an earlier study by Lee et al. (2017), participants with higher English proficiency can better use the source of concordance lines from the corpus. These studies suggest that learners with relatively higher L2 proficiency can benefit more from data-driven vocabulary learning.

### 2.4 The Application of Data-driven Vocabulary Learning in High School

### 2.4.1 Previous empirical studies in China

The study of data-driven vocabulary learning in high school has been growing in the recent 5 years with the development of English teaching and research in secondary school. In terms of vocabulary learning, many studies have attempted to use DDL to teach the meaning of words. Some studies aim at distinguishing the meaning of different words. For instance, in one study, the students were taught to distinguish adjectives with two different suffixes —"-ful" and "-able" by corpus. Some studies try to teach the meaning of words by providing context from the corpus. One example is to use the "keyword list" to build the semantic field for learning[9]. These studies show that with DDL, learners can explore large language data by themselves, thus acquiring an immersive learning experience[10].

### 2.4.2 The insufficiency of the current studies

First, given that DDL works better for learners with high L2 proficiency, currently, it is used more in English lessons in university. Not many studies on data-driven vocabulary learning in high school have been done yet.

Second, current data-driven vocabulary learning mainly focuses on the meanings of vocabulary, rather than the comprehensive usage of vocabulary, like collocation. In some cases, when teaching the meanings of words, teachers and researchers may also mention the collocations. In one study, the teacher asked students to observe the language data from the corpus and judge which among the three new words- -*island*, *farm* and *dormitory*, one cannot be used with *on*[11]. But generally speaking, in data-driven vocabulary learning, not much attention is given to collocations, and further studies are called for.

## 3. Methodology

### 3.1 Research Designs

With a view to the contemporary situations in China's DDL-mode instruction, this essay is expected to figure out the following questions:

(1) For high school EFL learners, is there any difference between the control group and experimental group after the application of DDL in verb-noun collocations?

(2) What are students' feedback and attitudes towards the utilization of corpus in learning and teaching high school vocabulary?

### 3.2 Participants

Ten Chinese senior-three high school trainees volunteered to join the experiment of the study (ten females). The target participants are EFL freshmen who have just graduated from high school but haven't received any English knowledge in the university. After assessing their current vocabulary and collecting basic information, they were averagely distributed to the control group (G1) and the experimental group (G2) concerning their competency. The average age of the participants was 18 years old, and each of them reached a vocabulary level of more than 6000 words.

### 3.3 Instruments

In terms of instruments, a pre-test (PrT) post-test (PsT), delayed post-test (DPT) interviews, and a questionnaire are utilized in the research. This study chose verb-noun collocations as the target material for vocabulary study. Before the practical instruction, 30 were selected from the high-frequency Gaokao

collocation list for Shanghai high school students. Under investigation and filtering, 10 of them were selected for subsequent experiments, excluding too familiar or difficult ones (see Table 1). During data collection, efforts will be made to maintain a balanced representation of controlled variables, while eliminating interference from irrelevant ones.

*Table 1: Ten Target Collocations Used in the Experiment*

| No. | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Verb-noun Collocations | appreciate the scenery | become a hit | take one's temperature | make an apology | lead a ... life |
| No. | 6 | 7 | 8 | 9 | 10 |
| Verb-noun Collocations | take office | make sth. a priority | attend a meeting | recreate one's glory | satisfy demand |

The research design encompasses a mixed approach. This study initially employs quantitative surveys to probe the RQ1, arranging comparative study offline experiments that lasted for about three weeks. Participants were averagely divided into two groups, the controlled and the experimental group regarding their vocabulary test results and Gaokao performance. Before the actual instruction, the participants underwent a vocabulary proficiency assessment using an authoritative online vocabulary testing platform (https://preply.com/en/learn/english/test-your-vocab). During the experiments, two groups received instruction of the target collocations in divergent ways: Group 1, as the controlled group, learn 10 collocations in a traditional way of grammar-translation, where the teacher directly shows them materials with both Chinese and English definitions of the vocabulary, while the experimental group learn them under the guidance of teacher and the assistance of a commonly used corpus COCA (The Corpus of Contemporary American English), with the teacher firstly introducing the concept of DDL to participants and then referring to a worksheet for them to learn the new mode and apply it under the induction of the material and their individual practice on COCA. Then, the semi-structured interview and attitudes survey are conducted in supplementation of previous statistics.

### 3.4 Research Procedure

Prior to the experiment, we designed a teaching plan for empirical practice, worksheet for reinforcement and feedback chart for evaluation. To ensure the reliability and validity of the study, three tests were controlled in similar difficulty for the evaluation of participants' learning effects: PrT and PsT composes multiple-choice and translation questions targeting the 10 verb-noun collocations to test their basic output of the vocabulary. Notably, the order of the questions is adjusted for the second test in avoidance of repetitive retention. Two groups undergo the experiment in separate spaces. Before the lesson, the participants are invited to complete the PrT in 20 minutes. During the instruction, teachers distribute prepared paper worksheets. Extra materials including videos and worksheets were supplemented for the experimental group to adapt to the new mode. In the last 20 minutes of the course, participants were asked to finish the PsT. DPT was conducted two weeks later with an extra writing task to test participants' long-term memory and further application of free productive knowledge based on The Ebbinghaus Forgetting Curve. Then, the participants completed the 10-minute feedback questionnaire online, assessing their satisfaction towards the course, teacher and test. Finally, each of them received a semi-structured interview in regard to record their feedback and attitudes toward their learning.

Advance with formal instruction, a one-sample Mann-Whitney test of the PsT result is made to ensure the consistency of the two group's competency. Upon completion of data collection, the researcher will conduct data analysis. Initially, an exploratory factor analysis was applied to both tests and the questionnaire. Any issue would be addressed promptly, including contacting participants for clarification if necessary. Subsequently, the IBM SPSS 27.0 was applied to analyze the reliability of data in corpus-based ESP vocabulary teaching. After the homogeneity test of variance, the Mann-Whitney U test and Wilcoxon sum-rank test were conducted to calculate significant differences between and within the two groups as an attribution analysis to the first question. Lastly, interview records were interpreted as word clouds to explain the results and underlying reasons for the third research question. All data quantification tasks will be computer-assisted.

## 4. Results

In consideration of the relatively small sample size (n=10), this research adopts a non-parametric test with IBM SPSS 27.0 to analyze the data. To further excavate whether the DDL approach presented more

effective than the traditional approach for learners in acquiring verb-noun collocation knowledge, two paired sample Wilcoxon tests are first conducted to investigate whether the two approaches respectively present significant progress in the short-term acquisition of verb-noun collocations, and then, an independent sample Mann-Whitney test of gains for collocation is conducted to analyse the effectiveness of DDL approach applied in learning verb-noun collocations. The findings reveal as follows:

### 4.1 RQ1: DDL presents effectiveness in high school students' acquisition of verb-noun collocations.

*Table 2: Independent Sample Mann-Whitney Test Result between Control and Experimental Group*

| Total Score (maximum = 40) | Group | M ($p_{25}$, $p_{75}$) | Z | P |
|---|---|---|---|---|
| Pre-test (PrT) | control group (n=5) | 26.0 (24.0, 32.0) | -0.757 | 0.449 |
| | experimental group (n=5) | 24.0 (23.0, 29.5) | | |
| Post-test (PsT) | control group (n=5) | 36.0 (34.5, 38.5) | -0.631 | 0.528 |
| | experimental group (n=5) | 35.0 (30.5, 39.0) | | |
| Delayed Post-test (DPT) | control group (n=5) | 33.0 (29.0, 36.0) | -0.325 | 0.745 |
| | experimental group (n=5) | 32.0 (31.5, 36.0) | | |

The score of PrT for both the experimental group and control group were compared first to ensure the participants obtained a similar level of verb-noun collocation knowledge before the treatment. Figure 1 demonstrates no significant difference (p=0.449) between the control group (M = 26.0) and the experimental group (M = 24.0). In addition, the result illustrates no significance in comparison to the two groups in regard to their three tests (p > 0.05).

*Table 3: Paired-sample Wilcoxon Signed-rank Test among Pre-test, Post-test and Delayed Post-test within the Control Group*

| Control Group | Post-test - Pre-test | Delayed Post-test - Post-test | Delayed Post-test - Pre-test |
|---|---|---|---|
| Z | -2.023[a] | -1.826[b] | -2.041[a] |
| P | 0.043** | 0.068* | 0.041** |

Note: *P<0.1, **p<0.05
a. Based on negative ranks
b. Based on positive ranks

*Table 4: Paired-sample Wilcoxon Signed-rank Test among Pre-test, Post-test and Delayed Post-test within Experimental Group*

| Experimental Group | Post-test - Pre-test | Delayed Post-test - Post-test | Delayed Post-test - Pre-test |
|---|---|---|---|
| Z | -1.753[a] | -1.219[b] | -1.753[a] |
| P | 0.080* | 0.223 | 0.080* |

Note: *P<0.1, **p<0.05
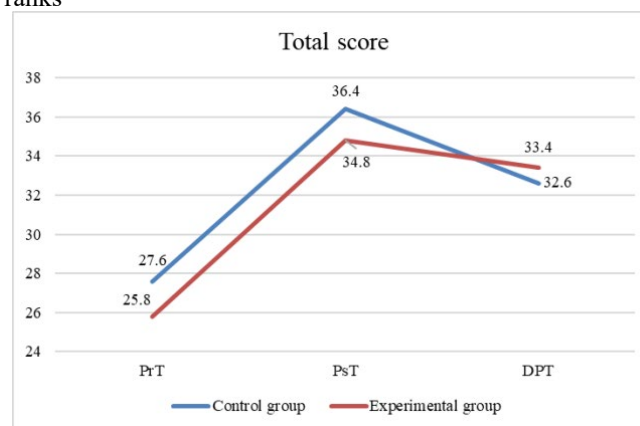a. Based on negative ranks
b. Based on positive ranks



*Figure 1: Line Chart on the Performance of Collocation Tests between the Experimental and Control Group*

Concerning the difference within the control and experimental group respectively, a paired-sample Wilcoxon signed-rank test was applied to the scores of PrT, PsT and DPT. As is shown in Tables 3 and

4, both present discrepancy after the instruction. Within the control group, there is a significant difference between PsT and PrT, PDT and PrT, (p<0.05), and a marginally significant difference between DPT and PsT, (p<0.1). While in Table 4, within the experimental group, there is a marginally significant difference between PsT and PrT, DPT and PrT (p<0.1), and no significant difference between DPT and PsT (p>0.1). Since the Z-score for DPT and PsT of the control group is based on positive ranks, it can be referred that the DPT score of the control group has inclined compared with the PsT score of the control group (as Tables5-7). But such a drop is not seen in the experimental group (see Figure1).

*Table 5: Descriptive Statistics of Control Group*

| Variable | N | Max | Min | Mean | SD | Median | Variance | Kurtosis | Skew | CV |
|---|---|---|---|---|---|---|---|---|---|---|
| Score of Pre-test(PrT) for Translation | 5 | 12 | 6 | 8.8 | 3.033 | 8 | 9.2 | -3.081 | 0.315 | 0.345 |
| Score of Pre-Test(PrT) for Multiple-choice | 5 | 20 | 18 | 18.8 | 1.095 | 18 | 1.2 | -3.333 | 0.609 | 0.058 |
| Score of Post-test(PsT) for Translation | 5 | 20 | 18 | 19.6 | 0.894 | 20 | 0.8 | 5 | -2.236 | 0.046 |
| Post-test(PsT) for Multiple-choice | 5 | 20 | 16 | 17 | 1.732 | 16 | 3 | 3.667 | 1.925 | 0.102 |
| Delayed Post-test (DPT) for Translation | 5 | 17 | 9 | 14.6 | 3.209 | 16 | 10.3 | 4.069 | -1.969 | 0.22 |
| Delayed Post-test (DPT) for Multiple-choice | 5 | 20 | 12 | 18 | 3.464 | 20 | 12 | 3.667 | -1.925 | 0.192 |

*Table 6: Descriptive Analysis of Composition in G1*

| Variable | N | Max | Min | Mean | SD | Median | Variance | Kurtosis | Skew | CV |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of Use | 5 | 9 | 4 | 6.8 | 1.924 | 7 | 3.7 | -0.022 | -0.59 | 0.283 |
| Number of Correct Use | 5 | 8 | 3 | 6 | 2 | 7 | 4 | -0.188 | -0.938 | 0.333 |
| Accuracy | 5 | 100 | 75 | 86.944 | 9.087 | 87.5 | 82.569 | 1.093 | 0.268 | 0.105 |

*Table 7: Descriptive Analysis of Composition in G2*

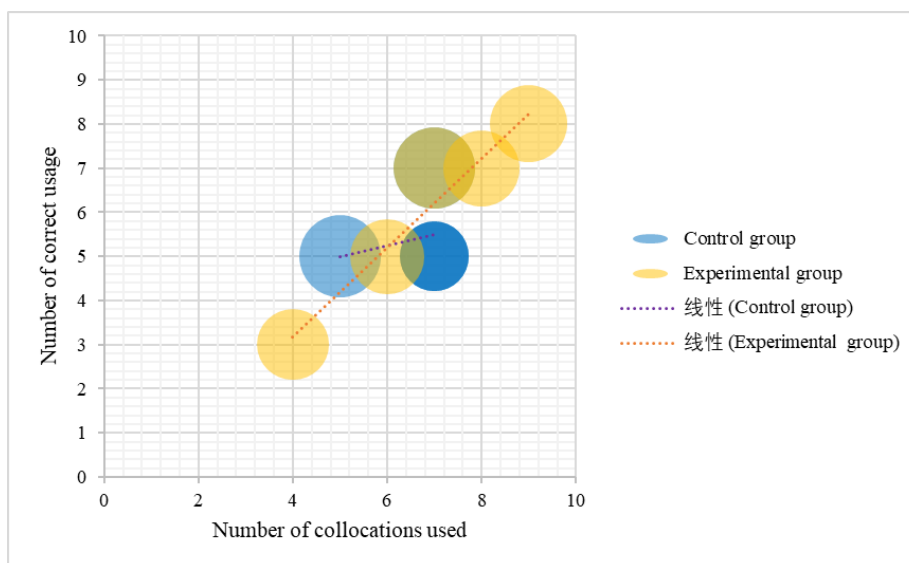| Variable | N | Max | Min | Mean | SD | Median | Variance | Kurtosis | Skew | CV |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of Use | 5 | 9 | 4 | 6.8 | 1.924 | 7 | 3.7 | -0.022 | -0.59 | 0.283 |
| Number of Correct Use | 5 | 8 | 3 | 6 | 2 | 7 | 4 | -0.188 | -0.938 | 0.333 |
| Accuracy | 5 | 100 | 75 | 86.944 | 9.087 | 87.5 | 82.569 | 1.093 | 0.268 | 0.105 |



*Figure 2: Bubble Diagram on the Performance of DPT between the Experimental and Control Group*

To further investigate the specific features of effectiveness in productive knowledge output between

two groups, a writing test was appended to the DPT. Figure2 illuminates the correctness rate concerning the total number of their applied target collocations and their correct usage. The size of the bubble positively correlates with the accuracy, and the darkness of the bubble is depending on the number of overlapping outcomes. From Figure2, the control group largely applied 7 collocations they acquired, while the experimental group demonstrated a more frequent usage of taught verb-noun collocation when writing the composition (mean=6.8). In addition, the yellow bubbles are relatively larger than the blue ones, illustrating that the average accuracy rate of the group (86.94%) is presented to be higher than the control group (82.86%).

### 4.2 RQ2: DDL receive averagely high satisfaction and support from high school students

The data of the in-depth face-to-face interviews with experimental group were examined from the following five aspects:

### 4.2.1 Participant's overall attitudes towards DDL by learning English collocations with the use of COCA

Overall, participants unanimously held a positive attitude towards the use of COCA for English vocabulary learning as having a positive effect on increasing engagement and depth of learning. Firstly, participants in all experimental groups generally recognized the positive effects of COCA in addressing the contextual nuances of words and providing a more comprehensive understanding of word usage and held positive attitudes towards this type of learning. Second, during the course of the experiment, participants expressed positive changes in their attitudes towards vocabulary learning, with a greater focus on contextualized learning and an emphasis on improving memory and comprehension through hands-on practice and exploration.

### 4.2.2 Participants' feelings of the advantages of learning collocations by DDL

Participants in experimental group found the most important advantages of DDL was that it emphasized the importance of independent learning and exploration through COCA, and the experience of looking up words through the corpus after the lesson and delving deeper into the content of the lesson made them more willing to actively engage in the learning process.

### 4.2.3 The problems of learning collocations by DDL pointed out by participants

Some participants mentioned operational difficulties that they may have encountered when they first encountered COCA, particularly concerning the login and registration process and the understanding of the system interface, suggesting that beginners may need better user guidance and introduction. Despite the operational challenges, most participants were satisfied with the classroom atmosphere created by the DDL, which they felt made the class more hands-on and contributed to better understanding and retention of vocabulary.

### 4.2.4 Participants' suggestions

Participants made several suggestions, including the provision of more detailed guides, introductions, and how-to guides, as well as better content screening to ensure accurate information when using COCA. In terms of timing, students consider that it might be more appropriate to begin the COCA learning style in their freshman and sophomore years of high school, when they have more time to learn and adapt, and for beginners who may need more guidance and translation.

In a nut shell, because COCA is a more innovative and scientific approach, the short-term level of assistance and classroom pleasure considered by the experiment is higher, which is also supported by all participants in the experimental group. However, at the same time, because they have not had much exposure to the course methods, they may excessively enjoy the formal learning methods rather than truly understanding classroom knowledge.

## 5. Discussion

From teaching plan to practice, this study lasts 1 year and aims to investigate the efficacy of corpus-assisted teaching methods compared to traditional approaches in English vocabulary analysis for high school students in China, designs to explore short-term and long-term learning outcomes, as well as their feedback and attitudes toward corpus application. We now address each research question with reflection and possible suggestions in turn.

Concerning RQ1, our results confirm that both DDL teaching mode and traditional grammar translation teaching mode is effective for students' learning verb-noun collocation. The delayed post-test was designed in exploration to the effects of students' long-term retention of new words (Al-Mahbashi et., 2017)[12]. For control group, an obvious drop of delayed post-test can be seen compared with post-test, which illuminates participants in control group had forgotten a large proportion of the collocation knowledge they learned during the two weeks before delayed post-test. Such phenomenon is not observed for experimental group, which indicates that DDL can promote longer memory duration for verb-noun collocation knowledge. Therefore, learning collocations from corpora leaves a deeper impression and allows for more efficient rectification in second language acquisition compared to traditional methods. Students who memorize the collocations directly from the substitution of meaning transferred from their mother language and the form of the vocabulary may underlyingly present the effect of traditional learning—repetitive, simple and direct. Such traditional approach fails to achieve the same effect because the collocations are acquired by learners discretely rather than in authentic context from corpora. Additionally, students may remember incorrect collocations influenced by their native Chinese language which are challenging to be corrected through rote learning. In contrast, the DDL mode gradually modifies students' linguistic patterns by exposing them to frequent usage in authentic contexts with abundant linguistic materials, indicating its potential in enhancement of student's learning autonomy, metacognitive capability, and other forms of long-term change in thought or action (Boulton & Cobb, 2017).

Compared with DDL, traditional instruction may fail to assist students in achieving effective learning due to the following drawbacks: Firstly, it only promotes a monotonous learning strategy, over-emphasizing rote-learning strategy in a relatively short time instead of bridging the second language in an authentic local context. This overlooks the significance of other key learning strategies such as cognitive, meta-cognitive and emotional strategies presented in the processing of corpus-assisted teaching. Second, it performed less attraction than DDL mode in terms of learning motivation.

The writing part of the delayed post-test crucially indicates the success of DDL. The experimental group showed significant improvement in accuracy and usage, demonstrating the practical application of collocation knowledge. The heightened accuracy in written compositions reflects DDL's impact on nuanced understanding and precise use of verb-noun collocations, achieved through engagement with authentic language data from corpora. Additionally, the increased utilization rates signify an effective transfer of learned collocations into practical linguistic contexts, revealing a profound assimilation of these lexical pairings. This highlights DDL's potential to shift from rote learning to functional language use, urging a shift in language education towards the holistic integration of vocabulary for authentic communication.

Regarding RQ2, we interviewed students about their previous vocabulary learning conditions and methods, students' feedback on experienced instructions, the influence of our course on them, their gains, attitudes and difficulties during the session. Qualitative outcome from interviews and questionnaires indicate that students exposed to corpus-assisted teaching demonstrated higher satisfaction levels in terms of course helpfulness and enjoyment. However, concerns were raised about the connection between corpus learning content and high school textbook content, as well as the adaptation of corpus. Although Student 10 indicated in the interview that she was unfamiliar with the corpus manipulation, the comparison of her pre- and post-test still showed significant improvement, especially the significant increase in her score in the translation (from 8 to 20), which showed an improvement in her productive knowledge. For example, she can correct the wrong collocation of "sense the scenery" to "appreciate the scenery".

The study acknowledges limitations related to sample size and suggests caution in generalizing results. This study therefore implemented interviews in an attempt to gain a more in-depth understanding of the sample to fill this gap. In addition, although this study focused on corpus manipulation during lectures, students still indicated during interviews that they needed more time to adapt and learn corpus manipulation, suggesting the importance of providing better user guidance and introductions for beginners (Leńko-Szymańska, 2014). Students lack exposure and preliminary knowledge of corpus, which means technological difficulties hinder them from acquiring collocations more fluently. Sometimes, mountains of data and concordance could be overwhelming for high school SLA learners since it takes more time and efforts for them to adapt to a new mode, while China's high school collocation study highlights scattering and repetitive input (Gu, 2018). Thus, both the training of especially in-service teachers and exploitation of more user-friendly corpus are expected for further exploration tackling such insufficiency with proper scaffolding to enhance the autonomy, data-driven perception, and language sense of learners (Bennett[13], & Dhonnchadha, 2023). Moreover, variations in

student engagement and attention during the experiment may introduce subjectivity into the results. To address this, future research should explore objective measures, such as monitoring eye movement (da Silva Soares Jr et al., 2023) or employing neurocognitive tools, to more accurately assess the impact of teaching methods on student attention and focus. Addressing these limitations will help us better understand how corpus-assisted teaching methods can be effectively applied in high school English vocabulary education.

## 6. Conclusion

In future studies, it would be advantageous to encompass a wider array of linguistic features rather than exclusively focusing on a single aspect. This is crucial because language learners are anticipated to achieve proficiency in a comprehensive manner. Besides, future research could explore adapting these methods for students at different language proficiency levels, especially in bridging high school and college English studies.

As an innovative exploration of the English vocabulary learning method, this study contributes to understanding English vocabulary analysis teaching models in high schools, especially in the Chinese context. The findings offer valuable insights for educators aiming to enhance teaching practices in high school English vocabulary instruction.

## Acknowledgement

## References

[1] Al-Mahbashi, A.A., Noor, N.M., & Amir, Z.The Effect of Multiple Intelligences on DDL Vocabulary Learning. International Journal of Applied Linguistics and English Literature, 2017, 6, 182-191.
[2] Da Silva Soares Jr, R., Barreto, C., & Sato, J. R.Perspectives in eye-tracking technology for applications in education. South African Journal of Childhood Education, 2023, 13(1), 1204.
[3] Demir, C. Lexical collocations in English: A comparative study of native and non-native scholars of English. Journal of Language and Linguistic Studies, 2017, 13(1), 75-87.
[4] Dugard, P., & Todman, J. Analysis of pre-test-post-test control group designs in educational research. Educational Psychology, 1995, 15(2), 181-198.
[5] Bennett, C., & Dhonnchadha, E. U.Becoming corpus literate: An in-service EFL teacher education framework for integrating corpora into EFL teaching. Applied Corpus Linguistics, 2023, 3(1), 100048.
[6] Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. Gauging the effects of exercises on verb–noun collocations. Language Teaching Research, 2014, 18(1), 54-74.
[7] Boulton, A.Data-driven learning: Reasonable fears and rational reassurance. Indian Journal of Applied Linguistics, 2009, 35, 81–106.
[8] Boulton, A., & Cobb, T.Corpus use in language learning: A meta-analysis. Language learning, 2017, 67(2), 348-393.
[9] Cotos, E., Link, S., & Huffman, S. Effects of DDL technology on genre learning. Language Learning & Technology, 2017, 21, 104-130.
[10] Crosthwaite, P. Retesting the limits of data-driven learning: Feedback and error correction. Computer Assisted Language Learning, 2017, 30(6), 447-473.
[11] Falahi, M., & Moinzadeh, A.Effects of receptive and productive tasks on Iranian EFL students' learning of verb-noun collocations. Journal of Language Teaching and Research, 2012, 3(5), 953.
[12] Boulton, A., & Vyatkina, N. Thirty years of data-driven learning: Taking stock and charting new directions over time. Language Learning & Technology, 2021, 25(3), 66–89.
[13] Chan, T. P., & Liou, H. C. Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations. Computer assisted language learning, 2005, 18(3), 231-251.