

International Journal of Learning, Teaching and Educational Research
 Vol. 23, No. 8, pp. 98-115, August 2024
<https://doi.org/10.26803/ijlter.23.8.6>
 Received May 25, 2024; Revised Aug 11, 2024; Accepted Aug 13, 2024

EFL Learners' Use of Data-driven Learning and their Attitudes in the Improvement of English-speaking Skills

Zhaoyi Pan* 

King Mongkut's University of Technology Thonburi
 Bangkok, Thailand

Abstract. The use of data-driven learning (DDL) to improve the English-speaking skills of learners of English as a foreign language (EFL) remains rare. Hence, this research examines whether two types of DDL – namely, hands-on DDL using the computer software directly and hands-off DDL using paper-based materials – can improve vocabulary production in the English speaking of EFL learners. The EFL learners' attitudes toward both types of DDL were also examined. A total of 45 Thai EFL learners were involved in this research; they were divided equally into two experimental groups, one using hands-on DDL and hands-off DDL, and one control group. A questionnaire and interviews were used to examine the EFL learners' attitudes toward DDL and a paired sample t-test and a one-way analysis of variance (ANOVA) were conducted. The results reveal that both hands-on and hands-off DDL approaches significantly improved vocabulary production in the EFL learners' spoken English. In addition, the hands-on DDL had a significant effect on the quantity (sig. = .000, $p < 0.05$), accuracy (sig. = .000, $p < 0.05$), and complexity (sig. = .000, $p < 0.05$) of the participants' vocabulary production, while the hands-off DDL only had a significant effect on the accuracy (sig. = .000, $p < 0.05$) of vocabulary production. Furthermore, although the EFL learners had relatively positive attitudes toward DDL, less enjoyable experiences were also noted. Experiences of boredom and stress while using DDL were reported, and the participants did not consider DDL to be suitable for all EFL learners.

Keywords: data-driven learning; vocabulary production; English speaking; learners of English; productive skill

1. Introduction

Data-driven learning (DDL) has recently been used as an approach to teaching and learning English as a foreign language (EFL) among diverse EFL learners in various educational settings (Mizumoto, 2023; Zare & Delavar, 2022). It is

* Corresponding author: *Zhaoyi Pan, zhaoyi.pan@sit.kmutt.ac.th*

considered to be an effective intervention for EFL learners that enhances various EFL skills.

To date, the communicative language teaching (CLT) approach has mainly been used to enable EFL learners to improve different aspects of their speaking skills (Hui & Yunus, 2023). Although the CLT approach can improve the speaking skills of EFL learners, its capacity for improving specific components of speaking skills, such as the accurate use of vocabulary, remains limited (Abdelmageed & Omer, 2020). On the other hand, despite the extensive use of DDL in various EFL educational settings, little research has yet been focused on the use of DDL to improve EFL learners' speaking skills (Ueno & Takeuchi, 2023). Thus, it is significant to combine DDL with the CLT approach with the aim of improving EFL learners' speaking skills.

Accordingly, the present research attempts to employ DDL to improve EFL learners' speaking skills by focusing on the three key aspects of vocabulary; namely quantity, accuracy, and complexity. First, the study examined whether hands-on and hands-off DDL approaches could significantly improve the vocabulary production of EFL learners when speaking in English; it then examined whether there was a significant difference between the use of hands-on and hands-off DDL approaches in improving the EFL learners' vocabulary productions. Finally, the study examined EFL learners' attitudes toward the use of DDL to improve their vocabulary production in spoken English. The three research questions are presented below.

- 1). Do hands-on and hands-off DDL approaches significantly improve EFL learners' vocabulary production when speaking in English?
- 2). Is there a significant difference in EFL learners' vocabulary production when speaking in English using the hands-on DDL approach, the hands-off DDL approach, and traditional approaches?
- 3). What are EFL learners' attitudes toward using hands-on and hands-off DDL approaches?

2. Literature Review

2.1. Development of Data-driven Learning

Defined as "an inductive approach that utilizes the tools and techniques of corpus linguistics for instructional objectives" (Zare et al., 2022, p. 2), DDL has been used as a teaching approach to enhance various aspects of EFL learners' English skills (Boulton & Cobb, 2017; Tosun & Sofu, 2023). The inclusion of DDL in the pedagogical process can be attributed to innovations and the development of relevant technologies in the field of corpus linguistics. Technological programs such as AntConc allow for the possibility of using DDL in the processes of teaching and learning. The two main types of DDL are hands-on DDL and hands-off DDL. Hands-on DDL is that in which "learners use corpora or concordancing software directly to discover the patterns of language use by themselves," whereas hands-off DDL is an indirect approach whereby teachers "simplify and modify the materials, and then present learners" with handouts that are tailored to match the learners' needs (Saeedakhtar et al., 2020, p. 3), such

as vocabulary handouts in English reading. Most of the research indicates that EFL learners experience significant improvements in a variety of EFL skills after using either type of DDL in the teaching and learning process compared to traditional methods (Lin, 2021; Pérez-Paredes et al., 2019; Zare et al., 2022).

One of the major benefits of using DDL, particularly hands-on DDL, is the facilitation of learner autonomy (Boulton & Cobb, 2017; Lee & Lin, 2019; Pérez-Paredes et al., 2019). Learner autonomy is reflected in the entire process of using DDL, including the training in the use of technological computer programs, the observation of concordancing, the inductive process of analyzing the concordancing in accordance with the EFL learners' focus, and the relevant follow-up practices that are designed by researchers or instructors (Karpenko-Seccombe, 2023; Muftah, 2023; O'Keeffe, 2021). In this process, the traditional teacher-centered style of classroom teaching is naturally transformed into a student-centered learning process, with the role of the teacher being that of an instructor who guides and assists EFL learners to understand concordancing and ensures correct observations of the data in different corpora (Murad et al., 2023).

It is clear that DDL should be considered as a learning method in the EFL input process. If DDL is mainly used to provide EFL input, productive skills must be linked to the input in this approach. Despite little of the existing research having highlighted this connection directly, one of the major contributions of DDL that has been identified in various studies is that, from the cognitive perspective, it allows for the retention of different types of linguistic knowledge in EFL learners' minds (Lay & Yavuz, 2020; Lee & Lin, 2019; Lin, 2021; Saeadakhtar et al., 2020; Zare et al., 2022). Retention can only be achieved via sufficient repeated encounters with particular EFL knowledge, and DDL can provide "a more interactive and context-heavy interface" through which EFL learners can increase the number of their encounters (Lay & Yavuz, 2020, p. 178). Although DDL can assist in retention via the contexts provided by concordancing and computer programs, it is essential for EFL learners to practice their productive skills independently. Hence, an eclectic teaching and learning process should be adopted to develop EFL learners' speaking skills as opposed to solely relying on DDL (O'Keeffe, 2021).

2.2. Communicative Language Teaching

Several recent studies on the effects of DDL on different EFL skills in the teaching and learning process have included experiments using mixed-method approaches (Corino & Onesti, 2019; Lusta et al., 2023; Zare et al., 2022; Zhang, 2022). A mixed-method approach suggests that DDL should be used in conjunction with other approaches in order to improve productive skills (Lusta et al., 2023). Since DDL states that a high number of encounters with specific EFL knowledge is mainly responsible for the retention of such knowledge when EFL learners experience the input process, it is clear that more processes or interventions should be included to improve productive skills. In EFL teaching and learning research globally, CLT has been mainly used in investigations of speaking skills, and has been "prioritised to teach spoken interaction" (Hui & Yunus, 2023, p. 1515). Given the increased focus on the actual language use of

non-native English speakers, CLT is able to promote performances in EFL teaching and learning processes by focusing on the interactions between EFL learners (Hirschi & Kang, 2024; Hui & Yunus, 2023; Pratiwi et al., 2024). In classroom activities, the CLT approach creates more opportunities for EFL learners to speak English, allowing them to engage in various tasks designed to improve their communication skills, such as role plays, group discussions, and debates (Corino & Onesti, 2019; Hui & Yunus, 2023), which encourage EFL learners to speak English. In the language classroom, a task refers to an activity “in which learners use language to achieve a real outcome” (Zare & Delavar, 2022, p. 2). Tasks are comparable to real-world interactions, thus encouraging EFL learners to use English in oral communication (Zare & Delavar, 2022; Zare et al., 2022). In this research, a mixed-method approach was employed via the combination of DDL and CLT in the EFL classroom.

Learner autonomy is the main aim of pedagogical and instructional approaches that include DDL and CLT (Hirschi & Kang, 2024; Mizumoto, 2023; O’Keeffe, 2021; Pérez-Paredes et al., 2019). With regard to achieving learner autonomy, both approaches can benefit the learning process by activating the EFL learners’ metacognition (Mizumoto, 2023). Metacognition refers to the learner’s “understanding of one’s own cognitive abilities and personal attributes as a cognitive processor” (Mizumoto, 2023, p. 3). Using DDL can increase EFL learners’ interest in the EFL knowledge on which they are attempting to focus, with EFL learners using their metacognition to evaluate whether they truly comprehend the EFL knowledge that they are investigating via concordancing (O’Keeffe, 2021; Pérez-Paredes et al., 2019). Furthermore, guided by CLT in EFL classrooms, various speaking tasks trigger the metacognition of each EFL learner to notice whether their retention of the relevant EFL knowledge can be translated into output in their verbal communication (Hui & Yunus, 2023; Pratiwi et al., 2024). Both steps rely on metacognition driving EFL learners to automatically input and output simultaneously when using DDL and CLT. By using mixed methods, the strengths of DDL can be harnessed in the EFL teaching and learning processes, while learner autonomy can be better achieved via productive results.

2.3. Speaking Skills and Data-driven Learning

EFL learners’ speaking skills cover a large range of concepts, including pronunciation, vocabulary, grammar, utterances, sequential orders, fluency, pragmatic use, and even gestures and eye contact (Zhussupova & Shadiev, 2023). Contrary to the use of DDL in previous research, the present research focuses on three key factors of vocabulary: quantity, accuracy, and complexity (Kızıl, 2023). Based on the findings of previous research, three indicators of mastery in productive skills were included in the assessment of speaking skills, including the quantity, accuracy, and complexity of vocabulary use (Kızıl, 2023). Quantity reflects the fluency and range of vocabulary use; accuracy demonstrates whether EFL learners have the ability to use correct vocabulary in situational interactions, and complexity refers to the EFL learners’ ability to use sophisticated and varied vocabulary (Hui & Yunus, 2023; Karpenko-Secombe, 2023). In the field of corpus linguistics, these three factors can be assessed based

on the tokens and types of vocabulary that EFL learners use in oral communication (Kızıl, 2023). Moreover, the EFL learners' attitudes toward the use of DDL for vocabulary improvement in speaking skills are also worthy of study.

When examining the efficiency of DDL, previous research has also examined the attitudes and emotions of EFL learners regarding the use of DDL. Most of the results have suggested that DDL is more accepted by EFL learners who have relatively high English proficiency, such as upper-intermediate and advanced levels based on the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2020; Pérez-Paredes, 2019; Tosun & Sofu, 2023; Zare et al., 2022). By contrast, some studies have found that EFL learners' attitudes toward the use of DDL may change over time (Hirschi & Kang; 2024; Tosun & Sofu, 2023). However, some EFL learners experienced negative emotions, such as stress and anxiety, when they observed the overwhelming concordancing and attempted to output everything that they had learned via DDL (Karpenko-Seccombe, 2023). Although recent research has attempted to use DDL with EFL learners who had relatively low English proficiency levels, these learners found using DDL to be stressful because they needed to be trained to use particular computer programs and to summarize the use of certain vocabulary by themselves based on their own observations, which appeared to be difficult (Lin, 2021; Saeedakhtar et al., 2020). Due to these differing results, the attitudes of EFL learners toward using DDL with CLT to improve their vocabulary in speaking skills were further examined in this research.

3. Methodology

3.1. Participants and Data Collection

As a result of the researcher providing details of this research to a university campus in Thailand, a total of 45 students voluntarily participated in this study. All of the participants were Thai nationals who were studying three different majors at two public universities in Bangkok, Thailand, during the data collection period. Prior to the data collection, they had been EFL learners for thirteen to fifteen years, although none of the participants had studied EFL in an English-speaking country. Each participant had previously obtained a valid score on an international English exam, such as IELTS. The participants' English proficiency was at the intermediate level, as were their speaking skills, based on the CEFR (Council of Europe, 2020). Intermediate-level EFL participants were selected for this research since it is the required level for Thai university students, based on the Thai education policy. Among the participants, 32 were male (71%) and 13 were female (29%). However, gender was not considered as a variable in this research.

Each participant consented to the entire research process prior to the data collection. Following previous research on DDL, the researcher conducted a quasi-experimental study, whereby the participants were randomly divided into three groups. The first group of 15 participants used the hands-on DDL and CLT approaches. The second group of 15 participants used the hands-off DDL and CLT approaches, while the third group of 15 participants used dictionaries and

online machine translations, such as Google Translate, and CLT approaches. Therefore, the first two groups constituted the experimental groups, whereas the third group represented the control group. It should be noted that the teaching and learning processes for the three groups in this research were not included in any of the participants' prescribed English courses. All of the participants were informed that the instructional processes used in their groups were aimed at improving their vocabulary output when speaking, and that all of the procedures in which they engaged during the data collection process were outside of the pedagogical requirements of their courses. Each participant was attending the English courses required by their majors, according to the curricula and pedagogies. Thus, despite any significant differences in the results of the three groups, the participants' usual EFL studies and progress – according to their pedagogical requirements – remained unaffected. The aim was to ensure compliance with ethical issues concerning fairness in the EFL learners' English studies in their majors.

The entire experiment lasted for ten weeks. Each weekend, the researcher, as the instructor, met each group when all the participants were available for approximately one-and-a-half hours. Therefore, each group received a total of 15 hours of the intervention to improve their vocabularies when speaking. Table 1 provides the details of the experimental process.

Table 1: Experimental process of this research

weeks	hands-on DDL	hands-off DDL	control group
1	pretest instruction of DDL and COCA	pretest instruction of DDL	pretest
2-9	DDL with COCA and CLT	DDL with the materials prepared by the researcher and CLT	Vocabulary learning by using dictionaries and online translation programs and CLT
10	posttest	posttest	posttest

In the first week, the 15 participants in each group were divided into three teams of their choice. Each team completed a pretest and a posttest in Week 1 and in Week 10, respectively. In both the pretest and in the posttest, each team was asked to conduct a group discussion about the four topics of daily life, friendship, social media, and high technology. Each group discussion was intended to last for 15 minutes, during which time the students were allowed to discuss anything that they wanted to mention about each topic. The pretest and the posttest were audio recorded using iPhone voice memos. Both of these tests were conducted to examine the quantity, accuracy, and complexity of the participants' vocabulary use. A detailed analysis will be provided in the following section. In the first week, the researcher guided each participant to study the basic concept of DDL. The Corpus of Contemporary American English (COCA; Davies, 2008) was selected because it is recognized as being the largest corpus, containing American English in various genres, including the spoken genre (Davies, 2008; Lusta et al., 2023). For the first experimental group, which used the hands-on DDL approach, the instructor guided each participant to

learn how to use the different functions of COCA, including searching keywords, keywords in context, frequency lists, and texts. For the second experimental group, which used the hands-off DDL approach, the researcher collected the vocabulary from the three teams in the second group according to their speaking performances each week and created hands-off DDL scripts for them. For example, based on the pretest, the researcher identified the vocabulary that the participants needed to improve in both experimental groups. The first group used COCA to input the vocabulary on their own but with the researcher's assistance when necessary, while a hands-off DDL script that the researcher had created from the concordancing in COCA was distributed to the participants in the second group for input, and the participants in the third group used dictionaries and online translation programs to input the vocabulary. From Week 2 to Week 9, the participants were requested to focus on each topic every two weeks. Each week, approximately 15 new words were input by each participant in the three groups. After they had completed the input process, the students were asked to use the vocabulary in group discussions each week. They were guided to use the vocabularies extensively in each group discussion to activate their metacognitive mechanisms; thus, they identified those vocabularies with which they still had difficulty using, and they were allowed to input them continuously each week. Following the posttest, a 10-item questionnaire that had been adapted from previous research was given to each participant to conclude the data collection, as presented in the following section (Tosun & Sofu, 2023). A short interview with each team in each group was also conducted and an extensive explanation of each question in the questionnaire was provided.

3.2. Data Analysis

To determine whether the use of hands-on and hands-off DDL could improve the vocabulary output in the English speech of EFL learners, a significant difference test should be conducted to examine the pretest and posttest scores for each experimental group. Hence, a paired-sample *t*-test was used. The test scores were the dependent variables. The scores were paired, with each participant's scores from the pretest and the posttest before and after the DDL in two experimental groups, respectively. To examine whether there was a significant difference in vocabulary production among the three groups, the independent variable was the teaching and learning approach (the categorical factor), while the dependent variable was the posttest result of each key factor (the continuous factor). Hence, a one-way analysis of variance (ANOVA) was conducted, followed by a post hoc test. The dependent variable, the posttest result of each key factor, was assumed to be normally distributed within each group. Homogeneity of variance was assumed, indicating that the variability in the posttest result of each key factor was similar across all groups.

Following the previous research and corpus linguistics, the three key factors mentioned above were measured as follows: The quantity of the use of the vocabulary was calculated based on the average number of vocabularies produced by each participant in each group in the pretest and in the posttest. To determine the accuracy, the ratio between the error-free vocabularies and the

total vocabulary was used (Kızıl, 2023). Finally, complexity was determined based on the ratio of the “total number of word types divided by the square root of two times the total number of word tokens” (Kızıl, 2023, p. 1386), namely $\text{word types}/\sqrt{\text{word token} \times 2}$. AntConc (2023) was used to calculate the tokens and types in the vocabularies that the EFL learners used in the pretest and in the posttest. Two raters, who were native English speakers and taught EFL, examined the errors in the use of the vocabularies to ensure reliability. The coefficient correlation for both raters was 0.946, indicating high reliability. Together, the raters discussed the different error markings until reaching a consensus. The Statistical Package for the Social Science (SPSS) Mac OS Version was used to obtain the results of both the significant difference tests mentioned above. SPSS was also used to calculate the mean scores for each question in the questionnaire to examine each EFL learner’s attitudes toward the use of DDL to improve their vocabulary output when speaking in English. The researcher categorized the interview responses that had the same semantic meaning using the keyword search function in AntConc.

4. Results

4.1. Quantitative Results

After the completion of the ten-week experiment, the descriptive statistics for the pretests and the posttests of the three groups were calculated to provide an overview of the vocabulary output in the EFL learners’ spoken English, as shown in Table 2 below.

Table 2: Vocabulary performance by EFL learners in three groups

test	group	quantity		accuracy		complexity	
		mean	SD	mean	SD	mean	SD
pretest	1	194	5.26	2.19	0.97	8.02	1.44
	2	213	5.60	1.95	0.84	8.76	1.56
	3	198	5.31	2.05	0.88	8.23	1.49
posttest	1	209	5.68	2.45	1.30	12.11	1.72
	2	216	5.72	2.31	1.21	10.04	1.68
	3	205	5.61	2.24	1.02	8.95	1.50

Based on the descriptive statistics in Table 2, it can be seen that each group made certain improvements in the quantity, accuracy, and complexity of their vocabulary output in spoken English. To further examine whether the hands-on and hands-off DDL approaches significantly improved the vocabulary output in the English speech of the EFL learners, Table 3 presents the results of the paired-sample *t*-test for the pretests and the posttests of Groups 1 and 2.

Table 3: Paired sample t-test of each key factor of groups 1 and 2

pretest-posttest	group	mean	SD	t	sig.
quantity	1	-15	4.29	-28.67	.000
	2	-3	2.17	-12.95	.815
accuracy	1	-0.26	1.94	-2.64	.000
	2	-0.36	3.37	-5.18	.000
complexity	1	-4.09	6.74	-10.75	.000
	2	-1.28	4.05	-8.67	.627

The results of the significant difference test between the pretests and the posttests in Groups 1 and 2 revealed the effectiveness of the use of the hands-on and hands-off DDL approaches regarding the EFL learners' vocabulary production in the English speech. With regard to the quantity of the vocabulary production, the hands-on DDL had a significant impact (sig. = .000, $p < 0.05$), while the hands-off DDL had no significant impact (sig. = .815, $p < 0.05$). By contrast, both the hands-on and the hands-off DDL approaches had a significant influence on the accuracy of the vocabulary production (sig. = .000, $p < 0.05$). Similar to the findings for the quantity of the vocabulary production, the hands-on DDL had a significant impact (sig. = .000, $p < 0.05$), while the hands-off DDL had no significant impact (sig. = .627, $p < 0.05$). Statistically, these results show that the hands-on DDL approach was able to improve vocabulary production in the English speech of EFL learners in all three key factors, whereas the hands-off DDL approach only significantly improved the accuracy of the EFL learners' vocabulary production.

As mentioned in the previous section, the one-way ANOVA test was conducted, followed by a post hoc test, to examine whether there was a significant difference in the results for the three key factors in the posttests of vocabulary production for either group. First, based on the result of the Test of Homogeneity of Variances, provided by SPSS, homogeneity of variance was sig. = .532, $p < 0.05$, thus indicating that the variance was equal in each group. The first one-way ANOVA test aimed to examine the influence of the different approaches on the quantity of vocabulary production in spoken English. Table 4 and Figure 1 illustrate the comparison of each approach and the quantity of vocabulary production when speaking in English.

Table 4: Multiple comparisons of different approaches to quantity

approach	approach	mean difference	std. error	sig. ^b
hands-on	hands-off	-7*	3.16	.012
	traditional	4*	4.27	.043
hands-off	hands-on	7*	3.16	.012
	traditional	11*	3.59	.000
traditional	hands-on	-4*	4.27	.043
	hands-off	-11*	3.59	.000

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

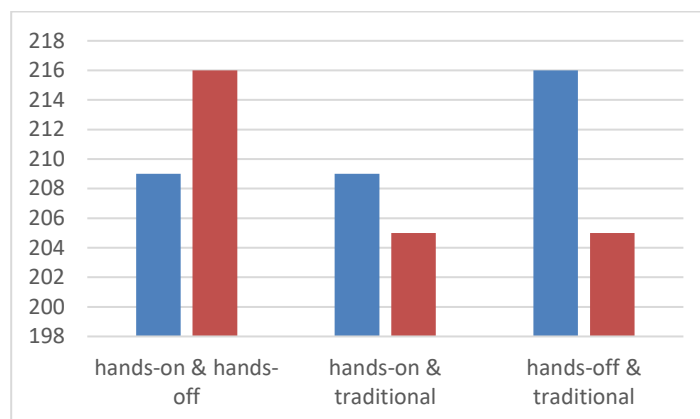


Figure 1: Comparison of mean difference in the vocabulary quantity in three groups

According to the results of the one-way ANOVA, there was a significant difference in the EFL learners' quantity of vocabulary production when speaking in English as a result of the different approaches that were employed (sig. = .036, $p < 0.05$). As can be seen in Table 4, there was a significant difference between various pairs of hands-on DDL approaches, hands-off DDL approaches, and the traditional approach. Furthermore, the hands-off DDL approach had the biggest significant impact on the quantity of vocabulary production in the EFL learners' spoken English because the mean difference between the hands-off DDL approach and the traditional approach was 11 (sig. = .000, $p < 0.05$), and the mean difference between the hands-off DDL approach and the hands-on DDL approach was 7 (sig. = .012, $p < 0.05$). Hence, the hands-off DDL approach can be considered to be the most effective approach for increasing the EFL learners' quantity of vocabulary production when speaking in English.

Table 5 and Figure 2 illustrate the comparison between each approach and the accuracy of vocabulary production when speaking in English.

Table 5: Multiple comparisons of different approaches to accuracy

approach	approach	mean difference	std. error	sig. ^b
hands-on	hands-off	0.14*	2.05	.000
	traditional	0.21*	2.58	.000
hands-off	hands-on	-0.14*	2.05	.000
	traditional	0.07*	0.87	.661
traditional	hands-on	-0.21*	2.58	.000
	hands-off	-0.07*	0.87	.661

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

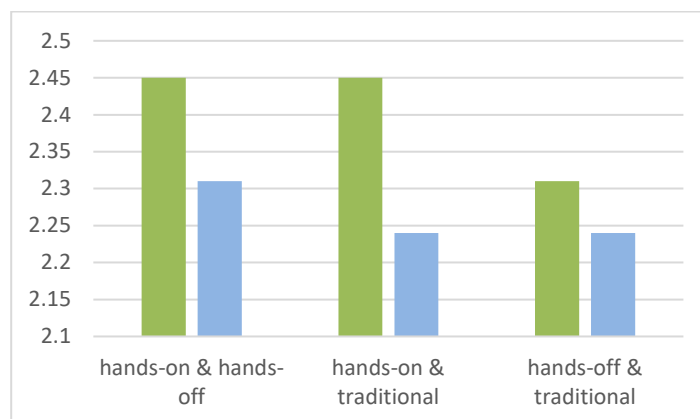


Figure 2: Comparison of mean difference in the vocabulary accuracy of three groups

The results of the one-way ANOVA showed that there was a significant difference in the accuracy of the EFL learners' vocabulary productions when speaking in English using different approaches (sig. = .000, $p < 0.05$). Based on the comparisons of the approaches, illustrated in Table 5, there was a significant difference in the EFL learners' accuracy in vocabulary production when speaking in English between the two pairs of the approaches; namely, the hands-on DDL and the hands-off DDL approaches (sig. = .000, $p < 0.05$), and the hands-on DDL approach and the traditional approach (sig. = .000, $p < 0.05$). By contrast, there was no significant difference in terms of the EFL learners' accuracy in their vocabulary production when speaking in English between the hands-off DDL approach and the traditional approach (sig. = .661, $p < 0.05$). The results suggest that the hands-off DDL approach may not significantly improve the accuracy of the EFL learners' vocabulary production when speaking in English, while the hands-on DDL approach could significantly improve the EFL learners' accuracy in vocabulary production in spoken English.

Finally, Table 6 and Figure 3 present the comparisons of each approach and the complexity of the participants' vocabulary production when speaking in English.

Table 6: Multiple comparisons of different approaches to complexity

approach	approach	mean difference	std. error	sig. ^b
hands-on	hands-off	2.07*	3.27	.000
	traditional	3.16*	4.86	.000
hands-off	hands-on	-2.07*	3.27	.000
	traditional	1.09*	1.43	.713
traditional	hands-on	-3.16*	4.86	.000
	hands-off	-1.09*	1.43	.713

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

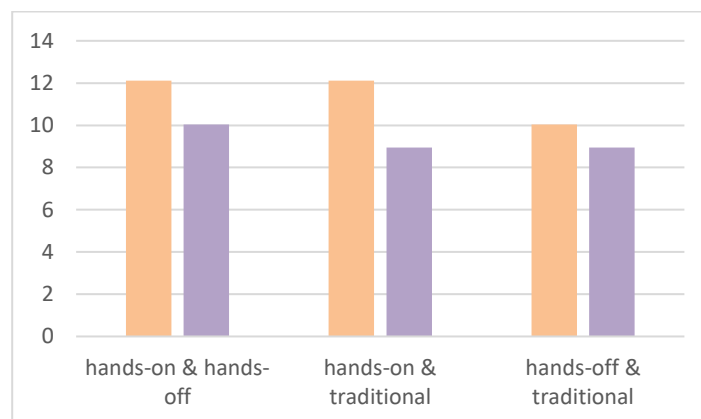


Figure 3: Comparison of mean difference in the vocabulary complexity of three groups

Based on the results of the one-way ANOVA, there was a significant difference in the complexity of the vocabulary production in the English speech of EFL learners when using different approaches (sig. = .000, $p < 0.05$). According to the results of the comparisons, presented in Table 6, two pairs of approaches to improving the complexity of the vocabulary production in the English speech of EFL learners produced significant differences; that is, between the hands-on and the hands-off DDL approaches (sig. = .000, $p < 0.05$), and between the hands-on DDL approach and the traditional approach (sig. = .000, $p < 0.05$). By contrast, there was no significant difference between the use of the hands-off DDL approach and the traditional approach for increasing the complexity of vocabulary production in the EFL learners' spoken English (sig. = .713, $p < 0.05$). Hence, these results indicate that the use of the hands-on DDL approach can increase the complexity of vocabulary production in the English speech of EFL learners most effectively, whereas the hands-off DDL approach may not be effective for increasing the complexity of vocabulary production in EFL learners' spoken English.

4.2. Analysis of the Questionnaire

The mean for each question in the questionnaire was first calculated, as illustrated in Table 7 below.

Table 7: Questionnaire adapted from the previous research (Tosun & Sofu, 2023)

Questions: I think...	Group 1		Group 2	
	mean	SD	mean	SD
1. studying vocabulary through concordance lines is enjoyable.	3.26	0.95	3.12	1.19
2. studying vocabulary through concordance lines helps improve my English.	4.25	1.52	3.97	1.34
3. using concordance lines improved my English speaking	4.46	0.84	3.95	1.06
4. using concordance lines is helpful for learning the usage of vocabulary.	4.72	0.93	4.15	0.96
5. using concordance lines in the learning of English vocabulary increased my confidence in learning English vocabulary.	4.39	0.82	4.27	0.97
6. learning vocabulary through concordance lines is more difficult than learning vocabulary through a dictionary or	4.16	1.32	4.38	0.82

online translation program.				
7. learning vocabulary through concordance lines is more boring than learning vocabulary through a dictionary or online translation program.	3.61	0.79	3.53	1.20
8. I prefer using concordance lines in learning of English vocabulary to using a dictionary or online translation program in learning of English vocabulary.	4.07	0.95	3.64	0.94
9. I recommend that teachers should use concordance lines so as to teach vocabulary in EFL classes.	3.54	0.83	3.28	1.02
10. I would like to do more concordance activities in class.	3.91	0.94	3.58	0.98

Based on the mean for each question in the questionnaire, certain discrepancies among different questions can be noted. With regard to the hands-on DDL approach, Items 4, 3, and 5 had the highest means, with 4.72, 4.46, and 4.39, respectively. According to the scores for these three items, the EFL learners in Group 1 found the hands-on DDL approach to be helpful and effective in improving their vocabulary production when speaking English. Moreover, the scores for Items 2 and 8, which had relatively high means (4.25 and 4.07, respectively), indicated that the use of the hands-on DDL approach was considered to be useful and superior to using a dictionary or online translation programs. However, the mean for Item 7 was low (3.61), indicating that the use of the hands-on DDL approach was more interesting than the use of a dictionary or an online translation program. Despite the effectiveness and usefulness of the hands-on DDL approach, it appears that the EFL learners found it difficult to use, as the mean for Item 6 was relatively high (4.16), leading to low recommendations and a need for more concordance activities in class, as the means for Items 9 and 10 illustrated (3.54 and 3.91, respectively).

With regard to the hands-off DDL approach, although the means for Items 4 and 5 were high (4.27 and 4.15, respectively), similar to the case of the hands-on DDL approach, Item 6 had the highest mean (4.38), suggesting that the EFL learners found the hands-off DDL approach difficult to apply. This result may also explain why, although the EFL learners found the hands-off DDL approach to be helpful for improving their vocabulary learning and production, as indicated by the means for Items 2 and 3 (3.97 and 3.95, respectively), the means for the preference and recommendation for using the hands-off DDL approach were relatively low, as Items 8, 10, and 9 demonstrated.

Despite the relatively positive attitudes toward both hands-on and hands-off DDL approaches, based on the means for the questionnaire items mentioned above, Item 1 had the lowest means in both groups, with 3.26 and 3.12, respectively. This result suggests that the EFL learners may not always have had favorable attitudes toward either type of DDL. This attitude was also reflected by the responses in the short interviews with the EFL learners, which will be described in the discussion section below.

5. Discussion

The first finding that emerged from the analyses mentioned above was that both the hands-on and hands-off DDL approaches are useful and effective for

improving EFL learners' vocabulary production when speaking in English. This finding is significant for the fields of DDL and English speech because the empirical experiment conducted in this research ascertained that the use of DDL can improve vocabulary learning and further expands the DDL approach in the field of EFL learners' spoken English. As speaking in English has been argued to be a comparatively difficult skill for EFL learners to acquire (Abdelmageed & Omer, 2020; Hui & Yunus, 2023; Murad et al., 2023), DDL can mitigate the difficulty of using vocabulary in spoken English. This finding is not only supported by the statistics that were presented in the previous section, but also by the responses to the short interviews with the EFL learners in the two experimental groups. The responses were input into AntConc, which revealed that the keywords "remember" and "memory" were the two most used in the keyword list that reflected a correlation with Items 3, 4, and 8, with raw frequencies of 42 and 29. Two examples of the responses are presented below.

Example 1: *I like this because I feel I can remember many words more clearly, and I can speak what I can remember.* (Participant 05 in group 1)

Example 2: *DDL gives me many contexts and I have many memories when I speak many words.* (Participant 18 in group 2)

Note: The utterances are originally spoken by the EFL learners; thus, certain errors occur.

The questionnaire and the responses to the interviews revealed that the major reason for the effectiveness of both types of DDL was the improved retention of vocabulary that the EFL learners demonstrated during the process (Hirschi & Kang, 2024; O'Keeffe, 2021; Saeedakhtar et al., 2020), thus leading to vocabulary improvements in their oral productions. According to their responses in the interviews, the EFL learners became aware that their vocabulary retention had improved because they could link "remembering words for a longer time to various examples" due to concordancing (Kızıllı, 2023, p. 184). This highlights the significant role played by the examples of vocabulary in context in order for vocabulary to be retained and used in speech production (Lee & Lin, 2019). However, this vocabulary retention due to the effectiveness of the use of both types of DDL differed from the detailed results.

The use of the hands-on DDL approach appeared to be more useful and more effective than the use of the hands-off DDL approach, as seen in three facets. First, the statistical analysis of the three key factors in vocabulary improvement clearly showed that the hands-on DDL approach was the most useful and effective approach for improving the quantity, accuracy, and complexity of vocabulary production, while the hands-off DDL approach only improved the accuracy of vocabulary production compared to the traditional approach. These results are in line with the findings of Saeedakhtar et al. (2020, p. 9), who noted that "both hands-on and hands-off DDL improved the accuracy" of English vocabulary. In this regard, the present research expands on the effectiveness of both types of DDL by identifying the differences in their effectiveness. In addition to the statistical evidence, the questionnaire provided evidence of the

differences. The means for Items 1 to 5 and 8 to 10 in Group 1 were higher than those in Group 2, indicating that the EFL learners had more positive attitudes toward the hands-on DDL approach than toward the hands-off DDL approach. Moreover, the responses to the interviews further supported the findings from the questionnaire. Apart from the two top keywords “remember” and “memory” mentioned above, the keywords “do” and “search” were listed third and fourth, with raw frequencies of 26 and 23. Both of these keywords were linked to the emerging theme of the benefits of the hands-on DDL approach, as the examples below demonstrate.

Example 3: *I think DDL let me do many things by myself, so I can study by myself now, and I know how to use the words in a good way.* (Participant 02 in group 1)

Example 4: *I just feel like I can do my own words, and search examples on [the] computer, so I can learn fast and use the words that I see from the example more.* (Participant 10 in group 1)

The different means for the items in the questionnaire and the examples of the interviewees’ responses show that the hands-on DDL approach encourages learner autonomy because learners must search, observe, analyze, and memorize vocabularies independently via concordancing (Corino & Onesti, 2019; Pérez-Paredes et al., 2019), also known as the inductive learning process (Lin, 2021; Zare & Delavar, 2020). This was the main reason for the greater efficacy of the hands-on DDL approach compared to the hands-off DDL approach. Since the instructor prepared the materials for the hands-off DDL, the EFL learners lacked experience of the searching process; hence, insufficient cognitive and metacognitive processes may have led to the hands-off DDL being less effective than the hands-on DDL approach (Boulton & Cobb, 2017; Mizumoto, 2023).

Finally, the EFL learners’ attitudes toward using the hands-on and hands-off DDL approaches to improving vocabulary production when speaking English are worthy of further discussion. Despite the mainly positive attitudes toward both types of DDL – as revealed by the means for the items in the questionnaire and the responses to the interviews, mentioned above – negative attitudes toward both types of DDL were also found. The means for Item 7 in the questionnaire were relatively low for both groups, at 3.61 and 3.53, respectively, suggesting that the EFL learners may have felt bored when using both types of DDL approaches. It was noted that the mean for Group 1 was higher than that for Group 2, indicating that using the hands-on DDL approach may have caused the EFL learners boredom despite its effectiveness. The responses to the interviews were further enlightening, as the keywords “boring” and “time” were used in the emerging theme of using DDL, with raw frequencies of 22 and 20, as illustrated in the examples below.

Example 5: *It’s just it’s the same, like checking different words on COCA, so, like, in these weeks, I feel less enjoy[ment] and [it’s] a little bit boring over time.* (Participant 05 in group 1)

Example 6: *I think if it’s going to be a long time, this learning is like others; like, [it] will be the same and sometimes a little stressful ‘cos [there’s] too many examples.* (Participant 19 in group 2)

The responses to the interviews with the EFL learners indicated that the EFL learners may have become bored by observing the concordancing over time and experienced some level of stress when the concordancing continued to be presented each week. This finding aligns with several previous studies of the use of DDL among EFL learners from different countries (Hirschi & Kang, 2024; Muftah, 2023; Zare et al., 2022). However, the reasons for the EFL learners' feelings of boredom and stress appear to vary. For example, Zare et al. (2022, p. 7) found that EFL learners experienced stress when using DDL because they "needed more help from the teacher." Some EFL learners mentioned experiencing technical problems when using hands-on DDL, thus resulting in feelings of boredom and stress (Tosun & Sofu, 2023; Zare et al., 2022). This researcher found reasons other than the duration of using DDL and the amount of concordancing causing boredom and stress for EFL learners.

In addition, the learners' feelings of boredom and stress led to the relatively low means for recommending and engaging in more concordancing activities in class, as Items 9 and 10 demonstrated; also, both boredom and stress led to relatively low enjoyment levels, as the mean for Item 1 in the questionnaire was the lowest in the two groups mentioned above. The keyword "suitable" was repeatedly found in this emerging theme, with 17 raw frequencies, as demonstrated by the examples below.

Example 7: *It may be hard [for] everyone to use it, so I'm not sure it's suitable [for] everyone.* (Participant 15 in group 1)

Example 8: *It really needs you to be, like, concentrat[ing] on the examples, so I'm not sure it's suitable always for me or for others.* (Participant 24 in group 2)

Similar concerns about whether DDL is suitable for all EFL learners have been discussed in previous research. Several researchers have contended that the use of DDL may require various considerations, such as the EFL learners' levels of English proficiency, and the previous English teaching and learning approaches applied (Karpenko-Secombe, 2023; Saeedakhtar et al., 2020). The EFL learners in this research indicated that the use of DDL can be challenging and requires high levels of learner autonomy, as illustrated in both examples. Hence, not all EFL learners may be ready for DDL. The discussions above revealed that although the EFL learners generally considered DDL to be effective for improving their vocabulary when speaking in English, they did not consider the use of DDL to be an enjoyable aspect of the learning process (Zare et al., 2022).

6. Conclusion

The results revealed that both hands-on and hands-off DDL approaches significantly improve vocabulary production in EFL learners' spoken English. In addition, the hands-on DDL had a significant positive effect on the quantity, accuracy, and complexity of their vocabulary production. Furthermore, although the EFL learners had relatively positive attitudes toward DDL, less enjoyable experiences were also noted. EFL learners experienced boredom and stress while using DDL, and did not consider DDL to be suitable for all EFL learners. The

findings of this research theoretically indicate that the use of hands-on and hands-off DDL approaches are useful and effective for improving EFL learners' vocabulary production when speaking in English. Since the hands-on DDL approach was more effective compared to the hands-off DDL approach and the traditional approach, it is recommended that the hands-on DDL approach should be used in EFL pedagogy. However, EFL learners need sufficient instructor support when engaging in concordancing activities. Furthermore, the use of DDL may not be appropriate for long periods, and a range of activities should be combined with the DDL approach to engage EFL learners' interest. This research contributes to existing knowledge on the use of DDL to improve EFL learners' speaking skills. Nevertheless, this research only examined the effectiveness of DDL in producing vocabulary improvements among EFL learners. Therefore, it is recommended that future research continues to study the use of DDL within this field, in order to further bolster existing knowledge on the subject.

7. References

- Abdelmageed, N. A. T., & Omer, M. A. A. (2020). The effectiveness of using communicative language teaching approach (CLT) in developing students' speaking skills from teachers' perceptions. *European Journal of English Language Teaching*, 5(3), 88-102. <http://dx.doi.org/10.46827/ejel.v0i0.3044>
- Anthony, L. (2023). *AntConc* (MacOS 10/11, 4.1.3 Version) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software/antconc/>
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393. <https://doi.org/10.1111/lang.12224>
- Corino, E., & Onesti, C. (2019). Data-driven learning: A scaffolding methodology for CLIL and LSP teaching and learning. *Frontiers in Education*, 4(7), <https://doi.org/10.3389/feduc.2019.00007>
- Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment-Companion volume. *Council of Europe Publishing*. www.coe.int/lang-cefr
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. [Data set]. <https://www.english-corpora.org/coca/>
- Hirschi, K., & Kang, A. O. (2024). Data-driven learning for pronunciation: Perception and production of lexical stress and prominence in academic English. *TESOL Quarterly* (in press), 1-27. <https://doi.org/10.1002/tesq.3302>
- Hui, S. M., & Yunus, M. M. (2023). Revisiting communicative language teaching approach in teaching ESL speaking skills. *Journal of Language Teaching and Research*, 14(6), 1515-1523. <https://doi.org/10.17507/jltr.1406.09>
- Karpenko-Seccombe, T. (2023). Data-driven learning: Aiming at the bigger picture. *Nordic Journal of English Studies*, 22(1), 144-181. <https://doi.org/10.35360/njes.798>
- Kızıllı, A. S. (2023). Data-driven learning: English as a foreign language writing and complexity, accuracy and fluency measures. *Journal of Computer Assisted Learning*, 39(7), 1382-1395. <https://doi.org/10.1111/jcal.12807>
- Lay, K. J., & Yavuz, M. A. (2020). Data-driven learning of academic lexical bundles below the C1 level. *Language Learning & Technology*, 24(3), 176-193. <http://hdl.handle.net/10125/44741>
- Lee, P., & Lin, H. (2019). The effect of the inductive and deductive data-driven learning (DDL) on vocabulary acquisition and retention. *System*, 81, 14-25. <https://doi.org/10.1016/j.system.2018.12.011>

- Lin, M. H. (2021). Effects of data-driven learning on college students of different grammar proficiencies: A preliminary empirical assessment in EFL classes. *SAGE Open*, 11(3), 1-15. <https://doi.org/10.1177/21582440211029936>
- Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon*, 9, 1-19. <https://doi.org/10.1016/j.heliyon.2023.e22731>
- Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics*, 3(3), 1-5. <https://doi.org/10.1016/j.acorp.2023.100074>
- Muftah, M. (2023). Data-driven learning (DDL) activities: do they truly promote EFL students' writing skills development? *Education and Information Technologies*, 28(10), 13179-13205. <https://doi.org/10.1007/s10639-023-11620-z>
- Murad, T., Assadi, J., & Badarni, H. (2023). Digital storytelling and EFL speaking skill improvement. *Journal of Language Teaching and Research*, 14(5), 1189-1198. <https://doi.org/10.17507/jltr.1405.06>
- O'Keeffe, A. (2021). Data-driven learning – a call for a broader research gaze. *Language Teaching*, 54(2), 259-272. <https://doi.org/10.1017/S0261444820000245>
- Pérez-Paredes, P., Guillamón, C. O., de Vyver, J. V., Meurice, A., Jiménez, P. A., Conole, G., & Hernández, P. S. (2019). Mobile data-driven language learning: Affordances and learners' perception. *System*, 84, 1-15. <https://doi.org/10.1016/j.system.2019.06.009>
- Pratiwi, W. R., Kuswoyo, H., Puspitasari, M., Juhana, J., & Bachtiar, B. (2024). Driving to communicative approach: the innovative teaching speaking methods in Indonesian English immersion program. *International Journal of Evaluation and Research in Education*, 13(1), 626-634. <http://doi.org/10.11591/ijere.v13i1.25420>
- Saeedakhtar, A., Bagerin, M., & Abdi, R. (2020). The effect of hands-on and hands-off data-driven learning on low-intermediate learners' verb-preposition collocations. *System*, 91, 1-14. <https://doi.org/10.1016/j.system.2020.102268>
- Tosun, S., & Sofu, H. (2023). The effectiveness data-driven vocabulary learning: Hands-on concordancing through a pedagogical corpus. *Journal of Language and Education*, 9(3), 177-191. <https://doi.org/10.17323/jle.2023.12426>
- Ueno, S., & Takeuchi, O. (2023). Effective corpus use in second language learning: A meta-analytic approach. *Applied Corpus Linguistics*, 3(3), 1-11. <https://doi.org/10.1016/j.acorp.2023.100076>
- Zare, J., & Delavar, K. A. (2022). Enhancing English learning materials with data-driven learning: a mixed-methods study of task motivation. *Journal of Multilingual and Multicultural Development* (in press), 1-19. <https://doi.org/10.1080/01434632.2022.2134881>
- Zare, J., Karimpour, S., & Delavar, K. A. (2022). The impact of concordancing on English learners' foreign language anxiety and enjoyment: An application of data-driven learning. *System*, 109, 1-12. <https://doi.org/10.1016/j.system.2022.102891>
- Zhang, J. (2022). Data-driven learning teaching model of college English based on mega data analysis. *Scientific Programming* (Special issue), 1-9. <https://doi.org/10.1155/2022/3490594>
- Zhussupova, R., & Shadiev, R. (2023). Digital storytelling to facilitate academic public speaking skills: case study in culturally diverse multilingual classroom. *Journal of Computers in Education*, 10(3), 499-526. <https://doi.org/10.1007/s40692-023-00259-x>