PASAA Journal
Volume 68, January – June 2024, 229–271

Effects of Data-driven Learning on Writing Complexity, Accuracy, and Fluency (CAF) of Thai EFL Learners

Punyapa Boontam, and Supakorn Phoocharoensil*
Thammasat University, Bangkok, Thailand
**Corresponding author: supakorn.p@litu.tu.ac.th*

Article information	
Abstract	In recent years, there has been growing interest in the use of data-driven learning (DDL) in L2 writing instruction. This paper examined whether and to what extent DDL activities could enhance the writing complexity, accuracy, and fluency (CAF) of 30 Thai EFL learners. The presentation of DDL in this study was hands-on concordancing with the assistance of the Corpus of Contemporary American English (COCA). The study was held in the context of an English writing workshop focusing on opinion paragraphs, divided into two phases: four weeks for the pattern-hunting activities and two weeks for the pattern-refining activities. Applying the CAF Profile of Larsen-Freeman (2006), the pretest and posttest writings were analyzed. The results showed that DDL can enhance learners' writing lexical complexity, accuracy, and fluency, whereas no statistically significant evidence was found in terms of grammatical complexity. Furthermore, the results from the questionnaire and the interview indicated that the learners had positive attitudes towards DDL as a useful reference resource to enhance their vocabulary acquisition, refine their writing, revise their compositions, generate fresh ideas for writing, and boost their

	writing confidence. These findings underscore the potential of DDL as a pedagogical tool in L2 writing instruction.
Keywords	data-driven learning (DDL), corpus linguistics, writing development, CAF measures
APA citation:	Boontam, P., & Phoocharoensil, S. (2024). Effects of data-driven learning on writing complexity, accuracy, and fluency (CAF) of Thai EFL learners. *PASAA, 68,* 229–271.

1. Introduction

Writing skills receive significant attention in university-level education and are crucial for academic success (Staples et al., 2016). In an increasingly global economy, written communication plays a vital role in businesses and organizations. English as a Foreign Language (EFL) learners often encounter challenges in mastering writing skills due to issues like limited grammatical and lexical knowledge (Boonyarattanasoontorn, 2017; Luo & Liao, 2015), difficulties with sentence connectors, and a lack of revising skills (Padgate, 2008). Traditional writing classrooms often rely on dictionaries, but these resources may not provide sufficient guidance for refining and revising writing due to space constraints (Phoocharoensil, 2020). Dictionaries can also fall short in teaching contextualized grammatical knowledge, whereas a corpus-based approach, known as data-driven learning (DDL), originally introduced by Tim Johns (1991), can effectively teach learners how and where to use words in sentences. Therefore, Thai EFL writing classes should consider implementing DDL to help learners overcome their writing difficulties (Luo & Liao, 2015).

Numerous studies have demonstrated the advantages of DDL in L2 writing, including idea generation, creative writing support (e.g. Kennedy & Miceli, 2010), reducing language interference from the learners' native language, fostering positive attitudes toward learning, and enabling self-correction during revision

(e.g. Alsolami & Alharbi, 2020; Birhan et al., 2021; Luo, 2016; Muftah, 2023; Quinn, 2015; Wu, 2021). Moreover, corpus examples effectively help learners with lexico-grammatical patterns, enhancing writing performance and reducing errors (e.g., Coxhead & Byrd, 2007; Flowerdew, 2010). Notably, corpora are also valuable in teaching phraseological patterns such as collocations, colligations, semantic preferences, and prosodies (e.g., Flowerdew, 2010).

Considering CAF measures as a key construct in L2 writing, there also has been an increasing awareness of the importance of complexity, accuracy, and fluency (CAF) in L2 writing among L2 writing scholars during the last decades (Biber, Gray, & Poonpon, 2011; Khushik & Huhta, 2019; Kyle & Crossley, 2018; Lu, 2011) as these measures can be used to find valid and reliable indices of L2 learners' language development. Thus, learners are expected to produce more complex, accurate, and fluent written production. Despite the significant framework of CAF, many previous studies have often focused on syntactic complexity measures without the company of accuracy and fluency as indices of proficiency (Khushik & Huhta, 2019; Kuiken & Veddar, 2019; Lu, 2011). As recommended by Polio (2017), more research should be conducted to examine writing performance in terms of accuracy and fluency.

However, the empirical research on DDL is still limited and narrowly focused, especially in the Thai EFL context. In addition to this, most studies have investigated exclusively the effectiveness of DDL on either learners' vocabulary knowledge or grammatical knowledge in L2 writing, and learners' perception. Consequently, an in-depth investigation aimed at increasing learners' writing complexity, accuracy, and fluency (CAF) through data-driven learning is largely absent from the literature (Yoon, 2011), thus leaving ample room for further empirical research to examine the effects of DDL in this regard. Therefore, to fill in the gap, this study aimed to investigate the effects of data-driven learning on the writing complexity, accuracy, and fluency (CAF) of Thai EFL learners, and

examined their attitudes towards using DDL activities as a reference to aid their writing and a tool to revise their work. In response to the objectives of the study, the following research questions were proposed:

1) What are the effects of data-driven learning on Thai EFL learners' writing complexity, accuracy, and fluency (CAF)?

2) What are Thai EFL learners' attitudes towards using DDL activities as a reference to aid their writing and a tool to revise their work?

2. Literature Review

2.1 The Role of Technology in L2 Writing

Technology supports teachers and students in writing courses by aiding the development of new writing forms and the writing process. Many technological tools and online resources are available to enhance student writing at both language and content levels (Stapleton & Radia, 2009). More significantly, the incorporation of corpus data into L2 writing classes has become valuable, providing teachers and learners with authentic examples of correct word and expression usage (Richards, 2015; Yoon & Hirvela, 2004). Consulting concordance resources allows students to discover and resolve language-related issues in their writing (Lee, 2017). Interestingly, concordancing aligns with assessment for learning (AfL) or assessment as learning (AaL), fostering independence and self-/peer editing skills (Lee, 2017). Properly trained and supported, concordancing serves as a valuable reference tool to enhance writing accuracy and learner autonomy (Yoon, 2011).

Technology's role in L2 writing extends to scoring through Automated Writing Evaluation (AWE), also known as automated essay evaluation or automated essay scoring (Deane, 2013). To elaborate, AWE uses the machine instead of relying on the human rater to generate electronic feedback in scores and/or comments on content, organization, and/or language use (Shermis & Burnstein, 2003). The implementation of AWE has gained popularity because it can

provide timely, valuable, and effective feedback on student writing, particularly in large classes, which can improve student writing outcomes as manifested in reduced error rates (El Ebyary & Windeatt, 2010; Lavolette, Polio, & Kahng, 2015; Liao 2016; Stevenson & Phakiti, 2014). However, Warschauer and Grimes (2008) have argued that AWE may not be reliable in generating feedback on content and rhetorical issues. Despite its limitations, teachers can still employ automated methods to maximize student learning (Shermis, Burstein, & Bursky, 2013).

2.2 Corpus Linguistics in English Language Teaching

Corpora are collections of texts stored in an electronic format allowing researchers to investigate the structure and regularity of authentic data through specific software called concordancers (Szudarski, 2018). According to Lindquist and Levin (2018), concordance can be defined as "a list of all the contexts in which a word occurs in a particular text" (p. 5). In most corpus-based/corpus-driven analyses, the data are often presented in keyword-in-context (KWIC) concordances in which the target word searched, called a node, is centered. Corpus linguistics is beneficial for linguists, researchers, ELT practitioners, and language learners, in terms of authenticity, reliability, and speed (Lindquist & Levin, 2018; Sinclair 1991). More importantly, corpus data can provide researchers and learners with information that is largely unavailable to linguistic intuition as Hunston (2002) posits that corpora are "a more reliable guide to language use than native speaker intuition" (p. 20). In addition, corpus linguistics can be described as the study of the tendencies of authentic data through frequency-based analysis (Szudarski, 2018). To elaborate, relying on frequency as evidence is possibly more reliable than relying solely on native speakers' intuition.

With regard to second language acquisition, it is noteworthy that there is a relationship between corpus linguistics and the noticing hypothesis (Schmidt, 1990) which emphasizes that learners' acquisition of linguistic input can be increased by their explicit attention to linguistic features. For example,

concordance-based tasks can engage learners to attend to recurring phrases in order to enhance their input through noticing (Gao, 2011; O'Keeffe, 2021; Zhu, 2021). Crucially, Schmidt (2010) postulates the concept of noticing the gap as he posits that "in order to overcome errors, learners must make conscious comparisons between their output and target language input." (p. 724). This statement contributes to the use of learner corpora as a corpus-assisted tool to self-correct their work. By the same token, exposure to a myriad of authentic language uses can deepen learners' understanding of particular uses of target words in various contexts and expand their L2 linguistic repertoire (Li & Li, 2022; Xue, 2021; Yoon & Hirvela, 2004).

2.3 Data-Driven Learning (DDL) and L2 Writing

Data-driven learning (DDL), introduced by Tim Johns (1991), is an inductive approach where learners analyze concordance lines to discover language patterns and rules. It involves corpus-designed activities that provide hands-on experience in working with real data (Bennett, 2010). DDL is grounded in authenticity and autonomy, offering learners access to authentic language data and promoting independent discovery (Timmis, 2015; Liu, 2013). Learners can benefit from this discovery-based approach, enhancing vocabulary retention and promoting autonomy (Gilmore, 2015). DDL also fosters metalinguistic and metacognitive awareness (Aston, 2001; Corino & Onesti, 2019). Notably, teachers act as facilitators, guiding learners in the discovery process and boosting their own confidence (Cobb & Bolton, 2015). Various DDL approaches exist, including identifying, classifying, and generalizing (Johns, 1991); illustration, interaction, and induction (Flowerdew, 2009); Chujo and Oghigan's (2012) four-stage approach; and Kennedy and Miceli's pattern hunting and pattern refining (2010).

Corpora's application in L2 writing instruction includes material development by teachers and direct student use in the classroom (Römer, 2008). DDL can be introduced at different writing stages, benefiting students'

understanding of conjunctions, connectors, and reporting verbs (Tseng & Liou, 2006; Bloch, 2009). More importantly, DDL aligns with the noticing hypothesis, enhancing learners' lexico-grammatical awareness and inductive language learning (Coxhead & Byrd, 2007; Fauzanz et al., 2022; Yoon, 2011; Yoon & Hirvela, 2004). Recent studies support DDL's positive effects on writing fluency, accuracy, and error correction (Ahsanuddin et al., 2022; Boone et al., 2023; Samoudi & Modirkhamene, 2020; Luo, 2016; Tono, Satake & Miura, 2014; Eak-in, 2015). Learners become more self-confident and proficient in the learning process (Johns, 1991; Gaskell & Cobb, 2004; Yoon, 2008). Moreover, DDL empowers learners to explore language independently, enhancing their writing skills and promoting lifelong learning (Eak-in, 2015; Gaskell & Cobb, 2004; Yoon, 2008).

2.4 Complexity, Accuracy, and Fluency (CAF) in L2 Writing

Wolfe-Quintero, Inagaki, and Kim (1998) investigated the connection between second language (L2) writing and CAF measures: complexity, accuracy, and fluency. These measures are central to L2 writing research (Biber, Gray, & Poonpon, 2011; Khushik & Huhta, 2019; Kyle & Crossley, 2018), reflecting writing quality. CAF is deemed a valid indicator of L2 performance (Lu, 2011).

Complexity is gauged by assessing learners' language sophistication, including the use of various grammatical forms and structures (Ortega, 2003; Skehan, 2009). This involves analyzing utterance length, subordination, coordination, and grammatical form diversity, indicating linguistic complexity or grammatical complexity (Norris & Ortega, 2009; Housen & Kuiken, 2009). Proficient L2 learners are expected to exhibit more intricate language and a wider array of syntactic patterns (Lu & Ai, 2015; Lu, 2011; Ortega, 2003). Lexical complexity, another aspect, pertains to the range and sophistication of a writer's vocabulary (Wolfe-Quintero et al., 1998). Accuracy refers to error-free language use, indicating control and error avoidance (Foster & Skehan, 1996; Foster & Wigglesworth, 2016; Skehan, 2009). Revision is crucial for error correction,

considering errors may become automated (Schmidt, 1992). Fluency is the ability to produce written content within a timeframe, involving real-time language processing (Ellis & Barkhuizen, 2005; Abdel Latif, 2013; Schmidt, 1992). It also relates to automaticity in language use, facilitated by data-driven learning, which enhances autonomy and grammatical awareness, aiding in fluency (Ellis, 1992). Fluency can also be measured by computing the words per text (W/Tx), T-units per text (T/Tx), and clauses per text (C/Tx) (Wigglesworth & Storch, 2009).

By utilizing T-units, a main clause plus any subordinate clauses (Hunt, 1965), in assessing written language productions, Larsen-Freeman (2006) has defined writing accuracy, fluency, and complexity as demonstrated in the following table:

Table 1

Larsen-Freeman's (2006) Rubric for Measuring CAF

Lexical Complexity	A sophisticated type-token ratio
Grammatical Complexity	Average number of clauses per T-unit
Accuracy	The proportion of error-free T-units to total T-units (in terms of lexical, morphological, and syntactic errors)
Fluency	Average number of words per T-unit

2.5 Previous Related Studies

There is no doubt that corpus consultation has potential for enhancing L2 writing at different stages as proven in the following previous related studies. To gain insights into ESL student attitudes regarding the use of corpora in L2 writing, Yoon and Hirvela (2004) explored corpus-based approaches in two ESL academic writing courses, employing the Collins COBUILD Corpus. Their questionnaire results indicated a positive impact, with learners expressing that corpus use enhanced their writing skills and provided valuable word usage patterns.

Interviews revealed that learners considered corpus use a confidence booster and a tool that fueled their enthusiasm for improving their writing. Yoon (2008) emphasized that corpus consultation aids L2 writers in resolving immediate writing and language issues while boosting their lexico-grammatical awareness and confidence. The study found that corpus tools encouraged learners to pay more attention to word combinations and collocational patterns in both reading and writing. Cotos (2014) investigated the use of learner corpus data to enhance academic writing skills in 31 ESL graduate students in the United States. Results revealed that both participant groups improved their knowledge of linking adverbials, demonstrating the effectiveness of DDL in enhancing L2 writing skills. Likewise, Tono, Satake, and Miura's study (2014) supported Yoon's (2008) findings, highlighting the effectiveness of using corpora in revision tasks to improve the accuracy of Japanese EFL learners' writing. Coupled with coded error feedback, learners corrected their errors using the BNC, which increased their lexico-grammar awareness and attention to specific collocational patterns during revision. In a similar vein, Huang's study (2014) revealed that paper-based DDL had positive effects on the acquisition of lexico-grammatical use of abstract nouns among Chinese EFL learners in L2 writing. Concordance activities helped promote usage-based learning, resulting in more accurate and complex sentences and better retention of acquired syntactic patterns. Most students responded positively to corpus-based instruction, although the format of cut-off concordance lines posed difficulties in understanding contexts. Eak-in (2015) investigated the effects of a corpus-based method on low-proficiency engineering students' abstract writing ability in Thailand. The study revealed that the corpus-based method enabled students to recognize significant linguistic features and apply them correctly in composing an abstract. Students improved in recognizing appropriate terms, collocations, tenses, voices, and patterns used in each move of abstract writing. Worth noting here is Samoudi and Modirkhamene's (2020) study of the application of both native-speaker and local learner corpora, attesting the effect of direct vs. indirect DDL activities on 39 EFL learners' development in CAF

measures of writing. Their findings provided evidence in favor of the effectiveness of DDL activities on L2 writers' CAF as indirect DDL in writing significantly helped the learners produce more accurate and fluent paragraphs. However, no statistical evidence was found as regards syntactic complexity. The study of Muftah (2023) also confirms the effectiveness of employing DDL techniques for revising essays to enhance the fluency and consistency of students' writing, yet there was no empirical evidence presented to demonstrate the utility of DDL in increasing the complexity of students' writing.

3. Methodology

In this study, a one-group pretest-posttest design was adopted. To enhance the feasibility of the study, a pilot study was conducted. The study focused on opinion paragraphs as the writing task. Participants completed both pre-test and post-test writing tasks on different topics. To ensure the effectiveness of DDL, two 2-hour training sessions were conducted before the treatment phase (Smart, 2014). As none of the participants had prior corpus consulting experience, the training was divided into two stages. Following Bernardini's (2004) recommendation, the first stage introduced hands-off DDL using selected concordance lines from COCA, where learners completed practice exercises like gap filling and error correction. In the second stage, hands-on DDL was implemented, allowing participants to independently consult COCA for assigned tasks. They were introduced to COCA's search functions, including part-of-speech tags, frequency, and collocational profiles, and practiced interpreting concordance lines to identify lexico-grammatical patterns. This instructional phase lasted eight weeks (3 hours per week), followed by a post-test writing task, a questionnaire, and an interview.

3.1 Participants

The population in this study was Thai EFL undergraduate students. As evidenced by Vyatkina's (2016) studies, DDL was also found to be effective with

low and intermediate learners. Consequently, the researchers recruited the potential participants based on the Common European Framework of Reference (CEFR) level of B1. As several studies revealed the success of DDL in small classes (e.g., Vyatkina, 2016; Yoon, 2008), to involve participants of the same level, an English placement test, i.e. *EF Standard English Test (EF SET)* (https://www.efset.org/), was used to randomly select 30 Thai EFL undergraduate students of English majors who had an average CEFR level of B1, classified as intermediate proficiency learners. This free adaptive online test evaluated listening and reading skills, aligning with the six proficiency levels of the Common European Framework of Reference (CEFR). EF SET scores were also converted to an internal EF scale from 1 to 100. Employing Computer Adaptive Multi-Stage Testing (ca-MST), the test adjusted its content in real-time based on the test takers' demonstrated comprehension levels, ensuring accurate measurement. EF SET's quality was asserted to be comparable to high-stakes tests like IELTS and TOEFL. Since the participants were selected through a convenience sampling technique, the generalizability of the findings was limited.

3.2 Materials and Instruments

To set the ground for putting the theoretical aspects of the current study into practical practice, the following materials were utilized:

3.2.1 Writing Pre-test and Post-test

The students were asked to write an opinion paragraph on different topics in the writing pre-test and post-test to minimize threats to validity. The topic was selected based on the B1 CEFR description for writing. According to the B1 CEFR description for paragraph writing, students should be able to produce short and simple paragraphs on topics of interest, use simple language to list advantages and disadvantages, and give and justify their opinions. Moreover, they should be able to confidently summarize, report, and give their opinion about accumulated factual information on

familiar routine and non-routine matters within their field. As such, the participants were asked to write a paragraph entitled: “How do movies or television influence people’s behavior?” in the pretest and “Is online learning better than traditional learning?” in the posttest. The task was limited to 40 minutes to evaluate fluency, assessing the ability to write within a set timeframe. Participants were not permitted access to dictionaries or online resources.

3.2.2 Questionnaire

The evaluation questionnaire with 5-point Likert-scaled questions was administered after the writing post-test. It was also designed to examine the participants’ attitudes towards using the corpus tool to develop their writing and elicit their difficulties in learning through DDL. In this study, the majority of items in the questionnaire were adapted from Huang (2014), and all of the questions were translated into Thai to prevent language barriers. The questionnaire consisted of 20 items on a scale of 1-5 (1: strongly disagree, 2: disagree, 3: neither agree nor disagree, 4: agree, and 5: strongly agree). Then, the Likert-type data obtained from the questionnaire were analyzed by computing the mean scores of the participants’ responses. In order to ensure the questionnaire's quality, a comprehensive validation process was implemented. Initially, three experts assessed the questionnaire to evaluate its content validity, followed by a pilot study with a sample of 20 participants who were not part of the main study to assess its reliability and practicality before its actual deployment. Regarding reliability, following the pilot study, the questionnaire’s consistency was assessed using Cronbach’s Alpha. The results indicated an acceptable reliability coefficient of 0.96, signifying a 96% level of consistency and reliability in the obtained scores.

3.3 Treatment (DDL)

3.3.1 DDL Procedures

In this study, DDL was presented through hands-on concordancing (computer-based DDL), encouraging learner autonomy and a student-led approach. The teacher-researcher acted as a facilitator to guide participants in their discoveries. Collaborative learning was emphasized due to the participants' intermediate proficiency and lack of prior DDL experience. The study used the Corpus of Contemporary American English (COCA), a well-balanced corpus representing various language genres, to assist participants. The instructional period spanned eight weeks (3 hours per week) and covered the introduction to opinion paragraph writing, hand-out DDL activities, and corpus-aided error correction in a computer-equipped room.

Table 2

Summary of DDL Activities

Week	Content	CAF Measure	Types of Activities
1	Introduction to opinion paragraph writing		Hand-on DDL through COCA based on corpus-based practice exercises (Pattern hunting)
2	Investigating word choices/lexical collocations	Complexity, Fluency	
3	Investigating grammatical collocations	Complexity, Accuracy	
4	Investigating formulaic language	Fluency	
5	Investigating sentence structures	Complexity, Accuracy	
6	Writing an opinion paragraph (1st draft)		

Week	Content	CAF Measure	Types of Activities
7	Revising drafts	Complexity, Accuracy, Fluency	Revision activities with the assistance of COCA (Pattern refining)
8	Revising drafts	Complexity, Accuracy, Fluency	

In this study, DDL activities for L2 writing were divided into two phases: four weeks for pattern hunting and two weeks for pattern refining (Kennedy & Miceli, 2010). The first week introduced the opinion paragraph structure. During the pattern-hunting phase, learners completed corpus-based worksheets investigating word choices, grammatical collocations, formulaic language, and sentence structures using COCA. These activities were conducted in L2 writing workshops, guiding them in incorporating corpus data into their writing. This enhanced their writing skills and awareness of lexico-grammatical patterns, known to improve fluency (Ellis, 2012; Fauzanz et al., 2022; Martinez & Schmitt, 2012). Collocational patterns and formulaic sequences were introduced in Weeks 2 and 4 (see Appendix). After discovery learning, participants wrote a four-paragraph opinion essay without dictionary or corpus help, using *ETS Criterion*® online writing evaluation in Week 6. The Criterion® Online Writing Evaluation Service is a web-based, instructor-led automated writing tool that helps students plan, write, and revise their essays. It offers immediate feedback, freeing up valuable class time by allowing instructors to concentrate on higher-level writing skills. The final two weeks focused on pattern refining, where students revised their drafts based on *Criterion®* feedback and COCA reference.

3.3.2 Corpus-based Materials

To support participants new to the intervention, DDL worksheets, adapted from Karpenko-Seccombe (2021), were provided. Designed to reduce difficulties and frustration, the worksheets were enriched with lexical chunks based on an Academic Formulas List (AFL) (Simpson-Vlach & Ellis, 2010) and the most common 4-word academic bundles (Hyland & Jiang, 2018) related to discussion and opinion expressions.

3.4 Data Analysis

In this study, both quantitative and qualitative data were analyzed to investigate the effects of the DDL activities on learners' writing development based on complexity, accuracy, and fluency (CAF); explore learners' DDL strategies; and examine their attitudes towards using DDL activities to consult their writing. Regarding the first research question, the scores of the writing pre-test writing and writing post-test were compared to investigate the effectiveness of DDL on the learners' writing CAF. As suggested in the literature, in order to score the two tests, T-units were analyzed in the first place, and then the Profile of Larsen-Freeman (2006) was applied to determine the scores of CAF. In doing so, the learners' performance was scored by two native speakers of English. The statistical procedures including descriptive statistics and paired samples t-test were run to compare their pre-test and post-test scores. In response to the second research question, the results obtained from the Likert-scale questionnaire were analyzed by using descriptive statistics, and the interview responses were analyzed using content analysis.

4. Results/Findings

4.1 The Effects of Data-Driven Learning on Thai EFL Learners' Writing Complexity, Accuracy, and Fluency (CAF)

In response to the first research question, the following three tables demonstrate the results of paired samples T-tests presenting the comparison of

compositions produced by the participants in the writing pre-test and post-test by utilizing the Profile of Larsen-Freeman (2006).

Table 3

Measure of Complexity for Paragraphs in the Pre-test and the Post-test

Complexity	Test	N	M	SD	95% confidence interval of the difference		t	df	Sig.(2-tailed)
					Lower	Upper			
TTR	Pre-Test	30	0.449	0.102	-0.126	-0.052	-4.955	29	.000
	Post-test	30	0.539	0.105					
C/T	Pre-test	30	1.155	0.256	-1.865	0.142	-1.756	29	.090
	Post-test	30	2.017	2.686					

TTR type-token ratio, C/T clauses/T-units

As depicted in Table 3, notable differences were found between the pre-test and post-test regarding writing complexity. To be more specific, the participants demonstrated an enhancement in their lexical complexity. This is evident from the results of a t-test conducted on the pre-test score ($x^- = 0.449$, SD = 0.102) and the post-test score ($x^- = 0.539$, SD = 0.105), which yielded a p-value of .000. This p-value was significantly lower than the 0.05 level of significance, signifying that the post-test score was substantially higher than the pre-test score. Thus, the result indicated the effectiveness of DDL in enhancing lexical complexity and fostering lexical sophistication in the writing of Thai EFL learners as the texts reflected a richer lexicon. However, it is possible that DDL did not yield evident beneficial impacts on the participants' writing development concerning grammatical complexity. This is because no significant difference was observed between the pre-test and the post-test (p = .090), despite an increase

in the mean score in the post-test ($\bar{x}$ =1.155, SD = 0.256) compared to the pre-test ($\bar{x}$ = 2.017. SD = 2.686).

Table 4

Measure of Accuracy for Paragraphs in the Pre-test and the Post-test

Accuracy	Test	N	M	SD	95% confidence interval of the difference		t	df	Sig.(2-tailed)
					Lower	Upper			
EFT/T	Pre-test	30	0.469	.317	-0.393	-0.162	-4.910	29	.000
	Post-test	30	0.746	.212					

EFT/T error-free T-units/T-units

According to Table 4, the results demonstrated a substantial improvement in writing accuracy among the participants in the post-test as the proportion of error-free T-units to T-units increased significantly from 0.469 to 0.746. These findings were supported by a p-value of .000, which was significantly less than the 0.05 level of significance, indicating that the learners made much fewer errors in the post-test, and wrote more accurate sentences after the implementation of DDL. Therefore, the result suggested that DDL served as scaffolding for the participants' writing.

Table 5

Measure of Fluency for Paragraphs in the Pre-test and the Post-test

Fluency	Test	N	M	SD	95% confidence interval of the difference		t	df	Sig.(2-tailed)
					Lower	Upper			
W/T	Pre-test	30	17.458	5.112	-6.782	-3.711	-6.988	29	.000
	Post-test	30	22.704	4.957					

W/T average number of words per T-unit

More importantly, as illustrated in Table 5, a significant difference existed between the pre-test ($\bar{x}$ = 17.458, SD = 5.112) and post-test ($\bar{x}$ = 22.704 and SD = 4.957) in terms of writing fluency which can be seen from the p-value at .000. This suggested that the participants had expanded their vocabulary size, and this was reflected in the increased length of the paragraphs.

4.2 Thai EFL Learners' Attitudes towards Using DDL Activities as a Reference to Aid their Writing and a Tool to Revise their Work

The qualitative analyses of the questionnaire responses were in response to the second research question, which aimed to explore the attitudes of Thai EFL learners regarding the utilization of DDL activities as a resource for improving their writing and as a means to refine their written work.

Table 6.1

Summary of the Questionnaire Responses

Item	Agree (%)	Disagree (%)	Neither agree nor disagree (%)	M	SD
1. I find that analyzing concordance lines is beneficial for me in learning English vocabulary.	100	0	0	4.82	0.39
2. I believe that Data-driven learning (DDL) is beneficial for learning vocabulary collocations.	100	0	0	4.73	0.45
3. I find that learning through concordance lines is useful for studying the lexico-grammatical patterns.	100	0	0	4.73	0.45
4. I think that data-driven learning (DDL) helps me understand how to use vocabulary better.	100	0	0	4.82	0.39
5. I think that the ability and expertise of the teacher are important in helping me understand concordance lines.	100	0	0	4.73	0.45
6. I think that receiving guidance from the teacher during learning helps me understand better.	100	0	0	.91	.29

As can be seen from Table 6.1, a vast majority of participants held a generally positive view of DDL as a valuable tool for enhancing their writing skills. Specifically, all of the participants (100%) expressed that analyzing concordance lines was beneficial for exploring English vocabulary and lexico-grammatical patterns. Interestingly, they also believed that DDL was effective for acquiring vocabulary collocations and improving their overall vocabulary usage. They noted

that DDL facilitated incidental learning of new vocabulary encountered within concordance lines. When it came to the role of teachers, all participants (100%) emphasized the importance of the teacher's abilities and expertise in aiding their comprehension of concordance lines and appreciated receiving guidance from the teacher in this regard.

Table 6.2

Summary of the Questionnaire Responses

Item	Agree (%)	Disagree (%)	Neither agree nor disagree (%)	M	SD
7. I find that data-driven learning (DDL) helps me incidentally learn new vocabulary encountered from concordance lines.	100	0	0	4.73	0.62
8. I find that learning through Data-driven learning (DDL) is beneficial for writing in English on related topics.	100	0	0	4.82	0.39
9. I find that Data-driven learning (DDL) helps me learn new phrases or expressions related to writing.	100	0	0	4.73	0.45
10. I find that Data-driven learning (DDL) helps me generate new ideas for writing.	100	0	0	4.64	0.48
11. Learning through Data-driven learning (DDL) increases my confidence in using vocabulary in my writing.	100	0	0	4.73	0.45
12. Overall, I feel that the corpus is a useful vocabulary-learning resource for writing.	100	0	0	4.91	0.29

Item	Agree (%)	Disagree (%)	Neither agree nor disagree (%)	M	SD
13. Overall, I feel that Data-driven learning (DDL) helps improve the quality of my writing.	100	0	0	4.82	0.39
14. I feel that the corpus is a good tool for improving my writing based on the feedback I receive.	100	0	0	4.82	0.39
15. I can use collocations and phrases from concordance lines in future writing on similar topics.	100	0	0	4.82	0.39

Regarding the utilization of corpora in writing as demonstrated in Table 6.2, all of the participants (100%) agreed that learning through DDL was advantageous for their English writing, particularly on topics relevant to their work. They found that DDL enabled them to acquire new phrases and expressions related to their writing topics. Furthermore, all participants concurred that DDL played a pivotal role in fostering the generation of new ideas for their writing, boosting their confidence in employing vocabulary effectively, and facilitating writing improvements based on feedback received during the revision stage. Significantly, all participants also expressed the belief that they could incorporate collocations and phrases obtained from concordance lines into their future writing on similar subjects.

Table 6.3

Summary of the Questionnaire Responses

Item	Agree (%)	Disagree (%)	Neither agree nor disagree (%)	M	SD
16. I have difficulty learning through concordance lines due to the time and effort required for data analysis.	27	30	43	3.18	1.40
17. I have difficulty learning through concordance lines because I encounter unfamiliar vocabulary in the data.	30	50	20	3.00	1.48
18. I have difficulty learning through concordance lines due to the large number of sentences.	40	60	0	2.73	1.60
19. I find learning through DDL enjoyable and challenging.	100	0	0	4.73	0.45
20. I feel that I can use the corpus to independently search for information to develop my writing.	100	0	0	4.82	0.39

When it came to the challenges encountered in learning through DDL, almost half of the participants (43%) remained neutral in acknowledging any difficulties stemming from the time and effort required for data analysis. Specifically, a portion of the participants (30% and 40%, respectively) admitted to facing challenges when they were confronted with unfamiliar vocabulary within the data and when they had to deal with a large number of sentences. Notably, Table 6.3 underscores that all participants (100%) perceived DDL as an enjoyable and challenging method. They recognized DDL as a tool that could enhance their autonomy, enabling them to independently search for information to enrich their writing skills in the future. In conclusion, the participants displayed positive

attitudes towards DDL, considering it both a valuable reference for enhancing their writing and a useful tool for revising their work.

Apart from the quantitative data obtained from the questionnaire, the qualitative analyses of the responses elicited from the semi-structured interview with ten purposively selected participants were taken into account to gain deeper insights into the learners' attitudes toward learning through DDL. It is worth highlighting that the responses gathered from the interview largely mirrored the findings from the questionnaire that all of them believed that DDL played a pivotal role in their discovery learning when it came to enhancing their writing skills. Thematic analysis of the interviews revealed six critical themes related to the participants' experiences with DDL and its impact on their writing skills as follows:

4.2.1 Vocabulary Enhancement

The participants consistently highlighted the role of DDL in expanding their vocabulary, allowing them to learn new words, replace existing ones, and select suitable collocations. They also credited DDL with enhancing writing quality by promoting proper sentence structure and employing diverse vocabulary. Importantly, they found pleasure in using it as a writing aid, considering it a valuable search engine and reference for future writing assistance, as revealed in the interviews. Therefore, it can be concluded that participants agreed on DDL's ability to improve their writing by acquiring new vocabulary from concordance lines and collocations in the COCA. Subsequent participant responses support this conclusion.

> S1: "I have learned new words that can substitute existing words, which has improved my writing. I can also choose the appropriate collocations."

S2: “I’ve learned various techniques for constructing sentences with a variety of words. I’ve expanded my vocabulary knowledge, which has helped me write in English better.”

S8: “I can bring what I’ve learned to enhance my writing. Properly structuring sentences and using a variety of vocabulary has improved the quality of my writing. I can use DDL as a future learning guide and research tool.”

4.2.2 Convenience and accessibility

Several participants appreciated the ease of accessing information through DDL, particularly during the revision stage. Convenience is also emphasized, enabling participants to thoroughly review and cross-reference their work with corpora, as drawn from the following insights shared in the interviews:

S3: “I like the ease of accessing information to enhance my own writing. It’s convenient and aids in refining my writing, particularly during the revision stage, as I can thoroughly review and cross-reference my work with corpora.”

4.2.3 Fun and challenge

Linking back to the interviews, the participants found using DDL enjoyable and stimulating, with a positive influence on their future writing. They appreciated the process’s enjoyment and the challenge presented by the abundance of sentence examples. Some viewed DDL as a method that could push their boundaries and potential. Despite overall positive feedback, a few participants faced difficulties, particularly with unfamiliar words and incomplete sentences.

S1: “Another thing I like is that DDL makes learning more about me. It focuses on what I need to improve and helps me work on those

specific things. This way, I'm not just doing general exercises; I'm getting better at the things I have been struggling with."

S5: "It's fun and challenging, and I've learned useful vocabulary that can be applied to my future writing. There are many sentence examples (concordance lines) that help me understand the usage of the words better."

4.2.4 Confidence building

It is apparent from the interviews that confidence in writing was highlighted as a benefit, particularly through the ability to search for new and unfamiliar words. Some participants found that discovering language patterns was seen as sparking creativity in writing.

S6: "Using COCA can increase my confidence in writing because I can search for new and unfamiliar words. The discovery of new language patterns sparks creativity in my writing. This is because I can access words that accurately mirror real-life usage within its context."

Therefore, confidence-building through tools like COCA and the application of learning in writing highlight the practical benefits of DDL in real-world language use.

4.2.5 Exposure to diverse examples

The participants observed significant improvements in their writing resulting from exposure to diverse examples from corpora. They emphasized the value of searching for node words in corpora, which refined and enhanced their writing skills, making it more sophisticated, precise, and expressive. Additionally, DDL was viewed as a tool that expanded horizons by aiding in the search for essential information and additional knowledge

beyond dictionaries. Some participants mentioned discovering new language patterns and expanding their understanding.

> S4: “DDL has broadened my horizon. It assisted me in searching for necessary information and additional knowledge beyond dictionaries.”

> S7: “My writing has improved significantly because I’ve encountered diverse examples from various sources simply by searching for a single word. This exposure enables me to refine and enhance my writing skills.”

> S10: “DDL has helped me edit and improve my writing from basic to advanced. It has elevated my writing, making it more sophisticated, precise, and expressive.”

4.2.6 Writing coach

As gleaned from the participants’ experiences, most participants utilized DDL as a language correctness checker during keyword searches. Interestingly, some students regarded corpora as their individual writing coach, turning to them for assistance with their writing challenges. Investigating keywords was described as challenging, providing an opportunity for self-assessment of English capabilities.

> S9: “Using DDL to search for keywords has assisted me in providing more examples and using it as a language correctness checker, like having a study partner who supports me. During the investigation, I felt I could test my English capabilities, which is very challenging.”

> S4: “It’s like a personal teacher that helps me work on the things I find difficult in writing.”

S7: “I like DDL because it feels like having a writing coach. It gives me advice on how to get better and encourages me to keep practicing my English writing skills.”

5. Discussion

5.1 The Effects of Data-Driven Learning on Thai EFL Learners’ Writing Complexity, Accuracy, and Fluency (CAF)

The first research question explored the effectiveness of data-driven learning on Thai EFL learners’ writing complexity, accuracy, and fluency (CAF). It is worth discussing these interesting results revealed by the comparison of compositions (Larsen-Freeman, 2006) produced by the participants in the writing pre-test and post-test. One of the pivotal aspects that can reveal a learner’s writing competence is complexity. The findings indicated that the use of DDL had a notable impact on improving the complexity of vocabulary and promoting more sophisticated vocabulary usage in the writing of Thai EFL students (type-token ratio) as their post-test scores showed a significant and substantial increase compared to their pre-test scores. Notably, corpus-based discovery appeared to encourage learners to be more attentive to word combinations and patterns of collocation during both pattern-hunting and pattern-refining phrases. The findings aligned with those studies indicating that DDL primarily led to improvements in vocabulary and increased lexical complexity, demonstrating that students possessing a more extensive vocabulary have the flexibility to choose words more diversely (Cotos, 2014; Coxhead, 2000; Gardner & Davies, 2014; Huang, 2014; Tono, Satake, & Miura, 2014; Wolfe-Quintero et al., 1998; Yoon, 2008). One additional factor that may have played a role in the positive outcomes of DDL in enhancing lexical complexity in this study was the level of familiarity with the topic as it was suggested in the literature that learners tend to produce more complex sentences when they are well-acquainted with the subject matter (Tedick, 1990). In this study, the participants were frequently exposed to technology-related topics, aligning with the B1 CEFR writing proficiency description, through the DDL

worksheet and writing tasks to ensure that all participants were familiar with this subject matter. Therefore, it can be concluded that the learners were able to identify and address language-related problems in their writing by referring to concordance resources (Lee, 2017), offering support for implementing frequency information obtained from corpora of the English language and academic word lists (Coxhead, 2000; Gardner & Davies, 2014) to enhance the lexical sophistication of more proficient L2 writers. Nonetheless, it is worth noting that the only exception was in terms of grammatical complexity (clauses/T-units), where no statistically significant difference was found between the pre-test and post-test scores ($p = .090$). This suggested that DDL did not appear to have a clear positive effect on the participants' development in terms of grammatical complexity in their writing, in comparison with previous studies where DDL yielded limited benefits (Luo, 2016; Muftah, 2023; Samoudi & Modirkhamene, 2020). One of the important factors that may have impeded the learners' improvement in grammatical complexity could be the insufficiency of time and corpus consultation skills. As suggested by Ortega (2003) and Storch (2009), achieving a higher level of syntactic complexity would likely demand a longer timeframe, possibly extending up to a period of 12 months. Hence, dedicating more time to practice and to maximize exposure to corpus consultation could potentially enhance the effectiveness of DDL in improving grammatical complexity in writing.

In terms of writing accuracy, the findings of this current study were consistent with those studies in demonstrating that DDL can increase learners' lexico-grammatical awareness, resulting in learners' ability to produce more accurate sentences as there was a significant difference in the participants' post-test scores (Eak-in, 2015; Huang, 2014; Tono, Satake & Miura, 2014; Samoudi & Modirkhamene, 2020; Yoon, 2008). Thus, DDL can be considered a valuable reference tool to enhance writing accuracy, confirming the effectiveness of concordancing as assessment for learning (AfL) or assessment as learning (AaL) (Yoon, 2011). More notably, this study seemed to be in line with prior research on

the application of DDL to revise learners' writing and correct errors in their writing, especially after receiving feedback (Coxhead & Byrd, 2007; Yoon, 2011; Yoon & Hirvela, 2004). As indicated by Schmidt (1992), the incorporation of the revision process in the writing activity is vital for learners to correct errors in their writing. During the revision stage in the pattern-refining phrases in this study, the learners had the opportunity to use a corpus-assisted tool to self-correct their work after receiving feedback, and this stage could contribute to the success of DDL in enhancing writing accuracy. The result, thus, provided evidence in favor of the noticing hypothesis (Flowerdew, 2008; Gao, 2011; Schmidt, 1990) as the investigation of abundant recurring phrases through KWIC concordance lines in the DDL activities can encourage learners to notice, analyze, and uncover accurate language patterns and help them reduce errors in writing, as reported by Cotos (2014), Luo and Liao (2015), and Tono, Satake, and Miura (2014).

The last criterion used to measure learners' writing performance in this study was fluency. The significant change in the participants' writing post-test led to the conclusion that the use of DDL could scaffold their writing development, resulting in the expanded vocabulary size and increased length of the paragraphs in the same amount of time. As discussed in the relevant literature (Crossley, Salsbury & McNamara, 2015; Schmitt, 2013), formulaic language plays a crucial role in enhancing the overall quality of language production in L2 learners, as it can improve fluency which is similar to that of native speakers and facilitate effective communication. Therefore, the findings of this current study illuminated the effectiveness of DDL materials in enhancing learners' understanding of phraseology and increasing learners' sensitivity to collocations as the learners could apply the newfound knowledge of word sequences in their writing, leading to a broader vocabulary range and more appropriate collocates, in line with previous studies (Ackerley, 2017; Geluso & Yamaguchi, 2014; Huang, 2014; Luo & Liao, 2015; Luo, 2016; Muftah, 2023; Samoudi & Modirkhamene, 2020; Tono, Satake, & Miura, 2014; Vyatkina, 2016; Wu, 2021). Accordingly, the incidental

acquisition of formulaic language through the DDL discovery can offer valuable insights into frequent patterns of authentic language (Bennett, 2010; Timmis, 2015; Liu, 2013) to EFL learners, resulting in significantly longer paragraphs written by the participants within the same timeframe, as the level of exposure to authentic language can enhance the strength of lexical priming and deepen learners' understanding of particular uses of target words in various contexts (Hoey, 2005; Yoon & Hirvela, 2004).

5.2 Thai EFL Learners' Attitudes towards Using DDL Activities as a Reference to Aid their Writing and a Tool to Revise their Work

The second research question explored Thai EFL learners' attitudes towards using DDL activities as resources to assist their writing and as a means to revise their work, as evidenced by the analysis of questionnaire and interview responses. The results of this study were consistent with previous studies in this field, demonstrating the learners' favorable outlook toward the incorporation of DDL activities for enhancing their second language (L2) writing skills. More precisely, all of the participants agreed that DDL was enjoyable, challenging, and useful for exploring English vocabulary and lexico-grammatical patterns and enhancing their writing skills, reinforcing the conclusions of Huang (2014), Luo (2016), Wasanam and Wasuntarasophit (2016), and Wu (2021). More importantly, this result tied well with previous studies wherein DDL activities can help learners generate new ideas for their writing and boost their confidence (Yoon & Hirvela, 2004; Yoon, 2008). Regarding learners' autonomy, it is evident that all of the participants agreed to incorporate collocations and phrases obtained from concordance lines into their future writing on similar subjects. They also indicated their readiness to autonomously seek out information through corpus-assisted tools to enhance their writing skills in the future (Gilmore, 2015; Sripicharn, 2002; Vyatkina, 2016; Yoon, 2011). The interview findings confirmed the questionnaire results and provided further insights into the learners' attitudes regarding the implementation of DDL activities. It is worth noting that the discovery of collocational patterns through the

COCA was perceived as one of the most beneficial corpus-assisted tools for the learners to improve their writing and revise their errors during the revision stage, supporting the results showing improved post-test scores in lexical complexity, accuracy, and fluency.

Nevertheless, the questionnaire results, aligned with those obtained from the interview responses, reflected some difficulties encountered during DDL learning which were cut-off sentences and unfamiliar words (Geluso & Yamaguchi, 2014; Huang, 2014; Luo & Liao, 2015). Therefore, in order to increase the successful use of DDL activities, it is essential to arrange DDL training sessions for both learners and teachers and design well-planned corpus-based activities (Quinn, 2015; Yaemtui & Phoocharoensil, 2019). The results also supported the existing literature (Cobb & Bolton, 2015; Römer, 2009), which suggests that a DDL teacher as a facilitator plays a vital role in assisting and scaffolding learners throughout the DDL discovery as all of the participants in the present study viewed that the teacher's abilities and expertise in aiding their understanding of concordance lines was essential. In conclusion, the participants displayed positive attitudes towards DDL, considering it both a valuable reference for enhancing their writing and a useful tool for revising their work.

6. Limitations and Recommendations for Future Research

As this study primarily highlighted the potential of DDL in enhancing learners' writing to advocate for corpus utilization, it is important to acknowledge the study's limitations, including the restricted sample size and the study design being a one-group pre-test-post-test design. Therefore, it might be difficult to ensure that the results of this study could be generalized to broader contexts. To establish more definitive conclusions regarding the effectiveness of DDL on learners' writing, future research should involve a larger pool of participants and incorporate a control group for comparison with the experimental group. Additionally, for a more comprehensive investigation, it would be beneficial to

consider incorporating additional indicators for measuring complexity, accuracy, and fluency (CAF) in future studies.

7. Conclusion

The present study investigated the effects of DDL on Thai EFL learners' writing development in terms of complexity, accuracy, and fluency (CAF). With the merit of corpus linguistics in terms of authenticity, reliability, and speed (Lindquist & Levin, 2018; Sinclair 1991), the results confirmed the positive impact of incorporating DDL activities in enhancing learners' writing lexical complexity, accuracy, and fluency. However, the data did not provide statistically significant evidence of an increase in the grammatical complexity of the learners' writing, suggesting that enhancing syntactic complexity requires more time and exposure. Furthermore, the results indicated that the learners had positive attitudes towards the use of DDL as a tool to enhance their vocabulary acquisition, refine their writing, revise their compositions, generate fresh ideas for writing, and boost their writing confidence. Notwithstanding its utility, a minority of learners faced challenges when using DDL, primarily stemming from unfamiliar words and cut-off sentences in concordance lines. Consequently, it is essential to consider incorporating additional training, enhancing teacher competence in DDL, and employing scaffolding techniques when conducting DDL activities. In conclusion, this study has illuminated the significance of using DDL as reference resources to enhance the writing skills of EFL learners as it has offered valuable insights into how DDL can effectively guide learners in refining and revising their writing, particularly when encountering limited spaces in conventional dictionaries.

8. About the Authors

Punyapa Boontam is an Assistant Professor at the Division of Foreign Languages, Faculty of Humanities and Social Sciences, Nakhon Sawan Rajabhat University, Nakhon Sawan, Thailand. She is also a Ph.D. Candidate in English Language Teaching at Language Institute of Thammasat University, Bangkok,

Thailand. Her areas of research interest include Corpus Linguistics and Critical Discourse Analysis.

Supakorn Phoocharoensil, PhD is an Associate Professor of English at Language Institute of Thammasat University, Bangkok, Thailand. He is also the Editor-in-Chief of *LEARN Journal: Language Education and Acquisition Research Network*, indexed in SCOPUS and ERIC. His areas of research interest include Second Language Acquisition, Corpus Linguistics, and English Collocations and Formulaic Language.

9. References

Abdel Latif, M. M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics, 34*(1), 99–105.

Ackerley, K. (2017). Effects of corpus-based instruction on phraseology in learner English. *Language Learning & Technology, 21*(3), 195–216.

Ahsanuddin, M., Hanafi, Y., Basthomi, Y., Taufiqurrahman, F., Bukhori, H. A., Samodra, J., & Wijayati, P. H. (2022). Building a corpus-based academic vocabulary list of four languages. *Pegem Journal of Education and Instruction, 12*(1), 159–167. https://doi.org/10.47750/pegegog.12.01.15

Alsolami, T., & Alharbi, A. (2020). Saudi EFL learners' perceptions of the use of corpora in academic writing teaching. *Studies in English Language Teaching, 8*(4), 94–111. https://doi.org/10.22158/selt.v8n4p94

Aston, G. (2001). *Learning with corpora.* Houston: Athelstan.

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers.* University of Michigan Press.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair, (Ed.), *How to use corpora in language teaching* (pp. 15–36). John Benjamins.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing

development? *TESOL Quarterly, 45*(1), 5–35. https://doi.org/10.5054/tq.2011.244483

Birhan, A. T., Teka, M., & Asrade, N. (2021). Effects of using corpus-based instructional mediation on EFL Students' academic writing skills improvement. *Theory and Practice of Second Language Acquisition, 7*(2), 133–153. https://doi.org/10.31261/TAPSLA.9993

Bloch, J. (2009). The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology, 13*(1), 59–78.

Boone, G., De Wilde, V., & Eyckmans, J. (2023). A longitudinal study into learners' productive collocation knowledge in L2 German and factors affecting the learning. *Studies in Second Language Acquisition*, 1–23. https://doi.org/10.1017/S0272263122000377

Boonyarattanasoontorn, P. (2017). An investigation of Thai students' English language writing difficulties and their use of writing strategies. *Journal of Advanced Research in Social Sciences and Humanities, 2*(2), 111–118.

Chujo, K. & Oghigian, K. (2012). DDL for EFL beginners: A report on student gains and views on paper-based concordancing and the role of L1. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 169–182). Masaryk University Press.

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber, and R. Reppen, (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478–497). Cambridge University Press.

Corino, E., & Onesti, C. (2019). Data-driven learning: a scaffolding methodology for CLIL and LSP teaching and learning. *Frontiers in Education, 4*(7), 1–12.

Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL, 26*, 202–224. https://doi.org/10.1017/S0958344014000019

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213–238. https://doi.org/10.2307/3587951

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*, 129–147.

Crossley, S., Salsbury, T., & Mcnamara, D. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics, 36*(5), 570–590, https://doi.org/10.1093/applin/amt056

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18*, 7–24.

Eak-in, S. (2015). Effects of a corpus-based instructional method on students' learning of abstract writing: A case study of an EAP course for engineering students. *Journal of English Studies, 10*, 1–41.

Ellis, R. (1992). Learning to communicate in the classroom: A study of two language learners' requests. *Studies in Second Language Acquisition, 14*, 1–23.

Ellis, R. (2012). *Language teaching research and language pedagogy*. John Wiley and Sons.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language.* Oxford University Press.

Fauzanz, A., Basthomi, Y., & Ivone, F. M. (2022). Effects of using online corpus and online dictionary as data-driven learning on students' grammar mastery. *LEARN Journal: Language Education and Acquisition Research Network, 15*(2), 679–704.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics, 14*(3), 393–417. https://doi: 10.1075/ijcl.14.3.05flo

Flowerdew, L. (2010). Using corpora for writing instruction. In M. McCarthy, and A. O'Keeffe, (Eds.), *The Routledge handbook of corpus linguistics* (pp. 444–457). Routledge.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299–323.

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics, 36*, 98–116.

Gao, Z. M. (2011). Exploring the effects and use of a Chinese-English parallel concordance. *Computer Assisted Language Learning, 24*(3), 255–275. https://doi.org/10.1080/09588221.2010.540469

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*, 305–327. https://doi.org/10.1093/applin/amt015

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*(3), 301–319.

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL, 26*(2), 225–242.

Gilmore, A. (2015). Research into practice: The influence of discourse studies on language descriptions and task design in published ELT materials. *Language Teaching, 48*(4), 506–530.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*, 461–473.

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL, 26*(2), 163–183. https://doi.org/10.1017/S0958344014000020

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. The National Council of Teachers of English.

Hyland, K., & Jiang, K. F. (2018). Academic lexical bundles: how are they changing? *International Journal of Corpus Linguistics, 23*(4), 383–407.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal*, *4*, 1–16.

Karpenko-Seccombe, T. (2021). *Academic writing with corpora: A resource book for data-driven learning.* Routledge.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology, 14*(1), 28–44.

Khushik, G. A., & Huhta, A. (2019). Investigating syntactic complexity in EFL learners' writing across Common European Framework of Reference levels A1, A2, and B1. *Applied Linguistics*, *41*(4), 506–532. https://doi.org/10.1093/applin/amy064

Kuiken, F., & Vedder, I. (2019). Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. *International Journal of Applied Linguistics, 29*(2), 192–210.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590–619.

Lavolette, E., Polio, C., & Kahng, J. (2015). *The accuracy of computer-assisted feedback and students' responses to IT, Language Learning & Technology, 19*(2), 50–68.

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts.* Springer.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387-417.

Li, J., & Li, M. (2022). Assessing L2 writing in the digital age: Opportunities and challenges. *Journal of Second Language Writing, 57*. https://doi.org/10.1016/j.jslw.2022.100913

Liao, H. C. (2016). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal, 70*(3), 308–319.

Lindquist, H., & Levin, M. (2018). *Corpus linguistics and the description of English* (2nd ed). Edinburgh University Press.

Liu, D. (2013). Using corpora to help teach difficult-to-distinguish English words. *English Teaching, 68*(3), 27–50.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing, 29*, 16–27.

Luo, Q. (2016). The effects of data-driven learning activities on EFL Learners' writing development. *SpringerPlus, 1255*(5), 1–13.

Luo, Q., & Liao, Y. (2015). Using corpora for error correction in EFL learners' writing. *Journal of Language Teaching and Research, 6*(6), 1333–1342.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299–320.

Muftah, M. (2023). Data-driven learning (DDL) activities: do they truly promote EFL students' writing skills development? *Education and Information Technologies, 28*. 13179–13205.

Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*, 555–578.

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.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492–518. https://doi.org/10.1093/applin/24.4.492

Padgate, W. (2008). Beliefs and opinions about English writing of students at a Thai university. *PASAA, 42*, 31–53.

Phoocharoensil, S. (2020). Collocational patterns of the near-synonyms *error, fault*, and *mistake*. *The International Journal of Communication and Linguistic Studies, 19*(1), 1–17. https://doi.org/10.18848/2327-7882/CGP/v19i01/1-17.

Polio, C. (2017). Second language writing development: A research agenda. *Language Teaching, 50*(2), 261–275.

Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal, 69*(2), 165–177.

Richards, J. C. (2015). *Key issues in language teaching.* Cambridge University Press.

Römer, U. (2008). Corpora and language teaching. In A. Lüdeling & M. Kyto (Eds.), *Corpus linguistics: An international handbook* (pp. 112–130). Mouton de Gruyter.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*, 141–163.

Samoudi, N., & Modirkhamene, S. (2020). Concordancing in writing pedagogy and CAF measures of writing. *International Review of Applied Linguistics in Language Teaching,* 1–24. https://doi.org/10.1515/iral-2020-2014

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi: 10.1093/applin/11.2.129

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357–385.

Schmitt, N. (2013). Formulaic language and collocation, In C. A. Chapelle (Ed.), *The Encyclopedia of applied linguistics.* Blackwell Publishing.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL, 26*(2), 184–201.

Sripicharn, P. (2002). Introducing data-driven learning: Theory, practice, and evaluation. *Journal of Liberal Arts, 2*(2), 100–111.

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33*(2), 149–183. https://doi.org/10.1177/0741088316631527.

Stapleton, P., & Radia, P. (2009). Tech-era L2 writing: Towards a new kind of process. *ELT Journal, 64*(2), 175–183. https://doi.org/10.1093/elt/ccp038

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51–65.

Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing, 18*, 103–118.

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research.* Routledge.

Tedick, D. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. *English for Specific Purposes, 9*, 123–143.

Timmis, I. (2015). *Corpus linguistics for ELT*. Routledge.

Tono, Y., Satake, Y., & Miura. A. (2014). The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL, 26*(2), 147–162.

Tseng, Y., & Liou, H. (2006). The effects of online conjunction materials on college EFL students' writing. *System, 34*, 270–283.

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology, 20*(3), 159–179.

Wasanam, T., & Wasuntarasophit, S. (2016). The effects of concordance-based self-correction on the reduction of collocational errors. *Journal of Graduate Study in Humanities and Social Sciences, 5*(2), 130–156.

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing, 26*(3), 445–466. https://doi.org/10.1177/0265532209104670

Wolfe-Quintero, K., Inagaki, S., & Kim, H. -Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity.* University of Hawaii Press.

Wu, Y-j. A. (2021). Discovering collocations via data-driven learning in L2 writing. *Language Learning & Technology, 25*(2), 192–214.

Xue, L. (2021). Using data-driven learning activities to improve lexical awareness in intermediate EFL learners. *Cogent Education, 8*(1), Article 1996867. https://doi.org/10.1080/2331186X.2021.1996867

Yaemtui, W., & Phoocharoensil, S. (2019). Effectiveness of data-driven learning (DDL) on enhancing high-proficiency and low-proficiency Thai EFL undergraduate students' collocational knowledge. *Asian EFL Journal, 23*(3), 290–314.

Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes, 10*, 130–139.

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology, 12*(2), 31–48.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes towards corpus use in L2 writing. *Journal of Second Language Writing, 13*, 257–283.

Zhu, F. (2021). Supporting EFL writing during the pandemic: The effectiveness of Data-Driven Learning in Error correction. E*nglish As a Foreign Language International Journal*, *1*(2), 8–28. https://doi.org/10.56498/66122021

10. Appendix

Example of DDL Supplementary Worksheet

Using Hedges and Boosters

Task 1: Browse the KWIC concordance lines in COCA to find examples of hedging devices and record them in Table 1:

Hedges	Examples
• Modal verbs may, might, can, could, should, would	
• Verbs to seem, to believe, to assume, to suggest	
• Nouns suggestion, possibility, probability, tendency, assumption	
• Adjectives possible, probable, un/likely	
• Adverbs of possibility perhaps, probably, apparently	
• Adverbs of frequency sometimes, often, commonly, occasionally	
• Phrases To the best of my knowledge, it is believed to be one of as far as we know, in some/many cases.	`

Task 2: Adverbs are often used in academic writing to strengthen or weaken a statement.

1) In COCA, put *ly in the search box, and specify the part of speech by choosing adv.ALL for adverb in the POS drop-down menu.
2) Search results show the adverbs in the order of frequency
3) Look at the search results and find which adverbs are used as hedges to show caution and which are used as boosters to show confidence. Add your results to Table 2.
4) Look for the verb collocates that mostly co-occur with each adverb.

Table 2 Hedging and boosting adverbs

Hedges	Verb Collocates	Boosters	Verb Collocates
arguably		undoubtedly	
approximately		obviously	

Using Evaluative Language

Task 3: In COCA, type it is * that as a search term and choose the academic section of the corpus. Investigate the evaluative constructions used with that clause and their frequencies. Then, write four sentences based on your research in Task 3 connected to your paragraph topic.