

EFL STUDENTS' LONG-TERM PRACTICE OF DATA-DRIVEN LEARNING

Natalina Asi¹, Akhmad Fauzan², Jean Seraf Yaspis³, Stepanus Saputra Ferry
Lui⁴, Ibnu Haikal Salasa⁵

^{1,2,3,4,5} Universitas Palangka Raya, Indonesia
a_fauzan@edu.upr.ac.id

Received: September 21, 2021 Revised: October 2, 2021 Accepted: November 5, 2021

ABSTRACT

Data-driven Learning (DDL) has been widely used in language classrooms and suggested to promote learner autonomy and retention. The current study would confirm if the findings of research in other countries are also found in Indonesian context within different time span. Therefore, this study investigates Indonesian EFL students' long-term practice of DDL after one semester has passed since they learned with DDL using corpus data. The participants of this study were university students specializing in English Language Education and the setting of this study was at one public university in Central Kalimantan Province of Indonesia where English is a foreign language. Survey research with one-shot design was conducted and questionnaire comprising of three dimensions (learners' retention of learning materials and current knowledge of DDL, current practice of DDL, and future plan) was distributed to collect data. The results show positive trends that the students had retention of the learning materials they learned in the previous semester through DDL, they currently employed DDL independently outside classrooms, and they would apply DDL in the future.

Keywords: *Data-driven Learning, EFL students, corpora*

INTRODUCTION

Corpus-based pedagogy, also known as Data-driven Learning (DDL), is characterized as a learning approach with an access to corpora or digitalized language data, in which learners become research workers who investigate the corpora and then come up with underlying rules of the language data they observe (Johns, 1991). It has gained popularity for the last three decades with various implementation in different contexts of language classrooms (Boulton & Cobb, 2017; Lee et al., 2019; Pérez-Paredes, 2019). A number of research have been conducted to investigate the use of corpora and DDL in language teaching and learning, such as in grammar (Lin, 2016; Lin & Lee, 2015; Nugraha et al., 2017), vocabulary (Daskalovska, 2015; Karras, 2016; Lee et al., 2019), and writing (Friginal, 2013; Luo, 2016; Tono et al., 2014).

In order to implement DDL in language classrooms, teachers can bring prepared sample of sentences for learners or directly provide them with corpora access to search the data by themselves. The underlying idea of DDL is similar to the concept of inductive method in general that teachers convert the classrooms into

student-centered by asking students to identify patterns, find rules from examples, and finally practice. The inculcation of DDL steps such as open-ended searches, observing and borrowing chunks, and copying models from a corpus to enrich written production were done to learners of Italian language and the results of interview revealed that this approach can have far-reaching value for developing independent language learners generally (Kennedy & Miceli, 2017). When a corpus-based approach to grammar instruction in IELTS writing classes was implemented to Vietnamese students, it was found to promote students' learning autonomy, learning motivation, and critical thinking skills (Pham, 2020). Thus, this approach allows learners to participate actively in the learning process and creates learner autonomy (Gilquin & Granger, 2010; Kennedy & Miceli, 2017; Pham, 2020).

Besides inductive approach, the activity in DDL also grounds in discovery learning. It is an approach to learning in which learners maximize their mental efforts to investigate samples of language use in order to make discoveries about how the target language is typically used (Tomlinson, 2018). The robust effect of discovery learning is that learners remember more often the greater detail of concept and knowledge they generate than materials provided by an instructor (Alfieri et al., 2011). As DDL was implemented to Korean students for L2 vocabulary learning, it is found out that L2 vocabulary proficiency, DDL-focused strategy use, and working memory were significantly associated with L2 vocabulary acquisition and retention (Lee et al., 2020). When Taiwanese teachers applied DDL to their students, they found that DDL successfully aroused their students' interest to learn grammar, increased their in-class participation and learning motivation, and it was believed to have helped the students to learn better and improved their long-term retention of the grammar acquired (Lin & Lee, 2015). Another study compared DDL with traditional grammar learning for the purpose of unlearning over-generated *be* in secondary-level EFL classrooms in Korea and the results pointed that the DDL group showed statistically significant retention as well as immediate effects in grammar learning (Moon & Oh, 2017). In the end, it has been confirmed that DDL should yield greater learning experience, comprehension, and retention (Lee et al., 2020; Lin & Lee, 2015; Moon & Oh, 2017).

Considering that DDL could create direct learner autonomy and enhance immediate retention as suggested by previous studies, the current study investigates learners' memory and independent practice of DDL in the long-run after one semester has passed since they last learned using corpus data. Furthermore, the setting of the current study is in Indonesia which is different from previous studies. Therefore, the current research seeks to answer the following research questions:

1. Do students have retention of past learning materials and steps of DDL?
2. Do students currently apply DDL independently?
3. Do students plan to use DDL in the future?

METHOD

The current study employed survey research with one-shot design and it was conducted at one public university in Central Kalimantan Province of Indonesia. The participants were Indonesian students who major in English Language Education Program and upon graduation they will become prospective English language teachers of primary and secondary schools. During their study at the ELE program, they study language components and skills as well as content subjects. For English grammar, it is taught for three consecutive terms beginning from the first up to the third semester.

When the participants of the current research were in the first semester of their study at the university (n=65), they were introduced to Data-driven Learning (DDL). They experienced the corpus-based pedagogy in English language grammar class by accessing digitalized language data in an online corpus namely Sketch Engine for Language Learning (SkELL). They spent four weeks studying present perfect tense and another four weeks studying passive voice by performing steps of DDL through (a) searching data of sentences, (b) identifying sentences, (c) analyzing sentences, (d) classifying the results of analysis, (e) generating rules, and (f) practicing sentence development. These six steps were done in each meeting with different sentences to analyze. As the students were in the second semester, they had to take another continuation of grammar class after they passed the first one in the first semester. However, they were not assigned to do DDL either inside or outside the lecture meetings. As a result, the current study had the chance to investigate whether or not the students performed DDL steps autonomously or

independently after they were no longer exposed to DDL. The data collection was done at the end of the second semester, or one semester after the students formally learned with DDL.

For the instrument of the current study, a questionnaire was developed containing closed-ended and open-ended questions. This is intended as a way to get definite answers through closed-ended questions and to get students' opinions through open-ended questions. For the closed-ended questions, students had to choose an answer based on the choices provided (Yes and No options). For the open-ended questions, students could write down their opinions as response to the questions. The questions in the questionnaire were grouped into three dimensions, namely students' retention of learning materials and current knowledge of DDL, current practice of DDL, and future plan. For the first dimension, since the respondents of this study had the experience of DDL in the previous semester, the questions checked their memory on the grammar materials they learned with DDL (present perfect and passive sentences) and the steps of DDL. For the second dimension, the questions sought information on the students' current practice of DDL steps that they did independently outside the lecture meetings. For the third dimension, the questions collected students' opinions on the future plan of using DDL steps in other learning materials or other subjects.

The analysis of data resulted from the questionnaire was carried out by calculating the percentage of students' answers on the closed-ended questions. The results in the form of percentages indicate how positive or negative their retention of learning materials and current knowledge of DDL, current practice of DDL, and their future plan of using DDL. Next, students' answers in the form of descriptions on the open-ended questions were used to collect their opinions towards the three dimensions of the questionnaire.

RESULTS

The results of the questionnaire are grouped into three dimensions, namely students' retention of learning materials and current knowledge of DDL, current practice of DDL, and future plan. Table 1 shows the data of students' grammar knowledge when they learned with DDL on the previous semester and their current

knowledge about DDL. Majority of the students stated that they recalled the steps of DDL they experienced in the first semester. They also admitted that they remembered the materials they learned through DDL such as characteristics of present perfect and passive sentences. Moreover, most of the students perceived the benefits of DDL.

Table 1 Retention and Knowledge on DDL

Questions	Answers	
	Yes	No
Do you remember what are the steps of DDL in the first semester?	88%	12%
Do you remember the characteristics of present perfect?	98%	2%
Do you remember the characteristics of passive voice	94%	6%
Do you obtain benefits from DDL?	97%	3%

Table 2 shows the students' current practice of DDL outside the classrooms or when they did not have meeting lectures. According to the students, when they studied independently beyond the classroom hours, they did the steps of DDL such as identification, analysis, generating rules, and sentence development practice. However, only more than half of the respondents mentioned that they accessed the corpus website.

Table 2 Current Practice of DDL

Questions	Answers	
	Yes	No
During the second semester, do you access SkELL?	63%	37%
During the second semester when you study independently outside lecture meetings, do you identify sentence structure (subject, predicate, object, complement) of sentences?	94%	6%

During the second semester when you study independently outside lecture meetings, do you analyze parts of speech of sentences?	78%	22%
During the second semester when you study independently outside lecture meetings, do you generate assumptions of certain grammatical rules?	71%	29%
During the second semester when you study independently outside lecture meetings, do you practice developing sentences?	82%	18%

As seen on Table 3, students show a very positive attitude on using DDL in the future. They believed that they would use the DDL steps in the future and use the corpus website as referencing tool. They also thought that the corpus could be used to support their study in other subjects.

Table 3 Future Plan

Questions	Answers	
	Yes	No
Other than grammar, could corpora be used to master other learning materials of different subjects?	88%	12%
Do you plan to employ DDL steps in the future?	98%	2%
Do you plan to use corpus (SkELL) as reference in the future?	95%	5%

Based on the results of the questionnaire in the tables above, it can be concluded that the participants in the current study had memory on the steps of DDL they experienced in the previous semester as well as the learning materials when they learned with DDL. The number of students who accessed the corpus website and performed the steps of DDL outside the lecture meetings was quite high. The

students also found the corpus website to be useful for other learning materials and planned to use it in the future.

DISCUSSION

In the first dimension of the questionnaire, when the students were asked if they remember the steps of DDL with SkELL website during the first semester, 88% of the students answered 'Yes' and 12% answered 'No'. Those who answered 'Yes' mentioned that the activity of finding and analyzing language data were the stages that they remembered the most; while generating theory and developing sentences were the stages that they least remembered. For those who chose 'No', they admitted that they forgot the stages of learning with DDL because they never applied DDL learning methods again in the second semester and they also did not access the corpus website anymore.

One of the learning materials that students learned by using DDL when they were in the first semester was the present perfect. According to the students, they could remember the learning material although one semester has passed. In the questionnaire, they could correctly mention the purpose of using the present perfect and its characteristics. Furthermore, another learning material that they learned with DDL was passive sentences. The students admitted that they still remembered the rules of passive sentences, although there were a small number of students who forgot the characteristics of passive sentences. It confirms the previous studies that DDL could enhance learners' memory towards learning materials (Lin & Lee, 2015; Moon & Oh, 2017).

Regarding the benefits they obtained from the learning experience with DDL, the students could increase their knowledge of grammar and vocabulary because they used the corpus website as reference. In addition, by getting a variety of new sentences offered by the website, they were able to identify sentences in various patterns so that they could practice distinguishing sentence patterns and eventually developing sentences with word choices that were not monotonous. These findings also support previous studies that learners perceived the positive effects of DDL on their writings (Chang, 2014; Larsen-Walker, 2017; Luo, 2016).

Based on data of the questionnaire's second dimension, when the students were asked about their current practice of DDL, 63% of the students acknowledged that they had accessed the corpus website when they were in the second semester. Their purpose of accessing the website was to find references for words and types of sentences as well as to check grammar. In addition, they also used the corpus website to assist them in doing assignments in writing class, such as in developing sentences and paragraphs. Some students also stated that they used the corpus website in their listening and reading classes, although they did not specifically state their purposes. On another research, learners used corpus data (COCA) to investigate words and phrases discovered in listening materials (Geluso & Yamaguchi, 2014) and to enhance lexical repertoire and thereby improve their reading comprehension (Gordani, 2013; Lee et al., 2020). The students who answered 'No' (37%) stated that they did not use the corpus website because they used other learning resources and the lecturers in the second semester did not instruct them to use the corpus websites.

To find out the students' independent learning in the second semester outside of the classrooms, they were asked if they identified sentence patterns by sorting out subject, predicate, object, and complement of the sentences. The majority of students did sentence structure identification. For them, by identifying the sentence structure, they found it easier to understand grammar, sentence types, sentence patterns, and eventually they could use the grammar correctly. This finding concurs previous research that corpus instruction contributed positively to the patterning of the frequencies and distributional data of linking adverbials, reporting verbs, and verb tenses in the students' research reports (Friginal, 2013).

However, when students were asked whether they identified sentence patterns by sorting out the parts of speech in the sentences, 78% of students answered 'Yes'. According to them, as they were in the second semester they had to learn clauses and phrases so that they had to know certain parts of speech in order to easily recognize sentences and understand the meaning of those sentences. Students who answered 'No' (22%) revealed that they were more likely to look for definitions of certain words and examples of their use. The fact that there were more students who did not analyze the language data by parts of speech might be due to lack of grammar knowledge on parts of speech as previous study states that one has

to specify the parts of speech of the words when working with corpus data and lack of sufficient grammatical knowledge may hinder learners with corpus work (Ebrahimi & Faghieh, 2017).

Based on the students' independent learning experiences outside of classroom hours, they were able to find underlying grammar patterns for certain learning materials such as phrases (noun, verb, adjective, preposition), clauses (noun, adjective, adverb), tenses, and sentence patterns (simple, compound, and complex). On the other hand, there is a number of students (29%) who admitted that they did not generate assumptions of certain grammatical rules because they were still confused in determining the grammar pattern for certain learning materials. It might be due to the inexistence of the teacher to facilitate the students in learning independently. In DDL, teachers should facilitate learning process by preparing learners to manipulate a corpus and helping them draw a conclusion from a corpus data (Boontam & Phoocharoensil, 2018; Chambers, 2010).

When it comes to students' independence in practicing developing sentences, a large number of students admitted that they practiced developing sentences so that in the future they could make more varied sentences and correct mistakes in sentences they develop. In addition, they also stated that developing sentences while learning grammar really helped them to master the learning materials. According to the students, in addition to grammar, corpus websites could be used to support mastery of learning materials in other courses, such as in writing because it includes good and correct grammar knowledge and in reading courses because they can consult parts of speech. The students' opinions are also in line with the findings of other studies which used DDL in writing (Chang, 2014; Larsen-Walker, 2017; Luo, 2016) and in reading (Gordani, 2013; Lee et al., 2020).

In the third dimension of the questionnaire, the students plan to continuously implement DDL learning stages such as identifying, analyzing, and developing sentences because these learning stages are very effective for them in enhancing grammar and vocabulary knowledge through finding errors in sentences and correcting them, and it can strengthen their memory of the learning material being studied. Meanwhile, regarding the students' plans to use corpus website such as SkELL in the future, they stated that these websites could be media that can provide many examples of word usage and examples in sentences so that it can help them

in vocabulary learning by looking at the use of words in context, help them in learning grammar by looking at grammar patterns, and help them in learning writing by looking at sentence patterns.

CONCLUSION

Though previous studies have seen the immediate effects of DDL in terms of learner autonomy or independence and retention, the present study investigated learners' long-term practice when they were no longer obliged to do DDL. The participants of the current research had the experience with DDL when they were in the first semester when the teacher formally taught them grammar using corpus data and employing DDL. When the participants were in the second semester, they were not instructed to access corpus website or implement DDL in the classroom. At the end of the second semester, it was then found out that they still had retention of the first semester grammar learning materials and the steps of DDL. The students also applied DDL steps during their study in the second semester not only in English language grammar class, but also in vocabulary, writing, and reading classes. In the end, the students positively view the potential of DDL for future practice. The current research leaves some gaps to fill for future research, such as a thorough investigation of long-term practice of DDL in vocabulary and writing classes.

REFERENCES

- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. <https://doi.org/10.1037/a0021017>
- Boontam, P., & Phoocharoensil, S. (2018). Effectiveness of English preposition learning through data-driven learning (DDL). *3L: Language, Linguistics, Literature, 24*(3), 125–141. <https://doi.org/10.17576/3L-2018-2403-10>
- 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>
- Chambers, A. (2010). What is data-driven learning? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 345–358). Routledge.
- Chang, J.-Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL, 26*(2), 243–259. <https://doi.org/10.1017/S0958344014000056>

- Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130–144. <https://doi.org/10.1080/09588221.2013.803982>
- Ebrahimi, A., & Faghih, E. (2017). Integrating corpus linguistics into online language teacher education programs. *ReCALL*, 29(1), 120–135. <https://doi.org/10.1017/S0958344016000070>
- Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes*, 32(4), 208–220. <https://doi.org/10.1016/j.esp.2013.06.001>
- Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL*, 26(2), 225–242. <https://doi.org/10.1017/S0958344014000044>
- Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359–370). Routledge.
- Gordani, Y. (2013). The effect of the integration of corpora in reading comprehension classrooms on English as a Foreign Language learners’ vocabulary development. *Computer Assisted Language Learning*, 26(5), 430–445. <https://doi.org/10.1080/09588221.2012.685078>
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal*, 4, 1–16. [https://lexically.net/wordsmith/corpus_linguistics_links/Tim Johns and DDL.pdf](https://lexically.net/wordsmith/corpus_linguistics_links/Tim%20Johns%20and%20DDL.pdf).
- Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL*, 28(2), 166–186. <https://doi.org/10.1017/S0958344015000154>
- Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning*, 30(1–2), 91–114. <https://doi.org/10.1080/09588221.2016.1264427>
- Larsen-Walker, M. (2017). Can Data Driven Learning address L2 writers’ habitual errors with English linking adverbials? *System*, 69(October 2017), 26–37. <https://doi.org/10.1016/j.system.2017.08.005>
- Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5), 721–753. <https://doi.org/10.1093/applin/amy012>
- Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the Establishment of a Data-Driven Learning Model: Role of Learner Factors in Corpus-Based Second Language Vocabulary Learning. *Modern Language Journal*, 104(2), 345–362. <https://doi.org/10.1111/modl.12634>

- Lin, M. H. (2016). Effects of corpus-aided language learning in the EFL grammar classroom: A case study of students' learning attitudes and teachers' perceptions in Taiwan. *TESOL Quarterly*, 50(4), 871–893. <https://doi.org/10.1002/tesq.250>
- Lin, M. H., & Lee, J.-Y. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, 69(3), 264–274. <https://doi.org/10.1093/elt/ccv010>
- Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *SpringerPlus*, 5(1255), 1–13. <https://doi.org/10.1186/s40064-016-2935-5>
- Moon, S., & Oh, S.-Y. (2017). Unlearning overgenerated be through data-driven learning in the secondary EFL classroom. *ReCALL*, 30(1), 48–67. <https://doi.org/10.1017/S0958344017000246>
- Nugraha, S. I., Miftakh, F., & Wachyudi, K. (2017). *Teaching grammar through Data-Driven Learning (DDL) approach*. 82(Conaplin 9), 300–303. <https://doi.org/10.2991/conaplin-16.2017.68>
- Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, 0(0), 1–26. <https://doi.org/10.1080/09588221.2019.1667832>
- Pham, Q. H. P. (2020). A Corpus-Based Approach to Grammar Instruction in IELTS Writing Classes. *RELC Journal*. <https://doi.org/10.1177/0033688220952170>
- Tomlinson, B. (2018). Discovery-based instruction. *The TESOL Encyclopedia of English Language Teaching*, 1–6. <https://doi.org/10.1002/9781118784235.eelt0057>
- 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. <https://doi.org/10.1017/S095834401400007X>

How to Cite (APA style):

Asi, Natalina. (2021, November). EFL Students' Long-Term Practice of Data-Driven Learning. *Exposure: Jurnal Pendidikan Bahasa Inggris*, 10(2), 390-401. <https://journal.unismuh.ac.id/index.php/exposure/article/view/6173>