Data-driven vs. Deductive Learning of English L2 Collocations Using Online Resources: A Convergent Mixed Methods Study

Ahmet ÇEKİÇ[a*]

a* Asst. Prof., Faculty of Education, Sivas Cumhuriyet University (https://orcid.org/0000-0002-7292-3301) * acekic@cumhuriyet.edu.tr

ABSTRACT

This convergent mixed methods study aims to compare effectiveness of using three online consultation resources entailing deductive vs. data-driven learning on learning and retention of 18 verb-noun collocations by EFL learners. The participants (N=45) randomly assigned to treatment conditions consulted three different online resources in different orders to complete the same online error correction tasks and gap-fill exercises in three sessions. The participants were given the Vocabulary Size Test (VST), and a productive collocation translation test (CTT) as the pretest, the posttest and the retention test. A sub-set of the participants also performed think aloud protocols during the treatment. After the treatment, the participants were given the CTT and responded a rating scale and an open-ended question. The results indicated that all of the resources led to significantly higher learning and retention rates with no significant differences among the resources. It was also found that the participants rate the resources differently and go through different cognitive processes when consulting resources. The VST scores positively correlated with the posttest scores, suggesting that participants' vocabulary sizes can be a moderating variable. The findings are discussed based on previous research and within the framework of data-driven learning.

Keywords: L2 collocations, online collocation dictionary, Google search, corpus, data-driven learning, convergent mixed methods design

Çevrimiçi Kaynakları Kullanarak İngilizce Eşdizimliliklerin Veri Güdümlü ve Tümdengelimsel Öğrenimi: Yakınsak Karma Yöntemler Çalışması

ÖZ

Bu yakınsak karma yöntemler çalışma, veri güdümlü ve tümdengelimsel öğrenme içeren üç farklı çevrimiçi İngilizce eşdizimlilik başvuru kaynağının, yabancı dil olarak İngilizce öğrenenlerince 18 adet İngilizce fiil-isim eşdizimlilik içeren etkinlikleri yapmak için kullanılmasının, hedef eşdizimliliklerin öğrenilmesine ve kalıcılık düzeylerine etkisini karşılaştırmayı amaçlar.. Bu amaçla, katılımcılar, aynı çevrimiçi hata düzeltme görevlerini ve boşluk doldurma alıştırmalarını tamamlamak için farklı sırayla üç farklı çevrimiçi kaynağa başvurdu. Uygulamadan önce, katılımcılara Sözcük Dağarcığı Testi (SDT) ve ön test olarak eşdizimlilik çeviri testi (EÇT) verildi. Uygulamadan hemen sonra ise, katılımcılar son test olarak aynı eşdizimlilik testini (EÇT) ve farklı kaynaklara yönelik yorumlarına yönelik açık uçlu bir soru da içeren derecelendirme ölçeğini yanıtladılar. EÇT'nin sonuçları, kullanılan üç kaynağın da, kaynaklar arasında önemli bir farklılık olmaksızın ön teste kıyasla anlamlı seviyede yüksek öğrenme ve kalıcılık oranları sağladığını göstermiştir. Sonuçlar katılımcıların kaynakları farklı derecelendirdiklerini ve farklı bilişsel süreçlerden geçtiklerini göstermiştir. SDT puanları, son test puanları ile pozitif bir korelasyon göstermiştir ve SDT puanının bir düzenleyici değişken olabileceğini düşündürmektedir. Nicel ve nitel bulgular, birbirlerine uyan ve farklılaşan noktalarıyla ve önceki araştırma sonuçlarıyla birlikte, veri güdümlü öğrenme çerçevesinde yorumlanmış ve tartışılmıştır.

Anahtar kelimeler: İngilizce Eşdizimlilik, Çevrimiçi Sözlük, Çevrimiçi Bütünce, Google arama motoru, Veriye Güdümlü Öğrenme, Yakınsak Karma Yöntemler Araştırması.

1 | INTRODUCTION

Although scholars underline the prominence of collocations for fluent, accurate and appropriate L2 production (Daskalovska, 2015; Li, 2017), even advanced L2 learners lack mastery in this respect (Nurmukhamedov, 2016). This is possibly because L2 collocations are not noticed as effortlessly as more idiomatic expressions (Nesselhauf & Tschichold, 2002) and finding an accurate collocation is an elusive and troublesome process for most EFL learners (Lew & Radłowska, 2010). Research suggests that resource consultation improved accuracy and variety of collocations used (Wu, 2015) and that L2 learners can substantially benefit from a variety of online consultation resources such as collocation dictionaries (Chen, 2017; Kim, 2017, 2018), digital corpus resources (Li, 2017; Rezaee et al., 2015), and search engines (Brezina, 2012; Han & Shin, 2017; Sha, 2010).

It has been revealed that using online or electronic collocation dictionaries, which can present increasing amount of data, increase L2 learners' ability to produce collocations (Nurmukhamedov, 2016) and autonomy when learners receive adequate dictionary consultation skill training (Chen, 2017; Kim, 2017, 2018; Laufer, 2011). However, few learners receive systematic dictionary consultation skill training (Kim, 2018). In addition to dictionaries, data-driven resources also become a part of L2 classes with the introduction of web-based corpora (Tsai, 2019). While using data-driven collocation consultation resources such as online corpus and Google search, which entail inductive learning with deeper processing (Craik & Lockhart, 1972) and greater learner involvement (Hulstijn & Laufer, 2001) in the construction (Cobb, 1999) of collocational knowledge and in converting input into intake through noticing of recurrent linguistic patterns (Flowerdew, 2015) represent an inductive approach, traditional exercises via dictionary look-ups represent a deductive approach in producing L2 collocations. Previous research indicates that data-driven learning via corpus resources increases accuracy of collocations (Daskalovska, 2015; Lee et al., 2019; Li, 2017; Mizumoto & Chujo, 2015; Sun & Wang, 2003; Wu, 2015), ensures active engagement of L2 learners and improves their self-editing skills (Han & Shin, 2017), increases linguistic awareness (Kotamjani et al., 2017), and leads to more learner autonomy (Boulton, 2010).

DATA-DRIVEN VS. DEDUCTIVE LEARNING OF L2 COLLOCATIONS

Previous research on DDL (Data Driven Learning) resources (corpus-based resources) also examined their effectiveness in comparison with deductive learning (Başal, 2019; Daskalovska, 2015; Frankenberg-Garcia, 2012; Huang, 2014; Li, 2017; Sun & Wang, 2003; Tsai, 2019) and few studies compared different consultation resources (Kotamjani et al., 2017; Nurmukhamedov, 2016). According to Tsai (2019), research overwhelmingly suggests that while deductive methods and inductive methods are equally conducive to the improvement of receptive word knowledge, inductive methods surpass deductive methods in converting this knowledge into productive knowledge, which is also longer lasting. Sun and Wang (2003), who used four different corpus tools to teach easy and difficult collocations via deductive vs. inductive learning, found that inductive learning was significantly more conducive to the learning of easy collocation, while both methods did not lead to any significant difference in the learning of difficult collocations. Huang (2014), who compared learning of five abstract nouns via using concordance lines and dictionary, showed that the corpus group produced language with higher variation in collocations and colligations and used the target nouns more accurately while also producing more accurate and complex language on average. Kartal and Yangın Ekşi (2018) investigated learning of verb+noun collocations via using corpus vs. completing traditional exercises such as matching, gap-fill using a dictionary, reported significant difference between the groups' collocation production in an essay writing task but not in the collocation knowledge test. Tsai (2019), who compared deductive vs. inductive learning of vocabulary knowledge via using dictionary and corpora in combination in different orders, found that the inductive approach (i.e. when the corpora is used first) was more conducive in collocation recall than the deductive approach but not in collocation recognition.

In another study, Li (2017), who compared the effects of BNC (British National Corpus) and COCA (Corpus of Contemporary American English) based instruction with teacher-led rule based traditional teaching (with access to dictionaries) in a linguistic course by 30 Chinese postgraduates over 15 weeks, found that the corpus group significantly showed better performance in the posttest (essay writing task) in terms of naturalness, the number of academic collocations and fixed phraseological patterns. Daskalovska (2015) found that learning verb+adverb collocations using an online concordancer yielded better results as compared to learning with traditional exercises. Similarly, Başal (2019) investigated collocations learning via traditional exercises such as multiple-choice and matching as compared to learning collocations using online collaborative collocation dictionary entries,

concordance and the Web, found that online resource yielded significantly better learning and retention scores. However, Nurmukhamedov (2016), who compared the Macmillan Collocation Dictionary (MCD, paper-based), the Longman Dictionary of Contemporary English (LDOCE, online) and www.wordandphrase.info (WPI, an online corpus tool) in essay-format error correction tasks, reported that the LDOCE (a deductive learning resource) and WPI (an inductive consultation resource) yielded similar results and both led to better performance than the MCD. In spite of such contradictory findings, previous research comparing the effects of deductive vs. inductive learning of lexical knowledge suggests that while both approaches are favorable to acquisition and retention of receptive vocabulary knowledge, inductive learning is more conducive to transforming receptive knowledge into productive knowledge and retention of such knowledge (Tsai, 2019). Likewise, recent meta-analysis research also indicates the superiority of data-driven inductive learning via using corpus as compared to the deductive approach (Boulton & Cobb, 2017; Lee et al., 2019) with the proficiency of learners being a moderating variable (Lew & Radłowska, 2010; Wu, 2015).

GOOGLE AS CORPUS

Recent research has also revealed that as an extension to data-driven language learning, Google Assisted Language Learning (GALL) (Chinnery, 2008) can transform the Web, which is referred as 'dirty corpus' (Robb, 2003), into an effective inductive consultation resource for L2 collocation production. Sha (2010) pointed out the possibility of using GS (Google Search) as corpus since conventional corpora can attest ineffective in the face of diverse writing needs L2. Han and Shin (2017) found that although teaching Google search techniques (GSTs) had positive effects on article use, it did not lead to improved collocation use, which was attributed to the lack of grammar and lexical knowledge of the participants. Emphasizing the comprehensiveness and potential of GS, Shei (2008) states that "No currently well-known corpus seems large enough to provide adequate instances of prefabricated chunks like this for closer investigation" (p. 67). Sha (2010) argues that web-based corpus can appeal to diverse needs in L2 writing, where traditional corpus often proves ineffective. Overall, research suggests that GS can be an effective collocation consultation resource. However, limited research compared GS with traditional corpus. Brezina (2012), who compared Google Scholar with the COCA academic sub-corpus, pointed out that Google Scholar better represented written academic language and register variation. In an empirical DDL study, Sha (2010), who compared GS with the BNC, found GS to be superior because of its promptness, usability, variety of solutions it offers and most importantly preference investigation, i.e. the possibility to check alternative collocations. Further, in a case study that examined the use of corpus tools including COCA, GS in combination with traditional consultation resources, i.e. online dictionary in by a postgraduate EFL student in academic writing tasks, it was revealed that all corpus resources used contributed to proofreading and editing and increased the learner's linguistic awareness (Kotamjani et al., 2017). Studies also underscore that with adequate training GS can enhance naturalness of learner writing (Acar et al., 2011; Geluso, 2013; Shei, 2008) and improve error correction (Geiller, 2014; Kvashnina & Sumtsova, 2018).

In spite of different affordances of various L2 collocation consultation resources, few studies compared them (Brezina, 2012; Kotamjani et al., 2017; Nurmukhamedov, 2016). In addition, there is no research that investigates how different online inductive consultation resources (i.e. online corpora vs. search engines) and a deductive online consultation resource (i.e. online dictionaries) compare in terms of learning and retention of L2 collocations and how each resource is comparatively rated by L2 learners. Besides, there is little research (Tsai, 2019) on how L2 learners actually construct word knowledge via deductive vs. inductive methods. To fill in these gaps, the current study compares how the employment of three different online resources, viz. SKELL, GS, which entail inductive learning (with different types and amount of information presented, though), and the LCDT, which involves deductive learning, in terms of producing, learning and retention of L2 collocation by completing error correction and gap-fill exercises. In addition, the study investigated how learners actually availed themselves of these resources through screencast recording of search process with each resource and how learners comparatively rate and comment on these consultation resources. The study also examined if the participants' English vocabulary size scores correlated with their overall collocation gains regardless of the digital resource used. Accordingly, the study has four one-sided alternative hypotheses:

1. The resources will not yield significant differences in error correction.

2.	Different consultation resources will lead to significantly different performances in learning and retention of L2 collocations in productive measures.

3.	There will be significant differences in the participants' ratings of the resources.

4.	The participants' vocabulary size will positively correlate with their collocation production, learning and retention test scores.

In addition to these hypotheses, the study seeks to answer the following research question:

1.	What are the comments of EFL learners on the three consultation resources?

2 | METHOD

Among the three mixed methods designs Creswell and Plano Clark (2018) identify (explanatory sequential, exploratory sequential and convergent designs), the current study, in which the two data sets were collected concurrently and separately (i.e. neither depended on the other), suits the convergent design as it is the only design where data is collected concurrently and separately. However, it differs from their typological categorization of convergent design, which *typically* give equal priority to both datasets, in that the current study prioritizes quantitative data in line with its research purposes. In accordance with Creswell and Plano Clark (2018), both data sets were collected from the same individuals (open-ended question in the rating scale) and subset of individuals (only three students who accepted to complete think aloud) as the aim was to corroborate findings and have a deeper understanding. The quantitative data was collected using four measures; namely the VST (Vocabulary Size Test), the collocation translation test, error correction exercise, and rating scale. The qualitative one included think aloud protocols and an open-ended question given at the end of the scale. Data were collected separately and analyzed independently. However, the results were merged in a discussion and interpreted in the discussion section. See Figure 1 for an illustration of data collection, analysis and interpretation procedure.
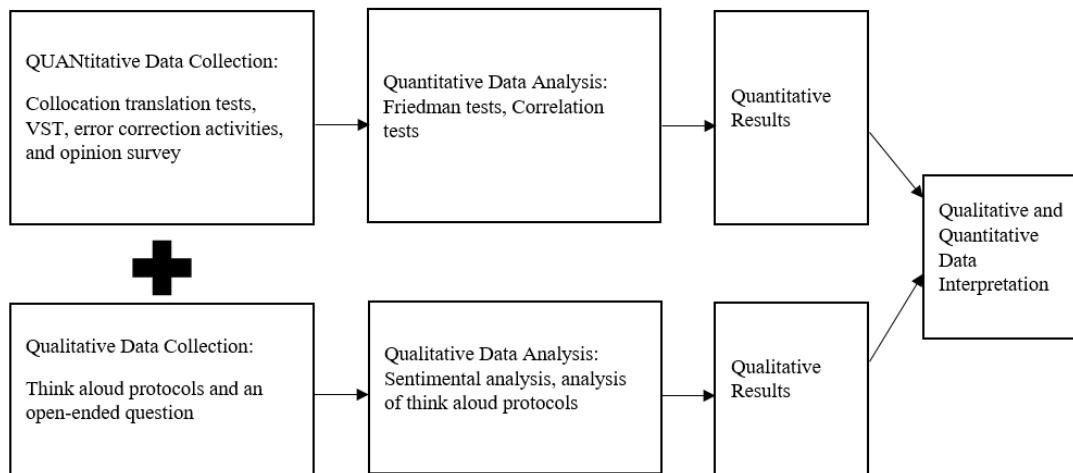

Figure 1. *Research Design*

PARTICIPANTS

The participants (N =48) in the study were from one intact class at an ELT (English Language Teaching) department at state university in Turkey. However, some of the participants missed a session or a test; therefore, data from a total of 45 participants (37 females and 8 males) were included in the analysis. They were all upper intermediate EFL learners and ELT majors and were randomly assigned to the groups.

MATERIALS

In this section, the digital collocation consultation resources, target collocations and instructional materials used in the study are described.

Digital resources. As the resource selection criteria, the researcher made sure that all these resources were free, accessible online, and all the target collocations were available in the resources. The first resource was the Longman Collocation Dictionary and Thesaurus (LCDT). 30-day trial version of the LCDT, which contains over 70.000 collocations including the Academic Collocation List, was used. The resource has a simple search bar and possible collocates for the searched collocate together with L2 definitions and example sentences. Moreover, the dictionary gives different meanings of the node; thus, students can evaluate the meaning of the collocation in the target context. In addition, the consultation resource provides collocation combinations, e.g., verb+ noun or adjective +noun collocations and the part of speech of possible collocates.

The second online resource used in the study is SKELL (Sketch Engine for Language Learning), which is a free online software corpus-based resource. It is based on over one billion words in 57 million sentences from various resources. The consultation resource allows its users to conduct part of speech sensitive search with the node word or the collocate as it displays patterns of collocation in different syntactical combinations with examples of real language use (Figure 2). The users are to deduce the meaning of collocations, distinguish between different senses of collocations to choose the collocation that fits the target context. The third resource was Google search (GS) engine, which can be accessed through the GS search interface with browser. The resource searches the indexes from billions of webpages. The users can choose to do simple or advanced search with specific search operators. The search term is displayed in bold type. Each page displays 10-100 results or more as set by the user. Examining the results of the search, users can find out whether certain collocations exist and deduce the meaning and suitability of a collocation for a given context. However, users of GS have to decide the part of speech information and deduce the meaning of the collocations from the search results.
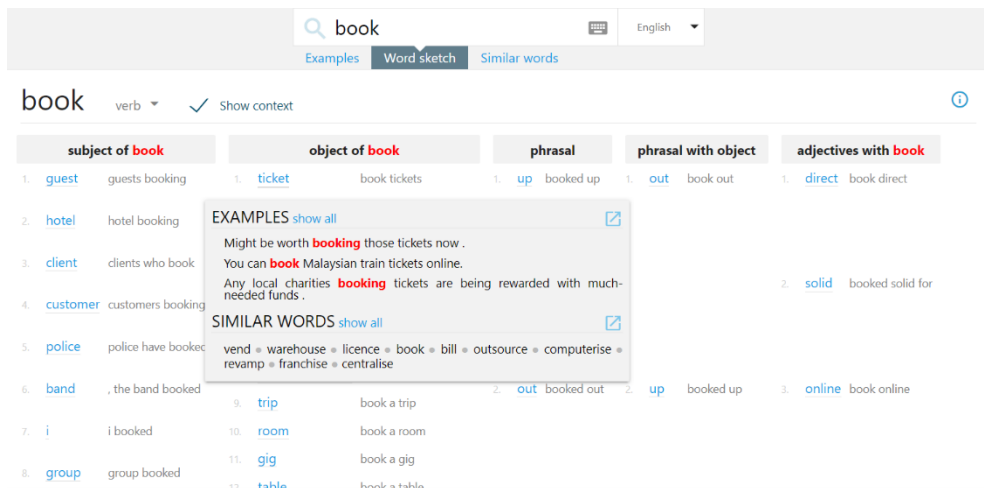

Figure 2. *The result screen from SKELL*

Target collocations and the CTT. The researcher used two main criteria to select the target collocations. First, to make sure that the participants already know the meaning of collocates and node words, the researcher included collocations with words that are within the frequency bands suitable for intermediate EFL learners, who know 3000 to 5000 word families as measured by the vocabulary size test. The collocates and the node words were submitted to Lextutor, a vocabulary profile analysis website (Cobb, n.d.), and all were in the first 3k word families according to the BNC-COCA corpus except for the words quench (10k) and headlines, (off-list). Next, as suggested by Peters (2016), the researcher did not choose the collocations with a higher level of congruency with L1 collocations to eliminate guesswork based on L1 collocation patterns.

Tutorial videos. The researcher and one of his colleagues teaching one of the sections that took part in the study prepared a screencast tutorial teaching for each consultation resource. The tutorials were in the L1 of the participants, i.e. Turkish. The tutorial for GS included the teaching of search operators asterisk, double quotation marks, "OR" and "AND", and minus sign and took longer. The tutorial for SKELL showed how to access the

resource through the website, how to do search and what to look for in the search screen. As for the LCDT, the participants were shown how to register for 30-day trial version and how to do search and evaluate the search results. The participants watched each tutorial before using each resource. In addition, three participants who voluntarily agreed to take part in the think aloud protocol were provided with tutorial videos about the protocol.

Exercises. A set of exercises, namely error correction and gap-fill exercise, were transformed into online exercises using Google Forms. In the error correction exercise, the participants were given 10 verb-noun collocations in sentences and then asked to correct any incorrect use of verb-noun collocations, which were highlighted in bold. The participants were also required to provide correct collocations in the gap-fill exercises. The sentences were chosen from online version of the Longman Dictionary of Contemporary English. The gap-fill exercise was given after the error correction exercises and the participants were not allowed to see the next exercise unless they finished the previous one.

PROCEDURE

First of all, the participants completed the pretreatment measures online. After that, the participants watched a tutorial video about the importance of collocations. They also viewed a tutorial video for each resource and completed practice exercises before using each one of the resources in actual treatment exercises. Later, they studied the same set of collocations each session by completing the same types of exercises in the three treatment sessions over the upcoming three days. The treatment procedure was designed in a way to allow all the participants experience using the three resources by completing the same exercises. Thus, the procedure had a 3 x 3 Latin square design with three groups of EFL leaners (which were subjected to the same treatment in different orders), three sessions (three sets of collocations) and three resources. The target collocations (N=18) were randomly assigned in three sets of six collocations. The order the participants studied the collocation sets, and the exercises were kept constant for each treatment session. However, each group employed different resources in order for each session. In other words, in the three treatment sessions, all the participants studied the same set of collocations in the same order by completing the same exercises using the same resources but in a different order (Table 1). After the exercises in the three sessions were completed, the researcher gave the immediate posttest and the rating scale. Two weeks after the posttest, the researcher also gave the participants the same test as the retention test. Meanwhile, three participants completed think aloud protocol during their search processes in the error correction task with each resource.

Table 1. The Treatment Procedure

	Session 1	Session 2	Session 3
Target collocations	catch fire; allocate resources; satisfy curiosity; follow an advice; achieve success; pose a risk	take a step; address an issue; make headlines; do a damage; quench thirst; make effort	set standards; draw conclusion; break the law; make distinction; follow instructions; make enemy
Group 1	LCDT	SKELL	GS
Group 2	GS	LCDT	SKELL
Group 3	SKELL	GS	LDCT

Note. Each session required studying a set of 6 collocations by completing error-correction and gap-fill exercises.

The researcher preferred a Latin square design rather than using a traditional within-subject design as the latter requires participants to go through the same treatment sessions (using the resources in this study) in the same order, which would have caused variances in terms of time gap between the treatment sessions and the posttest. This would most probably lead to increased rate of attrition for the first set of collocations, which would in turn render comparisons of the effects of resources doubtful. Alternatively, designing a between-subject study would lead to few participants in each group, which is not adequate for quantitative comparisons. Also, a between-subject design would not allow the participants to use different consultation resources and rate them comparatively. The

current design eliminates possible differences between the participants and the collocation sets while also ensuring that all the participants use the three resources. Thus, the scores received using the same consultation resource were brought together to create consultation resource performance scores (rather than those of the participants), so that the effects of digital resources, rather than those of groups, can be compared across the three testing points and in the error correction task.

DATA COLLECTION TOOLS

The data were collected via the collocation translation test (CTT), the vocabulary size test (VST), rating scale and think aloud protocol. The collocation translation test was developed by the researcher. As in active recall vocabulary tests, the meaning of the collocation was given in L1 and the participants were required to provide L2 verb collocate for the given L2 noun (see Appendix A). As a result of reliability analyses, 5 items with a discrimination index lower than .30 were omitted. These omissions amounted to 19 validated items, but to assign them to three groups with equal number of target collocations, one more item with the lowest index value was deleted. However, deleted items were included in the exercises and in the tests. The KR21 score of the remaining 18 items was .71, which means that the test had adequate reliability. Online version of VST (Nation & Beglar, 2007), which was adopted for the web by Cobb (n.d.), was given before the treatment and included the first five 1000 word levels the number of correct answers were multiplied by 100 to calculate the vocabulary size.

The rating scale was developed by the researcher to ascertain the participants' views regarding their experiences of the resources. It included questions about different aspects of the resource such as speed, ease, quality and comprehensiveness of search results, interface, and navigation. The scale was in L1 and it was checked for content and comprehensibility by two experts and necessary changes were made in content and wording. It included 10 questions, nine of which required the participants to rate each resource out of 10. The scale had a good internal reliability score (Cronbach's Alpha=.98), which was calculated after data collection as it was possible to find subjects who had used all the three consultation resources before. The last question in the scale was an open-ended question, which asked the participants to make overall comments on the resources. Their comments were subjected to sentimental analysis. The final data collection tool was screencasts of resource consultation. Three participants agreed to shot screencast videos of their own search processes with each resource while thinking aloud in their L1. They used Microsoft Teams to shoot their own screencast videos. The videos ranged between seven minutes and 30 min. depending on the participants and the resource used.

RESEARCH ETHICS

The researcher received endorsement of the ethics committee. The participants were informed about the stages and the purpose of the study. The participants volunteered to participate in the study were ensured that their data would be kept confidential and their names would be kept anonymous.

DATA ANALYSIS

Quantitative and qualitative data sets were analyzed separately. The CTT scores, which provided main quantitative data set, was scored by rewarding a point for each on the correct answer. The researcher marked the responses leniently, i.e. disregarding minor spelling or word formation mistakes. The pre-test, post-test and retention test scores were grouped according to the resource they used in each session. Global pre-test and post-test scores were also calculated by scoring each correct answer regardless of resource use. Thus, each participant had a GS score, a SKELL score and an LCDT score (each out of 6) and a global score out of 18. While the scores obtained with the same resource were added up and used to compare the resources, the global scores were used to see how they correlate with the participants VST scores.

As the data set from the CTT violated normality assumptions, a Friedman test was used to compare the participants' learning gains via using different resources. Wilcoxon signed ranks tests were conducted as post hoc tests. As for correlation analysis, since the datasets violated the assumption of normality a Spearman's rho test was conducted to see how the VST and pretest, posttest and retention test global scores correlated. Finally, the scale responses were compared using a Friedman test as the data set violated normality of distribution. Besides, the qualitative data from one open-ended question in the scale was submitted to sentimental analysis as there was only one question, which yielded limited qualitative data (Given, 2016, p. 13). Further, the researcher watched the

screencast videos of think aloud protocols during the error correction exercise and sought clues of how the participants availed themselves of different resources to construct collocation knowledge.

3 | FINDINGS

RESULTS FROM THE COLLOCATION TRANSLATION TEST

The results of the Friedman test were significant, $\chi^2(8) = 51.718$, $p < .001$. Dunn-Bonferroni post hoc test was conducted to locate significant differences. Pairwise comparisons (among the pretests, posttests and retention tests) at each testing point among the resources yielded no significant difference. As for the comparison between the pretests and posttests, there were significant differences between the LCDT pretest and LCDT posttest ($p < .001$), between GS pretest and GS posttest ($p = .003$), and between the SKELL pretest and the SKELL posttest ($p = .007$). Comparison between the pretests and retention tests revealed significant differences between GS pretest and GS retention test ($p = .004$), between the LCDT pretest and the LCDT retention test ($p < .001$), and between the SKELL pretest and the SKELL retention test ($p = .012$). Finally, there were no significant differences the posttest and retention tests.

ERROR CORRECTION

The results of the Friedman test indicated no significant differences in the median scores of the LCDT error correction, the SKELL error correction, and GS error correction $\chi^2(2) = 5.700$, $p = .058$.

CORRELATIONS WITH THE VST

The overall scores of the participants were used rather than categorizing the scores according to the resources. Furthermore, Cohen's standard was used to evaluate the strength of the relationships (Cohen, 1988). The results indicated medium significant positive correlations between total pretest scores and the VST scores of the participants ($r(43) = .37$, $p = .012$), and between total posttest scores and the VST ($r(43) = .32$, $p = .033$). There was a non-significant negative correlation between Error Correction rate and the VST scores ($r(43) = -.02$, $p = .922$).

RESULTS OF THE RATING SCALE

Answers of 32 participants were subjected to the rating scale analysis as some of the participants chose more than one rate for some of the questions. The results of the Friedman test indicated significant differences in the scores of the rating scale regarding resources, $\chi^2(2) = 23.870$, $p < .001$. Dunn-Bonferroni post hoc test was conducted to locate the direction of differences. The results indicated that scores of both SKELL and LCDT were significantly higher than the scores of GS ($p < .001$), however, there was no significant difference between the scores of LCDT and SKELL ($p = .617$). The results revealed that the LCDT received the highest scale scores than the other resources for all questions. While SKELL had similar scores as the LCDT, GS consistently lower scores for all questions.
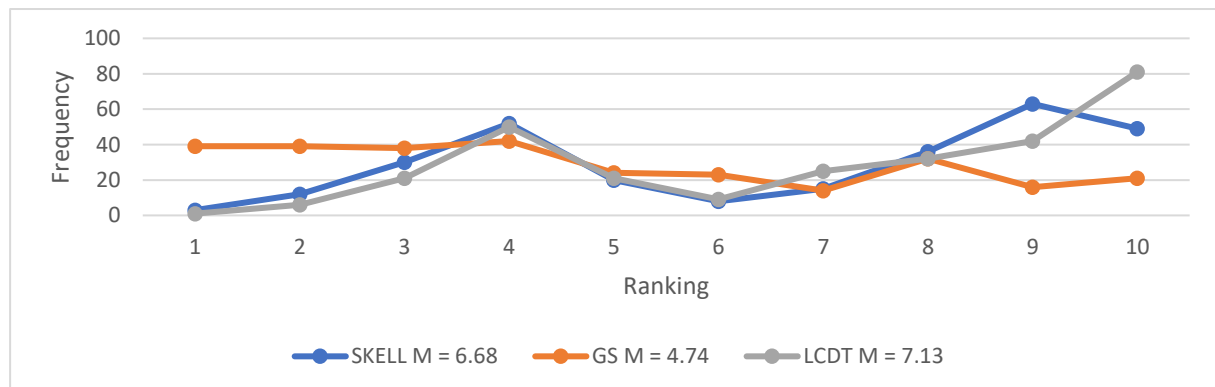

Figure 3. *Distribution of responses for each resource across 10 rating points*

As a total of 32 participants rated each resource for nine questions, so there were a total of 288 (9*32) responses for each resource. The distribution of the responses across ten rating points can be seen in Figure 3. For example, the LCDT was rated the highest rating (10) 81 times out of 288 responses, while SKELL was rated highest score 49 times out of 288 times, GS was rated the highest score 21 times out of 288 responses.

ANALYSIS OF QUALITATIVE DATA

As mentioned in the data analysis section, the qualitative data collected via the final open-ended question were subjected to sentimental analysis. 28 out of 32 participants provided data for the analysis as four of the participants did not answer the last question. Some participants made both positive and negative comments on resources. 21 participants made comments on SKELL, 19 of which were positive and 4 were negative ones. 21 participants made comments on LCDT and all comments were positive. Finally, 15 participants made comments on GS, and there were three positive comments and 12 negative ones. Some comments were inclusive of all of the resources; e.g. P5 stated that "Among the resources, SKELL and LCDT helped me a lot to find the collocation I'm looking for with a variety of examples they provided. But GS does not yield the same performance with unclear samples and inadequate explanations. I think I will use SKELL for my further search for collocations because I can find the collocation I am looking for easily and the search results are clear". Some participants stated their willingness to use SKELL and the LCDT in their future studies especially in writing compositions. For example, P1 stated that "SKELL and the LCDT will definitely be the resources that I will use when writing my essays or for any other (language related) issue." Views on resources generally included comments on the usefulness of the resources. In addition, P28 discussed the format of GS by stating "I had a hard time searching with Google because the font size, having more and different information than the information I was looking for distracted me and caused me to spend more time".

ANALYSIS OF THE THINK ALOUD PROTOCOL

The researcher asked all the participants if they could think aloud and shoot screencast videos of their experiences while completing the error correction exercise. Only three participants agreed to do so. Given to the limited data, the researcher could not conduct any theme analysis; therefore, the think aloud was not transcribed. The researcher watched the screencast recordings and took notes on the participants' search process. It was seen that they spent more time with GS, which was followed by SKELL and the LCDT. P1 spent 7 m. 45 s. with SKELL, 9 m 29 s. the LDCT, and 10 m.11 s. with GS, P2 spent 13 m 11 s., 11 m. 29 s., 32 m. 33 s., and P3 spend 19 m. 07 s., 9 m.18 s. and 19 m. 34 s, respectively. It was also seen that in some cases, the participants just read the sentences and their decisions aloud without giving any details about the underlying cognitive processes of how they constructed knowledge about the target collocations.

As the LCDT displayed all search results in one screen and categorized them according to the part of speech information, the participants could swiftly and easily locate the collocate verb they were looking for by scrolling down. The participants could quickly scan search screen as it was neatly designed thanks to bold characters, different coloring and highlighting (see Figure 4) Most of the time the P1 and P2 read the example sentence first and overlooked the L2 definition in the parenthesis if they could understand the meaning from the context. In some cases, P1 also eliminated other possible collocations by evaluating if the collocate verb's meaning can fit into the context. However, it was seen that in some cases P2 failed to understand the context and thus find an appropriate collocation because of her low proficiency. It was also noticed that she probably misunderstood the instruction and tried to correct accurate cases of collocations even if she had found examples in LCDT that verify collocation accuracy.
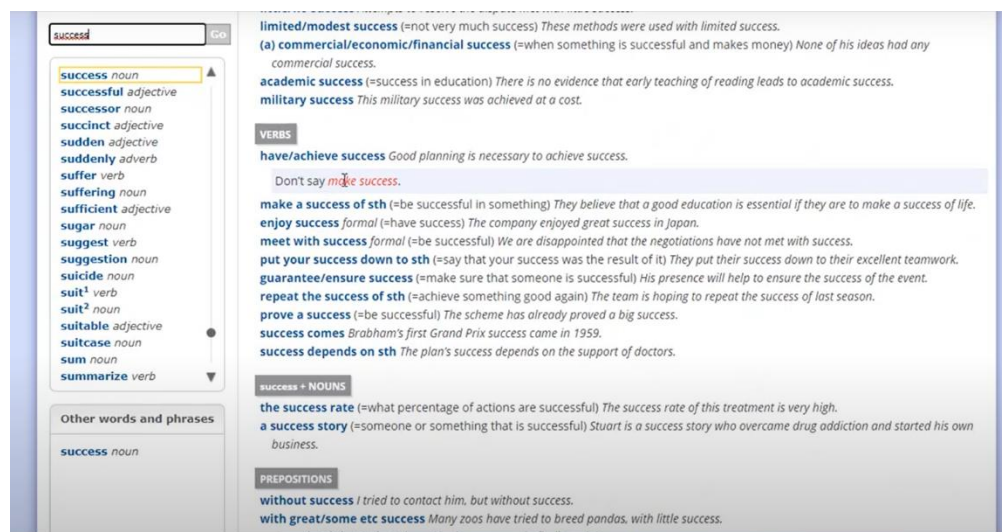
Figure 4. *Search results displayed by LDCT*

With SKELL, the participants quickly directed their attention to the appropriate part of speech and syntactical structure as the resource displayed the results categorized according to syntactical structures. The participants first translated the context and the meaning of the collocation into their L1. Then they checked if the collocation is correct by searching the node. Later on, if they found the collocation in SKELL search screen, they examined the examples if the collocation meets the intended meaning and also confirming its accuracy relying on their L1. They did not translate L2 collocations into L1 literally but used L1 to think about intended meaning and thus accurately translated them. Next, they evaluated if the collocation has the meaning required by the context by examining the examples provided by SKELL. If they think that it does not meet the intended meaning, they evaluated search results by examining examples provided for each collocation. They hovered their cursors on the most possibly correct collocations to see examples and examined the examples. In this respect, the participant's vocabulary knowledge helped them to evaluate the context and search result and thus better avail himself of the resource. On the other hand, even if P2 was able to filter the search results to fit the syntactic structure required by the context, she was not able to decide which collocation provided intended meaning. She was not able to understand the example sentences in some cases probably because of her lower proficiency. Further, she tried to correct every collocation even if the instruction asked them to correct the false ones. Yet, she was able to find the correct collocation in most of the cases as there were limited alternatives with examples. P3 also translated the L2 collocation and search for the collocate verb that can suit the context. He was able to evaluate search results syntactically and semantically. He spent more time with SKELL compared to the other participants because he has slower internet connection. He also evaluated more search results.

While using GS, the participants tend to pay attention to the explicit deductive resources such as dictionaries, which already provided meaning of the target collocation in L1 or L2 and sometimes an example sentence. P2 particularly took very long time to complete the error correction task with GS as she tried to correct collocations in spite of GS provided evidence for their accuracy, possibly because she misunderstood the instruction. In some cases, the participants failed to evaluate if any search result was a true collocation or just a coincidence. Further, they were not able to evaluate if the meaning of the collocation in the GS results meets the intended meaning in the sentence. P1 decided if she found the correct collocation just based on the frequency of collocation in search results. In addition, P1 and P2 translated collocations and searched for their L1 equivalent in GS, which directed them to bilingual dictionaries. In sum, the result of the think aloud protocol indicates that employing GS, a data-driven resource, does not necessarily guarantee that the participants are involved in data-driven learning. It was also seen that differences in proficiency and their level of attention played a determining role in the extent the participants availed themselves of the resources. In general, it was seen that while the participants could use the LDCT and SKELL in line with the tutorial video, they were not so good at using GS. It was also noted that lower collocation awareness and ability to evaluate the context of the sentences semantically caused them to perform

poorly with the resources especially with GS. In sum, it was seen that not only the quality of the resource used but also the individual participant's overall linguistic proficiency (and collocation awareness) determined the success in error correction.

4 | DISCUSSION

In this section, the results will be discussed for each hypothesis or research question one by one. For each hypothesis or research question, first, qualitative and quantitative results related to each research hypothesis will be merged in a discussion and then the interpretation will be elaborated on based on previous research and established theories. As for the first hypothesis, the analysis of the error correction rates (quantitative data) revealed no significant difference in terms of the resource used in the error correction task, so the first hypothesis is rejected. However, the results of the rating scale indicated that the learners rate them differently. The contradiction between these quantitative results can be explained by the analysis of think aloud protocols, which indicated that the participants could find correct collocations even if they had difficulty in using the resources and spent more time with GS. There is limited research that compared the performances yielded by different consultation resources and finding of a similar study (Nurmukhamedov, 2016), which found no significant difference between using an online dictionary (the LDOCE) and with an online corpus-based resource (WPI) in an error correction task, corroborates the current study. This finding can be attributed to the limited set of collocations studied in both studies and to the fact that all of target collocations were accessible via all of the resources used.

The second hypothesis argued that there will be significant learning gains and retention rates among the resources. It was based on the theoretical assumptions that the inductive resources (SKELL, GS), which entail more learner involvement, will yield better performance in collocation learning and/or at least in retention as compared to a deductive resource (the LCDT). However, there was not a significant difference in the posttest and the delayed posttest, so the second hypothesis was rejected. However, the results of the rating scale indicate that the participants rate LCDT and SKELL significantly better than GS, which is also supported by the qualitative analysis of the open-ended question. The contradiction between these quantitative measures can be interpreted to mean that positive ratings of resources does not always translate as superior performance. However, the results of the think aloud protocol revealed that in some cases, the participants used GS like a deductive resource. Thus, as has been argued in the analysis of the think aloud protocol, the employment of a data-driven resource does not necessarily mean that the participants will be involved in the data-driven learning process. This analysis raises the possibility that the lack of difference cannot be interpreted to mean that consulting data-driven resources does not lead to significantly better learning and retention of L2 collocation.

The results of the posttest and the retention tests are in conflict with theoretical underpinnings of DDL and body of empirical research. In line with the theoretical tenets of DDL, the participants using GS and SKELL were expected to inductively study language samples with greater involvement in the process, and thus notice recurrent patterns and construct more and longer lasting lexical, grammatical and semantic information about collocations. However, at first view the finding of the current study seems to contradict with these assumptions based on learning theories. This finding also conflicts with the findings of previous research (Başal, 2019; Boulton & Cobb, 2017; Daskalovska, 2015; Huang, 2014; Kartal & Yangın Ekşi, 2018; Lee et al., 2019; Li, 2017; Tsai, 2019) indicating the superiority of data-driven learning in learning collocations over traditional treatments.

The lack of significant difference in the current study contrary to research and learning theories behind DDL can be attributed to three main differences: measures, different methodological understandings of DDL and the length of the study. Although previous research which measured both collocation recognition and recall separately in comparison of the deductive approach vs. the inductive approach, found that the latter is particularly more conducive to collocation recall whilst being equally effective in collocation recognition (Cobb, 1997; Frankenberg-Garcia, 2012; Huang, 2014; Kartal & Yangın Ekşi, 2018; Tsai, 2019), different perceptions of productive collocation knowledge measures can make comparisons less straightforward.

The productive measures entailed error correction in ten grammatical collocations (prepositional colligations) in Frankenberg-Garcia (2012) and recall of nodes for target verbs (only five) in Tsai (2019). Therefore, the conflict between the two studies and the current study, which also used a productive measure of collocation knowledge (i.e. collocation translation test), can be attributed to different understandings of productive collocation knowledge

and also the number of test items. Some researchers used even more productive measures, which require the participants to apply collocation knowledge in new contexts, i.e. in writing. Kartal and Yangın Ekşi (2018) reported that inductive learning via using corpus as compared to deductive learning via completing traditional exercises with access to dictionary led to significantly more collocation production in writing but not in recognition of a specific set of collocations. Their finding suggests that deductive learning and inductive learning can be equally conducive in terms of the learning a specific set of collocations at receptive level but the inductive method (via. corpus-tools) is superior in the application of collocation knowledge in writing. Furthermore, Huang (2014) revealed that the participants in experimental group, who studied five abstract nouns via using concordance lines as compared to consulting dictionary, noticed more lexical collocations and showed significantly increased accuracy and variety in writing. Similarly, Cobb (1997) found that while vocabulary was learned via inductive and deductive methods, only learners in the inductive condition significantly improved their ability to apply vocabulary knowledge in new contexts. Thus the lack of significant difference in the current study can be explained with the lack of more productive measures which required leaners to use their collocation knowledge in new context, viz. via writing.

Furthermore, methodological differences also make it difficult to compare results. In her rigorous and in-depth study of how learners avail themselves of different resources while constructing different dimensions of vocabulary knowledge, Tsai (2019) allowed the participants in both groups to consult the corpus and the dictionary (in reverse orders), which does not only lack ecological validity since L2 learners would consult either one of the resources based on their perceived practicality and/or usefulness but also casts doubts on her comparison. Their access to both types of resources makes it difficult to attribute the superiority of the inductive group in collocation recall to the nature (quality) of cognitive process underlying the inductive approach or to the different order in the implementation of both methods in combination.

Li (2017), who examined the improvement of productive vocabulary improvement over a term in an essay writing task, found that although both groups improved their writing, the corpus group outperformed the traditional group in collocation production. Thus, previous research indicates that superiority of corpus consultation can be particularly observed in the implementation of collocation knowledge in new contexts (i.e. via writing), which can at least partially explain the contradictory result of the current study, which measured a limited set of collocations in a collocation translation test. The lack of difference between the resources in the current study can also be attributed to the length of the study. The current study was conducted over a period of three weeks, including the training and the delayed posttest. However, Li (2017), who compared corpus-based activities with traditional teacher-led activities with access to dictionaries in a study over 15 weeks, found that the corpus group performed significantly better. This is also supported by Boulton and Cobb (2017), who argued that the effect of DDL could be seen in the long period of study. In sum, the outwardly contradictory finding of the current study can be attributed to different perceptions and measures of productive collocation knowledge, varying methodological understandings of how the inductive vs. deductive divide, and the length of treatment.

The lack of difference between GS and SKELL, both of which entail inductive learning, can be simply interpreted to mean that similar consultation resources yield similar results. However, the finding contradicts with the results of previous studies, which suggested the superiority of GS compared to other corpora such as COCA (Brezina, 2012) and the BNC (Sha, 2010). The discrepancy between the findings can be attributed to the limited number of collocations (which were all accessible in all resources) in the current study as compared to other studies. Furthermore, they (Brezina, 2012; Sha, 2010) did not compare the performances of the participants with difference resources but compared different characteristics of the resources. Brezina (2012) compared Google Scholar and with COCA-academic and found that the former encompasses larger amounts of academic language, better represent the shades of academic language and allows for comparison among different disciplines. He concluded that it is a better tool in corpus-driven English for Academic Purposes (EAP) research. Sha (2010) also found GS is superior than the BNC in terms of usability, speed, comprehensiveness and preference investigations. However, the current study compared GS with other resources in terms of learning and retention of a specific set of collocations. So although GS did not lead to better performance with the limited set of collocations in this study, it can be more effective with larger sets of collocations or when writers want to find answers during their writing tasks (Sha, 2010).

The contradiction of the finding with existing theories and research can be better explained via an analysis of the actual processes the participants had gone through in-depth via think aloud protocol. Think aloud protocol reveal the reason why GS did not lead to significant results, contrary to the expectation based on studies indicating its superiority over corpus resources (Brezina, 2012; Sha, 2010). As already pointed out, it has been revealed that the learners did not necessarily engage into data-drive learning and evaluation process with GS just because the resource displayed large amounts of search results. Think aloud protocol indicates that while the participants effectively used SKELL and the LCDT, they were not so good at using GS. When using GS for some collocations, all of the three participants turned to dictionary results with explicit meaning and examples in the same result screen, which cause GS to lose its advantage as a resource involving DDL. Furthermore, in some cases they failed to evaluate if the search result is a true collocation or a coincidence. They chose collocations based on their frequency without evaluating if its meaning fits the target context. Thus, think aloud protocol revealed that the lack of difference in favor of GS can be attributed to the inadequacy of training via screencast the tutorial video and the participants' low level of awareness about collocations. However, it must also be noted that although it took longer for the participants to correct mistakes with GS compared to SKELL and the LCDT, the participants were able to correct similar number of mistakes using GS. Based on the analysis of the think aloud protocols, the lack significant difference between SKELL, which is a resource that require DDL, and the LCDT, can be attributed to the fact that the participants were able to locate similar number of collocations in both resources and they were able to infer the meanings from SKELL as well.

The third hypothesis was accepted as the results of the rating scale indicated that the participants rated the LCDT and SKELL significantly more favorable than GS in all aspects. This is also corroborated by the analysis of the think aloud protocols, which showed that the participants spent more time to locate collocations in GS and had more challenging experiences. Sentimental analysis of the responds to the open-ended scale question also suggested the same. P14 stated that "Compared to the others, searching with GS was more tedious and it was more challenging to find the collocation that I was looking for." This difficulty can be attributed to inadequate grammar and lexical knowledge of the participants (Han & Shin, 2017) and their need for more rigorous GS training. Furthermore, searching with GS might have been a more tedious task for them because GS, which is not particularly developed for locating L2 collocations, displays huge numbers of search results in a number of pages without lexico-grammatical and semantic information and without filtering coincidental co-occurrences. Thus, it can also be argued that GS was rated by far the least favorable resource because the participants did not have enough collocation awareness to evaluate search results by GS and thus find it challenging to use (also indicated by think aloud protocols). In sum, the results of the rating scale indicate that L2 learners can rate a DDL consultation resources, i.e. SKELL as favorable as a payable deductive consultation resource, i.e. LCDT, unless it is too challenging for them to use (i.e. GS in this study). It is also interesting to note that in spite of the fact that think aloud protocols revealed that the participants had difficulty in using GS and rated it the least favorable resources in the rating scale, their performance with GS was not significantly worse than with the LCDT and SKELL. This finding can be interpreted to mean that although the participants availed themselves of these resources equally (as indicated by posttest, the retention test, error correction measures), they rated them differently. In spite of the fact that the participants consistently performed better (tough not significantly better) with GS than SKELL in error correction, in the posttest and the retention test, they rated GS lower than the LCDT and SKELL in responds to all questions in the scale.

Finally, the fourth hypothesis was accepted as the VST scores positively correlated with the posttest and with the retention test. This finding suggests that learners with larger vocabulary size benefitted more from the resources, which is also supported by the analysis of the think aloud protocol. The participants with lower vocabulary size had difficulty to understand example sentences or evaluate the required context. This finding is compatible with the findings of previous studies, which indicated that level of proficiency can determine how much learners can benefit from consultation resources (Lew & Radłowska, 2010; Wu, 2015). The research question with regard to the participants' evaluation of the resources revealed that the learners made more positive comments on the LCDT and SKELL. This qualitative finding is corroborated by the results of the quantitative dataset from the rating scale, which indicated that the participants rated LCDT and SKELL higher than GS. In sum, although the quantitative measures of the participants' performances with each resources did reveal no significant difference in learning gains and retention, interpretation of the quantitative dataset from the scale and

qualitative data from the open-ended question in combination suggests that L2 users differ in the way they avail themselves of these resources depending on their linguistic proficiency and rate them differently.

5 | CONCLUSION

The findings in the current study indicate that online resources can be used to promote L2 more accurate collocation production and using them can lead to learning and retention of collocations. Their designs do matter at least from a qualitatively perspective. Therefore, resources that are particularly developed for language learning should use more user-friendly interfaces with more precise definitions, example sentences in search results so that learners can more swiftly locate the collocations and better understand their meanings. Furthermore, the example sentences or samples from real language use should be clear cut so that DDL can be facilitated. The study also points out the effectiveness of GS, a resource which is not intended for locating collocations, as a DDL resource for locating and learning collocations. Therefore, EFL practitioners should not disregard its potential as an L2 collocation consultation and learning resource. The results revealed that not only the resource consulted but also the collocation awareness level and overall proficiency level of the learners can determine success in locating and learning collocations. Another implication of the current study is that ELF practitioners should not suffice with introducing these resources but plan instructional activities that will give L2 learners hand-on-experience in effectively using these consultation resources. Furthermore, activities that will raise L2 learners' collocation awareness and lexico-grammatical awareness should be a part of L2 classes so that learners can better avail themselves of such resources.

This study is not free from limitations. First, the number of the participants was small, and several participants could not provide adequate data because of missing some of the sessions. Next, the tutorial on GS was not adequate as revealed by the think aloud protocols. Furthermore, the type (verb+noun) and number of collocations and the experimental period were also limited. In addition, it was not possible to collect think aloud data from more participants as only three participants agreed to so. Finally, the study employed productive collocation translation test as the only measure. Learning of different types of collocations such adjective+noun collocations via different corpus resources such the COCA, the BNC and GS can be compared. For a more robust quantitative evaluation, studies that employ more sensitive and multidimensional measures of L2 collocation knowledge can be conducted to capture improvement in receptive and productive L2 collocation knowledge through different resources. In addition, more versatile qualitative data via different methods such as interviews, focus groups and think aloud protocols from more participants can be collected to shed more light on affective and cognitive variables involved. Studies can also investigate user-friendliness and quality of interface, involvement loads of tasks and resources together with depth of processing they induce in locating and learning L2 collocation. The findings of the current study call for further studies with longer periods of treatment, more rigorous training of the participants in terms of resource consultation. Besides, higher number of participants with varying degrees of linguistic and collocation awareness and more dimensional and productive measures of collocation resources in more extended L2 tasks are warranted to gain a better insight into the affordances of DDL vs. deductive L2 collocation consultation resources.

REFERENCES

Acar, A., Geluso, J., & Shiki, T. (2011). How Can Search Engines Improve Your Writing*? CALL-EJ*, *12* (1), 1-10. Retrieved from http://www.callej.org/journal/12-1/Acar_2011.pdf.

Başal, A. (2019). Learning collocations: Effects of online tools on teaching English adjective-noun collocations. *British Journal of Educational Technology*, *50*(1), 342-356. https://doi.org/10.1111/bjet.12562

Boulton, A. (2010, December). Language awareness and medium-term benefits of corpus consultation. In *Proceedings of Eurocall 2009 conference: New trends in CALL-working together,* (pp. 39-46). Macmillan ELT.

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

Brezina, V. (2012). Use of Google Scholar in corpus-driven EAP research. *Journal of English for Academic Purposes*, *11*(4), 319-331. https://doi.org/10.1016/j.jeap.2012.08.001

Chen, Y. (2017). Dictionary use for collocation production and retention: A CALL-based study. *International Journal of Lexicography, 30*(2), 225-251. https://doi.org/10.1093/ijl/ecw005

Chinnery, G. (2008). You've got some GALL: Google-assisted language learning. *Language Learning & Technology, 12*, 3−11. http://dx.doi.org/10125/44126

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*(3), 301-315. https://doi.org/10.1016/s0346-251x(97)00024-9

Cobb, T. (1999). Breadth and Depth of Lexical Acquisition with Hands-on Concordancing. *Computer Assisted Language Learning, 12*(4), 345-360. https://doi.org/10.1076/call.12.4.345.5699.

Cobb, T. (n.d.). Vocabulary Profilers [computer program]. Available at https://www.lextutor.ca/vp/

Cohen, J. (1988). *Statistical power analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: SAGE.

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

Flowerdew, L. (2015). Data-driven learning and language learning theories. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning*, *69*, 15-36. https://doi.org/10.1075/scl.69

Frankenberg-Garcia, A. (2012). Learners' use of corpus examples. *International Journal of Lexicography, 25*(3), 273-296. https://doi.org/10.1093/ijl/ecs011

Geiller, L. (2014). How EFL Students Can Use Google to Correct Their" Untreatable" Written Errors. *The EuroCALL Review, 22*(2), 26-45. Retrieved from https://files.eric.ed.gov/fulltext/EJ1055226.pdf.

Geluso, J. (2013). Phraseology and frequency of occurrence on the web: native speakers' perceptions of Google-informed second language writing. *Computer Assisted Language Learning, 26*(2), 144-157. https://doi.org/10.1080/09588221.2011.639786

Given, L. M. (2016). *100 questions (and answers) about qualitative research*. SAGE Publications.

Han, S., & Shin, J.-A. (2017). Teaching Google search techniques in an L2 academic writing context. *Language Learning & Technology, 21*(3), 172–194.

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL, 26*(02), 163–183. https://doi.org/10.1017/S0958344014000020

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language learning, 51*(3), 539–558. https://doi.org/10.1111/0023-8333.00164

Kartal, G., & Yangın Ekşi, G. (2018). The effects of using corpus resources on EFL student teachers' learning and production of verb-noun collocations. *PASAA: Journal of Language Teaching and Learning in Thailand, 55*, 100–125. Retreived from https://eric.ed.gov/?id=EJ1191743

Kim, S. (2017). "Minimum input, maximum output, indeed!" Teaching collocations through collocation dictionary skills development. *Lexikos, 27*, 265-286. 10.5788/27-1-1403

Kim, S. (2018). EFL learners' dictionary consultation behaviour during the revision process to correct collocation errors. *International Journal of Lexicography, 31*(3), 312-326. https://doi.org/10.1093/ijl/ecx009

Kotamjani, S. S., Razavi, O. F., & Hussin, H. (2017). Online Corpus Tools in Scholarly Writing: A Case of EFL Postgraduate Student. *English Language Teaching, 10*(9), 61-68. 10.5539/elt.v10n9p61

Kvashnina, O. S., & Sumtsova, O. V. (2018). Using Google to search language patterns in web-corpus: EFL writing pedagogy. *International Journal of Emerging Technologies in Learning (iJET), 13*(3), 173-179. https://doi.org/10.3991/ijet.v13i03.7712

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

Lew, R., & Radłowska, M. (2010). Navigating dictionary space: The findability of English collocations in a general learner's dictionary (LDOCE4) and special-purpose dictionary of collocations (OCD). In A. Ciuk, & K. Molek-Kozakowska (Eds.), *Exploring Space: Spatial Notions in Cultural, Literary and Language Studies, Volume 2: Space in Language Studies* (pp. 34-47). Cambridge Scholars Publishing.

Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography, 24*(1), 29–49. https://doi.org/10.1093/ijl/ecq039

Li, S. (2017). Using corpora to develop learners' collocational competence. *Language Learning & Technology, 21*(3), 153-17. 10125/44625

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies, 22*, 1-18.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The language teacher, 31*(7), 9–13.

Nesselhauf, N., & Tschichold, C. (2002). Collocations in CALL: An investigation of vocabulary-building software for EFL. *Computer Assisted Language Learning, 15*(3), 251-279. https://doi.org/10.1076/call.15.3.251.8190

Nurmukhamedov, U. (2016). The contribution of collocation resources to collocation correction in second language writing. *International Journal of Lexicography, 4*(8), 1–29. https://doi.org/10.1093/ijl/ecw031

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research, 20*(1), 113–138. https://doi.org/10.1177/1362168814568131

Rezaee, A. A., Marefat, H., & Saeedakhtar, A. (2015). Symmetrical and asymmetrical scaffolding of L2 collocations in the context of concordancing. *Computer Assisted Language Learning, 28*(6), 532-549. https://doi.org/10.1080/09588221.2014.889712

Robb, T. (2003). Google as a quick 'n dirty corpus tool. *TESL-EJ, 7*(2). Retrived from: http://www.tesl-ej.org/wordpress/issues/volume7/ej26/ej26int/?wscr=

Shei, C.-C. (2008). Discovering the hidden treasure on the Internet: Using Google to uncover the veil of phraseology. *Computer Assisted Language Learning, 21*(1), 67−85. https://doi.org/10.1080/09588220701865516

Sha, G. (2010). Using Google as a super corpus to drive written language learning: a comparison with the British National Corpus. *Computer Assisted Language Learning, 23*(5), 377-393. https://doi.org/10.1080/09588221.2010.514576

Sun, Y.-C., & Wang, L.-Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning, 16*, 83–94. https://doi.org/10.1076/call.16.1.83.15528

Tsai, K. J. (2019). Corpora and dictionaries as learning aids: inductive versus deductive approaches to constructing vocabulary knowledge. *Computer Assisted Language Learning, 32*(8), 805-826. https://doi.org/10.1080/09588221.2018.1527366

Wu, Y. J. A. (2015). *Utilizing corpus resources accompanied by other consultation resources in enhancing collocation accuracy and collocation richness in L2 writing*. [Unpublished doctoral dissertation, University of California] ProQuest Dissertations Publishing.

Appendix A

a. Sample Items from the Collocation Translation Test

Şaka kaldırmak: TAKE a joke (This was given as an example)

_____

1. (birinin ya da kendinin) merağını gidermek: _____one's curiosity
2. Standartları belirlemek:_____ the standards
3. Çaba göstermek:_____ an effort