Language Learning & Technology
ISSN 1094-3501

October 2017, Volume 21, Issue 3
pp. 172–194

ARTICLE

Teaching Google search techniques in an L2 academic writing context

Sumi Han, *Hallym University*

Jeong-Ah Shin, *Dongguk University*

Abstract

This mixed-method study examines the effectiveness of teaching Google search techniques (GSTs) to Korean EFL college students in an intermediate-level academic English writing course. 18 students participated in a 4-day GST workshop consisting of an overview session of the web as corpus and Google as a concordancer, and three training sessions targeting the use of quotation marks (“”) and a wildcard (*). Each session contained a pre-test, a 30-minute training, and a post-test, and each training session focused on one of the three key writing points: articles, collocations, and paraphrasing. Two questionnaires for demographic information and GST learning experiences were conducted. The results showed a statistically significant effect for the overall gain score. In particular, participants' use of articles greatly improved after the training—in contrast to their use of collocations and paraphrasing. Lack of grammar and vocabulary knowledge seemed to hinder their data-driven learning, especially for collocation use and paraphrasing. The questionnaire data showed that all students found the GSTs beneficial, mostly because they were easy to use for confirmation and correction. Overall, both quantitative and qualitative data suggest that teachers' meticulous guidance and vigilant individualized feedback are necessary to facilitate L2 self-directed Google-informed writing.

Keywords: *Computer-Assisted Language Learning, Data-Driven Learning, Google, L2 Writing*

Language(s) Learned in this Study: *English*

APA Citation: Han, S., & Shin, J.-A. (2017). Teaching Google search techniques in an L2 academic writing context. *Language Learning & Technology*, *21*(3), 172–194. Retrieved from http://llt.msu.edu/issues/october2017/hanshin.pdf

Introduction

Recent developments in technology have had a notable impact on college composition classrooms. In computer-assisted language learning (CALL) environments, more teachers and students engage in real-time electronic writing, with more actual student writing and immediate, online teacher feedback (Egbert, 2005; Kern, 2006). Students write using software such as Microsoft Word, and teacher feedback is provided by email or on students' written files, thereby saving time and triggering active communication between teachers and students. In the meantime, online resources such as Google and other search engines have been integrated into writing instruction because of their easy accessibility and representativeness of real-life language usage (Kilgarriff & Grefenstette, 2003; Shei, 2008). Many corpora are also available for language research and instruction, such as the Corpus of Contemporary American English (COCA; Davies, 2008) and the British National Corpus (BNC; BNC Consortium, 2007). Currently, large amounts of linguistic resources are accessible through computers and smartphone applications, with no great regional or temporal restrictions.

The use of these linguistic resources has made data-driven learning (DDL) possible in second language (L2) instruction. Boulton (2010) explains the potential benefit of DDL as "encouraging noticing and consciousness-raising, leading to greater autonomy and better language learning skills in the long term" (p. 535). However, some concerns about DDL exist because it can be intensive and cost-ineffective. For

instance, learners may need to invest considerable time and effort in order to identify correct language usage while learning with DDL. Further, the available computers are often not sufficient for the large population of students in many schools (for a detailed discussion of potential barriers to the use of DDL, see Boulton, 2010).

To the best of our knowledge, most writing courses in EFL settings such as Japan and Korea are based on traditional writing instruction, where the teacher is the only language expert. The workload of writing teachers is often daunting, and instruction and feedback consequently tend to be less tailored for individual college students with differing writing profiles. Most importantly, traditional teacher-centered instruction can take away priceless learning opportunities from our students, who as digital natives grew up and live surrounded by digital technology and instant messaging. Integrating technology into L2 writing classrooms, if optimally introduced, potentially reduces the burden on writing teachers while students actively engage in writing tasks, thereby increasing students' confidence as writers (Chapelle, 2003; Milton, 2006). As writing teachers, we can raise the following questions: What type of technology do we need to introduce into our L2 writing instruction? And how can we make it support our students?

To answer these questions, the present study focuses on Google-informed L2 writing, which can lead to inductive (and potentially deductive) learning in L2 writing instruction (e.g., Acar, Geluso, & Shiki, 2011; Geluso, 2013). More specifically, this study examined the effectiveness of teaching two Google search techniques (GSTs) in a Korean EFL college academic writing context with novice writers at varying levels of English proficiency: quotation marks (" ") and the asterisk as wildcard (*). Google consultation in the study focused on three of the most challenging writing points for Korean EFL learners: usage of articles, collocations, and paraphrasing. A detailed teaching procedure for the GST training workshop is described in the Methods section, which is often lacking in existing research but is very informative in L2 writing pedagogy.

Literature Review

Data-Driven Learning

Recently, DDL has been more widely adopted as a way to teach and learn foreign languages with the advance of information technology (Granger, Hung, & Petch-Tyson, 2002; Reppen, 2010). Johns (1991), who coined the term DDL, defined it as an inductive approach to language learning that makes learners explore linguistic data from a corpus (i.e., a collection of authentic language use) on their own. Typical language data consist of concordance lines that reveal the uses of target words or phrases with the support of corpus software called a concordancer. In fact, DDL also supports deductive learning with teacher-guided materials where students are directed to examine particular language points (for an in-depth discussion on DDL theories of language learning, see Flowerdew, 2015).

In DDL, the roles of the teacher and student differ from those in traditional language instruction. Teachers who take a DDL approach act as facilitators of learners' self-directed learning rather than as the sole language experts in the classroom while L2 learners are guided to engage in problem-solving activities where they raise their awareness of language and become independent. In the context of L2 writing instruction, for example, concordancing can be used to help learners edit their own writing; and while honing self-editing skills (i.e., self-confirmation and self-correction) through corpus consultation, L2 learners become autonomous writers (Liu & Jiang, 2009; Reppen, 2010; Todd, 2001).

One of the major concerns of DDL is that it may not be ideal for all students. Flowerdew (2015) argued that "certain types of learners (because of their learning styles, previous learning background, etc.) may resist this kind of exploratory, process-oriented learning environment" (p. 18). For instance, using concordancing as a cognitive tool, intermediate and advanced students seem to benefit most from DDL. Liu and Jiang (2009) only targeted intermediate and advanced L2 learners, excluding lower-level students because "corpus-based learning was deemed to be too difficult for them" (p. 63) Relatively little research has examined low-level students' DDL-based L2 writing (e.g., Chujo, Oghigian, & Akasegawa, 2015; Yoon

& Jo, 2014). Likewise, learning styles, search strategies, or learning goals varying across students have also been examined as potential determinants of DDL-based learning gains (e.g., Yoon, 2016; Yoon & Hirvela, 2004). The question then is, if traditional forms of DDL are not appropriate for all learners, would the use of the web as corpus and Google as a concordancer can help to open it up to a wider population?

Web as a Corpus and Google as a Concordancer

Over the last decade, the web has become a powerful and growing resource of information, and such data have been introduced into the DDL literature as web corpora (see also Kilgarriff & Grefenstette, 2003). Google and Yahoo as search engines as well as online corpora and concordancers have been widely used for simple and quick information searches. Recently, Davies and his colleagues launched three web-based English corpora, which grow every day: the Newspapers on the Web corpus (NOW; Davies, 2013a), with 3 billion words; the corpus of Global Web-based English (GloWbE; Davies, 2013b), with 1.8 billion words for more than 20 world Englishes; and the Wikipedia corpus, with 1.9 billion words from more than 4.4 million Wikipedia articles (Wikipedia; Davies, 2015). L2 teachers and practitioners have introduced such online language data to their classrooms (e.g., Franken, 2014; Wu, Franken, & Witten, 2009), but using the web as an authentic online corpus seems to also be promising.

Among online web-based resources, Google has received the most attention from language teachers and practitioners to date, partly because of its user-friendly interface with an immense amount of language usages in various registers from personal blogs to academic journals. Google provides a simple, intuitive interface that does not require much linguistic or metalinguistic knowledge and skills (Boulton, 2013). Typing any query into the search box for this free, ubiquitous, and gigantic corpus allows users to retrieve a considerable amount of relevant search results in a fraction of a second, including texts, photos, and videos. This is the reason why Google has been by far the most popular search engine with a 64% share of web search volume worldwide as of August 2015 (comScore, 2015). Maximizing the use of Google in language learning has been called Google-Assisted Language Learning, or GALL, as part of DDL-based language learning (Chinnery, 2008). Especially in L2 writing, Google as a corpus and concordancer is particularly of interest to teachers and researchers. Like other DDL materials, Google consultation also helps L2 learners to increase their awareness of written language usage and to self-edit their words, structures, organization, and even content information for improved writing, thereby enhancing discovery skills and learner autonomy (Panah, Yunus, & Embi, 2013; Robb, 2003; Shei, 2008).

Google-Informed L2 Writing and Empirical Research

The implementation of Google in L2 writing instruction is a teaching practice for training L2 learners to search for language use on Google for their own real-time electronic writing purposes (Geluso, 2013). The success of such Google-informed L2 writing instruction lies in how successfully L2 writers are taught to use Google as a language reference tool. Because such online data (not solely from Google) include erroneous or unnatural expressions, self-driven discovery learning of language usage is not optimal without a certain amount of teacher guidance. With appropriate intervention, Google as a quick and dirty corpus tool has the potential to be useful in L2 writing (Boulton, 2013; Robb, 2003).

Previous empirical studies of Google-informed L2 writing instruction reveal various study purposes and designs, but can be classified into two groups according to their scope: those focusing solely on Google consultation and those comparing Google with other online resources such as specialized corpora, concordancers, or dictionaries. Common linguistic points to teach via Google consultation include grammatical accuracy (e.g., articles or spelling), writing clarity or naturalness (e.g., synonyms or collocations), and sentence or pattern usage.

The first group examines the effectiveness of teaching GSTs such as the use of quotation marks and wildcards for error correction or the use of natural expressions. Acar et al. (2011) examined how quotation mark searches helped eight engineering majors to correct their written errors such as articles and prepositions following a 20-minute training session. The results showed that their English as a Second Language (ESL) students subsequently improved 24% of their original sentences in writing clarity (degree

of naturalness) and grammatical improvement (fewer grammatical errors). Geluso (2013) also focused on quotation mark searches for the use of more natural formulaic sequences (FSs) in L2 writing. After a 40-minute session of Google search training, 25 Japanese freshmen used the search technique to revise half of 334 learner-generated phrases based solely on the frequency of occurrence of FSs. The mean score from four raters (English native speakers) was 0.83 higher (see Table 2) for Google-informed phrases than for non-Google-informed phrases for perceived naturalness. Geiller (2014) explored whether training in use of quotation marks and a wildcard led to successful error correction by 17 French college EFL students, using Google's Custom Search Engine with 28 British and American websites to eliminate less reliable search results. The advanced learners were taught how to use GSTs and derive frequency evidence, and they accurately self-corrected around 52% of their written errors.

The second category of Google-informed L2 instruction research focuses on the relative effectiveness of Google consultation compared to other online reference tools such as COCA, the BNC, or multiple DDL tools. Brezina (2012) examined the potential of Google Scholar as a linguistic resource in English for academic purposes research. Targeting two disciplines (applied linguistics and physics), Brezina argues that the advantages of Google Scholar over the COCA academic subcorpus lie in its wide representativeness of written academic language and register variation in academic patterns. Sha (2010) conducted an experimental study on the effectiveness of the aid of Google in DDL in comparison with the BNC and found Google searches were superior and preferred because of its usability, search speed, and varying solutions. Yoon (2016) investigated how six Korean ESL graduate students used different DDL resources in writing and found that they were used for different purposes: a bilingual dictionary mostly for synonyms, COCA for collocations, and Google for simple confirmation. Despite the varied characteristics of these studies, the overall findings are that (a) most L2 learners find GSTs more beneficial for their academic writing, (b) a combination of Google consultation and other resources is needed for more language usage evidence, and (c) a small amount of time and effort was enough to train L2 learners in GSTs.

Two main limitations of previous research on Google informed L2 writing are noteworthy: the little training in search phrase formulation, and frequency-based evaluation of Google search results. As discussed above, deciding what to search for on Google was a daunting task for most of the learners (e.g., Geiller, 2014). Davies (2013c) argues that COCA-based searches in academic writing are done quickly and easily. In addition, frequency (as measured by the estimated number of hits for the search phrase) can help decide its appropriateness as a valid phrase, but in many cases frequency information fails to provide conclusive evidence or is difficult to interpret. In contrast to online corpora such as COCA and the BNC, Google search results are not sorted as concordance lines with the search phrase centered, which seemed to make the learners feel it was difficult to explore evidence for the target phrase (for an example of Google results, see Figure 1). In this regard, to maximize the effectiveness of Google consultation in L2 writing, it is critical to offer detailed step-by-step instructions for formulating phrases with specific GSTs and for interpreting search results.

To address the gaps in Google as a DDL approach, the present study, with its mixed-methods research design, addressed the following four research questions:

1. What is the overall effectiveness of teaching GSTs to Korean EFL students in using articles, collocations, and paraphrasing?
2. Compared to a bilingual dictionary, how much use did the students make of GSTs for articles, collocations, and paraphrasing?
3. How did the students use GSTs in their own draft writing?
4. What are students' perceptions and attitudes toward GSTs?

Note that the choice of the three writing components (i.e., articles, collocations, and paraphrasing) was practically motivated in order to provide additional aid in learning these traditionally difficult linguistic points of English writing.

Methods

Settings and Participants

The study was conducted in an intermediate-level academic English writing course, *College English Writing II*, at a university in Seoul, South Korea, over a five-week winter semester. In a computer-mediated learning setting, the class met three times per week for three hours. The participants were 20 Korean students enrolled in the writing course. They signed the informed consent form and voluntarily participated in this study, but the data analysis was based on 18 students because two of them missed one of the three training sessions.[1]

Table 1 lists the demographic profiles of the 18 students (13 male, 5 female) from a pre-treatment questionnaire with 14 STEM majors (Science, Technology, Engineering, and Mathematics). The student age ranged from 19 to 25, with the majority being in their third year at the university. Their self-rated English scores on the four skills (i.e., listening, reading, speaking, and writing) using a 5-point Likert scale (1 = very poor, 2 = poor, 3 = acceptable, 4 = good, 5 = very good) indicated that they felt that their skills were either poor or very poor, except for their acceptable reading skills. Though the course targeted intermediate-level students, their self-reported English proficiency scores for the Test of English developed by Seoul National University (TEPS) showed that the class included low- to high-intermediate-level students. Only one student had studied abroad in an English-speaking country for eight months; the rest had no such experience. None had prior experience using GSTs.

Table 1. *Profiles of Participants (N = 18)*

Characteristic	***n***	***M***	***SD***	***Range***
Gender				
Male	13			
Female	5			
Major				
STEM[a]	14			
Non-STEM	4			
Academic Year				
Freshman	1			
Sophomore	3			
Junior	9			
Senior	5			
Age		22.67	1.53	19–25
Proficiency test score[b]		659	100	420–780
Self-rated skills[c]				
Listening		2.78	1.00	2–5
Reading		3.22	0.88	2–5
Speaking		1.89	0.90	1–4
Writing		2.17	0.86	1–4

Note. [a] Science, Technology, Engineering, Mathematics; [b] TEPS (score range: 0-990); [c] 5-point Likert scale

GSTs in this Study

Among the Google advanced search operators, two were targeted in the study: quotation marks (“ ”) and the asterisk as a wildcard (*). As described above, the use of the quotation marks helps ensure a search for the phrase as a whole string rather than individual lexical words in any form (inflected or not) in any order. The use of the wildcard shows common collocates within a word string. The asterisk only works in place of one or more whole words (with a space either side), unlike the asterisk in many corpora where it can either be a whole word or (when attached to a word) any string of letters as part of a word. It also works best within quotation marks. Based on the literature review, the authors expected the application of the two GSTs to be most beneficial to L2 learners in their use of articles, collocations, and paraphrasing for academic writing (e.g., Acar et al., 2011; Park, 2010; Shei, 2008; Yoon, 2016). In the following, application of the GSTs to teaching each of the three writing components is explained and modeled with the same sample sentence: *I will conduct a research on economic influences of air pollution*. All Google searches reported in the study as examples were conducted in July of 2016.

First, the quotation marks are particularly useful to EFL learners for correcting or confirming their own article usage. To write the above sample sentence accurately, learners may wonder if an article should be used with the noun *research*. In this case, Google Scholar searches of *“conduct a research”* and *“conduct research”* with the quotation marks can help them quickly choose the latter as a more frequent and thus probably more natural expression. Figure 1 shows screenshots of the search phrases and the relevant information: the number of hits (15,600 vs. 135,000) and the context (in bold). These present evidence for choosing the phrase without the indefinite article, *conduct research*.[2]

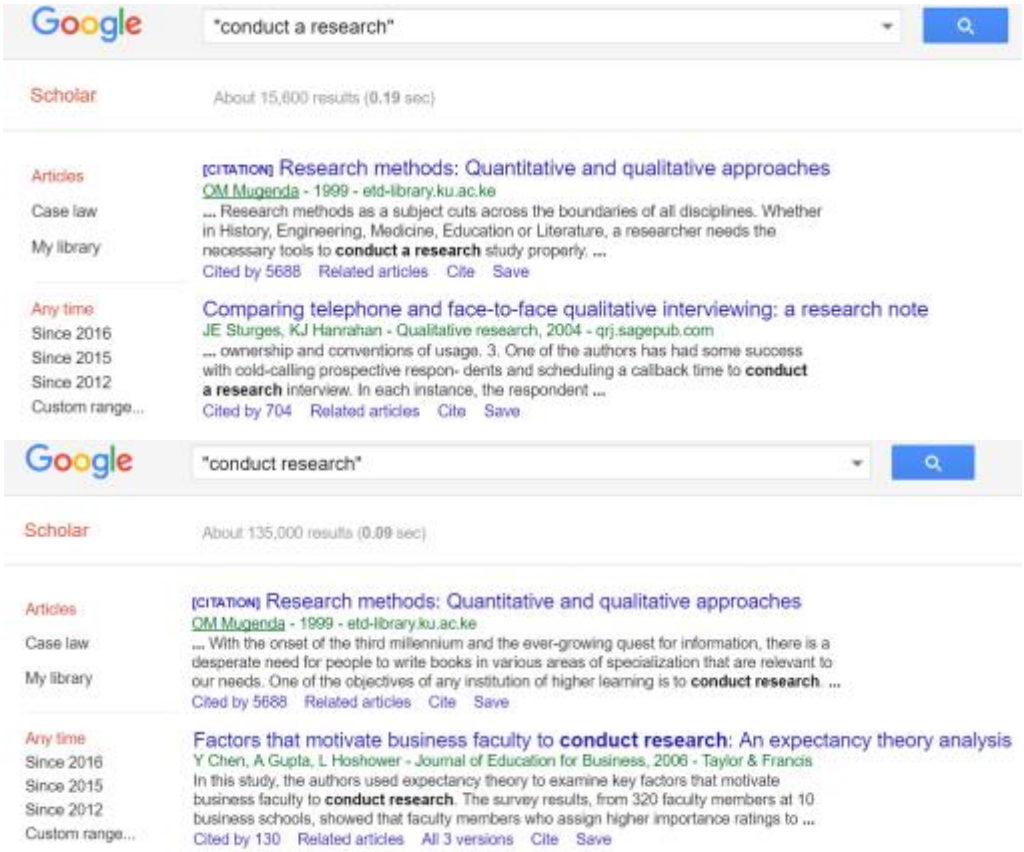

Figure 1. Screenshots of *“conduct a research”* (top) versus *“conduct research”* (bottom), using quotation marks

Second, the wildcard in conjunction with the quotation marks supports the retrieval of words that are frequent collocates in the context. In academic writing, EFL learners often find it challenging to use appropriate collocations. Figure 2 shows a screenshot of the search results of *"will * research"* in Google Scholar, suggesting three verb collocates of the noun *research*: *stimulate*, *advance*, and *conduct*. Other verb collocates retrieved include *do*, *carry out*, *focus on*, and *inspire*. Applying the two GSTs, learners can gain more collocational knowledge while browsing search results further, and ultimately increase the diversity of their expressions.

Figure 2. A screenshot of "will * research" for using wildcard (within quotation marks)

Combined use of the two GSTs demonstrated in Figure 1 and Figure 2 can have various benefits, including improving paraphrasing skills, as explored in the third session. *Paraphrasing* means reshaping the original source with different grammar structures and vocabulary while maintaining the original meaning. In particular, academic writing requires this skill to avoid plagiarism with structure change and source citations, but most importantly to deliver the main idea through the topic sentence and concluding sentence in paragraph writing. In this regard, search results with the support of the two GSTs quickly and convincingly suggest natural grammar structures or collocations when paraphrasing the original expression.

Procedure

Figure 3 shows the procedure for the 4-day workshop with the topics and the materials used. The first author, as the course instructor, offered the training sessions. On Day 1, the background information of the participants given in Table 1 was obtained using a pre-treatment questionnaire. An overview session on using Google as a web corpus and concordancer was also offered to increase the students' knowledge about corpus consultation for their academic writing. They learned how to obtain evidence with different Google *subcorpora*: Google News, Google Books, and Google Scholar.[3] Subsequent days featured GST Session 1 for articles, GST Session 2 for collocations, and GST Session 3 for paraphrasing. The training sessions were offered in an incremental manner, beginning with the use of quotation marks first introduced for article usage on Day 2. To confirm if they understood the first session and to evaluate how they used the first search technique, a take-home assignment was given so that the students' second drafts of a one-paragraph writing task (*What is your life goal and why?*) would include underlined expressions for which they used the quotation marks.[4] On Day 3, the wildcard was introduced for collocation use, and on Day 4, both GSTs were combined to help with paraphrasing sentences.

For each writing component, 30 minutes of training (with a practice exercise) was offered with a pre-test administered two days before and a post-test immediately after each session. In particular, Pre-tests 1 and

2 and Post-tests 1 and 2 were divided into two parts with five items in each part: Part A (error correction) and Part B (Korean-to-English translation). Pre-test 3 and Post-test 3 included only five paraphrasing items (for samples of the three exercises used in the sessions, see Appendix A; for Pre-tests and Post-tests 1, 2, and 3, see Appendix B, Appendix C, and Appendix D, respectively). All the examples and sentences used in the exercises and tests were common mistakes taken from the students' written work in quizzes, exams, short reflection assignments, or emails to control for the level of vocabulary and grammar structures used in the instructional materials. In a norming session administered ahead of the workshop, the equal difficulty for each pair of pre- and post-tests was validated. Finally, a post-treatment questionnaire was conducted to obtain the students' experiences and perceptions pertaining to the GST workshop (see Table 4). The students completed all the work electronically using Google Docs.

Day 1	Pre-treatment questionnaire (10 minutes)
⬇	Pre-test 1: Articles (20 minutes)
	Overview of Google as a web corpus and concordance (20 minutes)
Day 2	GST Session 1: Articles (30 minutes)
	Post-test 1: Articles (20 minutes)
⬇	Pre-test 2: Collocations (20 minutes)
	Take-home Assignment
Day 3	GST Session 2: Collocations (30 minutes)
⬇	Post-test 2: Collocations (20 minutes)
	Pre-test 3: Paraphrasing (20 minutes)
Day 4	GST Session 3: Paraphrasing (30 minutes)
	Post-test 3: Paraphrasing (20 minutes)
	Post-treatment Questionnaire (25 minutes)

Figure 3. Procedure of GST Workshop

To help the reader better understand exactly what was done during the training sessions, a sample question from the exercise in GST Session 1 is provided in Figure 4, with search phrases and results. With the sentence provided, each student identified any unnatural item and underlined it, and then the whole class discussed it with a focus on the article usage. Next, each student presented his or her own correction. They could search any unknown words in an online Korean–English bilingual dictionary offered by the Korean search engine Naver. To confirm or correct their first answer, the instructor modeled formulating search phrases with and without the indefinite article *a* and the quotation marks on Google Books. The search results of first two search phrases without the quotation marks provided the students with hints for more possible search phrases. Students were then guided to search for each possible phrase with the quotation marks and check the numbers of hits and the context (e.g., Does the indefinite article modify the noun *research*?). Based on the two pieces of Google information, they either confirmed or corrected their first answer.

I will conduct a research on English learning.
My first answer(s): conduct research

Search phrases	Search Results (Hit numbers / Context)
1. conduct a research	1. NA (Not Applied)
2. conduct research	2. NA
3. "conduct a research"	3. 57,700 / ~ study, proposal, project
4. "conduct research"	4. 229,000 / conduct research

My final answer(s): conduct research

Figure 4. Sample question from GST session 1 (Part A) with search phrases and results

Data Analysis

The data of this mixed-method study in a quasi-experimental design included both quantitative and qualitative analyses of the students' performance on the tests as well as their responses to the questionnaires. For the test data, the answers of Pre- and Post-tests 1 and 2 were coded as 1 for correct answers and 0 for incorrect answers, and descriptive statistics and effect sizes were computed. In contrast, for Pre- and Post-test 3, a coding sheet was developed and used to quantify and classify the students' paraphrased sentences. Appendix E presents three coding categories: grammar and vocabulary appropriateness (0 for inappropriate, 1 for appropriate), the degree to which the paraphrased sentence had the same meaning as the original sentence (0 for unsuccessful, 1 for somewhat successful, 2 for successful), and types of paraphrasing (1 for grammatical structure, 2 for vocabulary use, and 3 for both grammatical structure and vocabulary use). The points from the first two categories were used as scores for Pre- and Post-test 3 (15 points each), while the points from the third category were used to further examine how the students focused on grammar or vocabulary or both. For the data from the pre- and post-tests, an effect size for learning gains was calculated and interpreted based on the benchmarks for L2 instructional research using within-groups designs by Plonsky and Oswald (2014): small (0.60), medium (1.00), and large effect (1.40). The first author and a graduate student coded the test data and the inter-coder reliability (i.e., the Pearson correlation coefficients based on the percentage of agreement) was 0.88; any disagreement was resolved after discussion.

The other data were coded and analyzed as follows: First, the students' self-reported percentage of usage of the GSTs and the bilingual dictionary pre-test was analyzed to see how much they used the search techniques. Second, the underlined expressions in the second drafts were analyzed and classified into four types of changes compared to those in the first drafts: confirmation, correction, revision, and new information. These four types of changes demonstrated how each student employed the quotation marks in process-oriented writing for confirming their original expressions, correcting any grammar or vocabulary errors, fully revising the expression, or adding more information with new words. Frequency counts for these four types were computed per student.

Finally, the responses to the nine questions on the 5-point Likert scale post-treatment questionnaire were strongly disagree (1), disagree (2), neither disagree nor disagree (3), agree (4), and strongly agree (5). The questions were split into three categories: computer use and knowledge (Questions 1–2), training in GSTs (Questions 3–6), and use of GSTs (Questions 7–9). For the last question, the students were asked to write their reflection on the GST training in English.

Results

This section provides the results of both quantitative and qualitative analyses of the measures used for the present study. The data collected included the test scores and responses to the post-test questionnaire.

Table 2 presents descriptive statistics for the pre- and post-tests administered to measure learning gains in

each of the three GST training sessions. On average, the post-test scores were higher than the pre-test across the three sessions. In particular, the use of English articles (Session 1) improved significantly ($p < .01$) and with a medium effect ($d = 1.13$) after the training. However, the gains in the use of collocations (Session 2) and paraphrasing (Session 3) were less pronounced. The overall gain score (Total) was 3.89, which is statistically significant ($t = 3.84$, $p < .01$, $d = 0.86$).

Table 2. *Descriptive Statistics of Pre-test and Post-test by Session*

	Pre-test		Post-test					
Session	***M***	***SD***	***M***	***SD***	***t***	***p***	***d***	**95% of CIs**
1	4.17	1.82	6.44	2.18	3.70	0.00	1.13	0.43–1.83
2	6.61	1.88	7.11	7.11	1.14	0.26	0.29	-0.37–0.95
3	11.06	3.19	12.17	12.17	1.17	0.09	0.43	-0.23–1.09
Total	21.83	4.99	25.72	4.04	3.84	0.00	0.86	0.17–1.53

Note. Maximum scores: 10 for each test in Sessions 1 and 2, and 15 for each test in Session 3.

Figure 5 shows the percentage for using the two online tools, Google and Naver, while completing Post-tests 1, 2, and 3 for article usage, collocations, and paraphrasing, respectively. Both Post-tests 1 and 2 included Parts A and B (error correction and Korean-to-English translation, respectively), while Post-test 3 included only one part. Overall, the students reported that they used GSTs more than the dictionary (71% vs. 29%). More specifically, they heavily relied on GSTs rather than the dictionary to correct any erroneous expressions in Part A of both Post-tests 1 and 2. To translate the Korean sentences into their English counterparts, they relied relatively more on Naver, but still used GSTs approximately twice as much as the dictionary. For Post-test 3, they used GSTs slightly more than the bilingual dictionary, but the difference was the smallest among those of the three post-tests (58.8% vs. 42.4%).

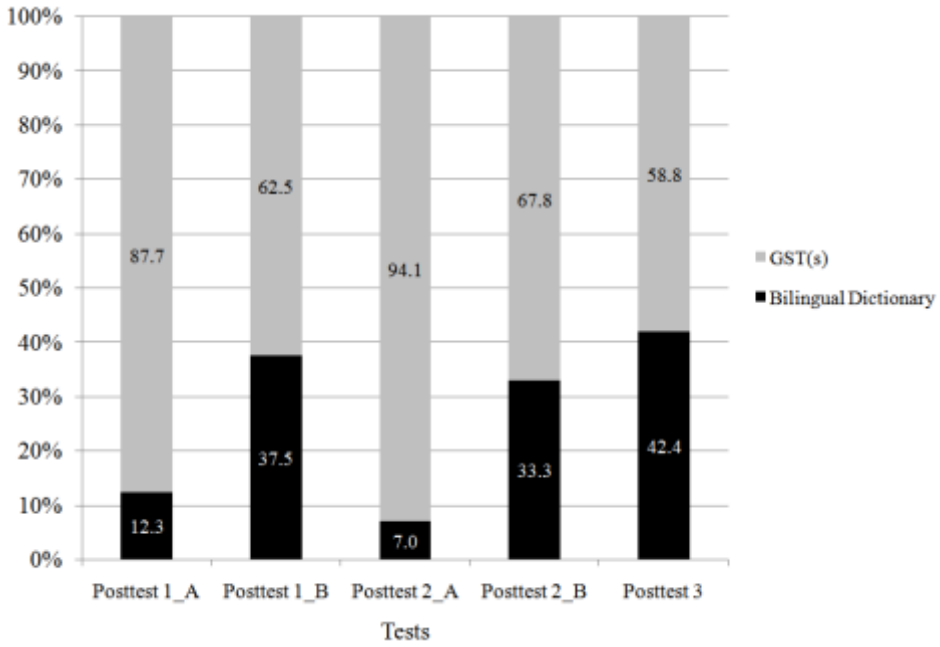

Figure 5. Percentage use of GSTs and the bilingual dictionary across post-tests

Table 3 presents the frequencies of the four types of Google searches on the revised expressions, which the participants underlined when they used the quotation marks in Google from Draft 1 to Draft 2 (three of the 18 students did not turn in the homework). The remaining students made a total of 129 changes to their

second drafts, including 36 for confirmation, 23 for correction, 33 for revision, and 37 for new information.[5] The average frequencies in type range from 1.5 for correction to 2.5 for new information, totaling 8.7. Although the frequencies of the four types vary from student to student, many students with intermediate-level to advanced-level proficiency attempted to use the quotation marks for the different types; Seongjoon and Saemi made considerable use of the search technique for confirmation.

Table 3. *Frequencies of Four Types of Google Searches (from Draft 1 to Draft 2)*

		Types of Google Searches				
Names	**Proficiency**	**Confirmation**	**Correction**	**Revision**	**New Information**	**Total**
Jinho	intermediate	1	0	5	3	9
Seonho	upper-intermediate	2	3	2	4	11
Sejin	advanced	0	0	1	1	2
Min	low	_[b]	_[b]	_[b]	_[b]	_[b]
Woobin	intermediate	_[b]	_[b]	_[b]	_[b]	_[b]
Moosung	intermediate	0	1	0	0	1
Woojae	low	3	6	1	2	12
Dongjin	intermediate	5	3	3	1	12
Seongjoon	intermediate	9	1	3	2	15
Hyeok	intermediate	1	1	2	4	8
Dongjae	intermediate	0	1	1	4	6
Saemi	upper-intermediate	12	0	1	2	15
Hoon	intermediate	1	2	4	0	7
Joongi	intermediate	0	0	2	3	5
Sooji	advanced	_[b]	_[b]	_[b]	_[b]	_[b]
Inha	intermediate	0	1	3	3	7
Song	upper-intermediate	0	0	1	8	9
Yumee	upper-intermediate	2	4	4	0	10
Total (*M*)		36 (2.4)	23 (1.5)	33 (2.2)	37 (2.5)	129 (8.7)

Note. [a] All student names are pseudonyms. [b] No homework submission

Table 4 presents descriptive statistics of the students' answers to the nine questions on the post-treatment questionnaire. Overall, the students reported that their computer use and skills were fair to some extent (M = 3.83, SD = 0.68), and that the GST training was very effectively administered (M = 4.45, SD = 0.63). As a result, they said they could and would use the GSTs for English writing (M = 4.50, SD = 0.65).

Table 4. *Descriptive Statistics for Responses to Post-Treatment Questionnaire*

Questions	***M***	***SD***
Computer use and knowledge		
1. Before this semester, I used computers frequently.	4.22	0.65
2. I have good computer skills and knowledge.	3.44	0.70
Total	3.83	0.68
Training in GSTs		
3. The training helped me improve my writing in general.	4.61	0.50
4. The training helped me correct my expressions and improve my writing.	4.67	0.49
5. The training helped me confirm my expressions and improve my writing.	4.44	0.51
6. The training time was enough to learn about GSTs.	4.06	1.00
Total	4.45	0.63
Use of GSTs		
7. I can do Google searches with and without quotation marks (“ ”).	4.39	0.78
8. I can use the wildcard (*) for Google searches.	4.39	0.61
9. I will use the GSTs for my writing in the future.	4.72	0.57
Total	4.50	0.65

Note. 5-point Likert scale, from 1 for strongly disagree to 5 for strongly agree.

The student’s open-ended comments or feedback on the GST training were also collected. Most students found the training informative and effective for their English writing, but a few still preferred the online dictionary to the GSTs. Comments written by students at different proficiency levels include the following, reproduced verbatim:

> It was much easier to paraphrase using the google search technique. Especially, * is really useful when I do not know which verb or adjective to use with a certain noun. It was also easier to find an appropriate synonym with the search technique (Sejin, advanced level).
>
> It was good to know Google search techniques. It has been helpful when I wrote paragraphs. I think I will use these techniques after this class. It is because it helps me to write in English more naturally (Saemi, upper-intermediate).
>
> Google search technique helps me when I try paraphrasing. I could correct the details by using the technique. However, I still have trouble in changing vocabulary when I only use this technique (Seonho, upper-intermediate level).
>
> I think it is more convenient for me to use Naver online dictionary rather than Google search. Because I have difficulty in using new vocabulary to paraphrase. In Naver dictionary, I can find vocabulary easier than in Google (Song, intermediate level).
>
> I think it is very useful. After training, finding synonym and collocation is much easier. But paraphrasing is also challenging for me (Min, low level).

Discussion

The four research questions in the present study are addressed below based on the test and questionnaire data.

Research Question 1: The Overall Effectiveness of Teaching GST in L2 Writing

The overall results showed that the GST training yielded a small-to-medium effect size (d = 0.86) for teaching the three writing components: articles, collocations, and paraphrasing. More specifically, teaching the use of quotation marks for article usage made a significant contribution to the overall gain score, resulting in a medium effect size (d = 1.13). In contrast, training on the quotation marks and wildcard helped the students use collocations and paraphrasing better, but the learning gains were not statistically significant.

The discrepancy in the learning gains among the three writing components reflects the relative difficulty in learning fundamental knowledge or skills in order to become fluent writers. Note that an incremental learning progress was considered in presenting the three writing components to these Korean EFL learners. As expected, the use of quotation marks was extremely easy and helpful in error correction of grammar and vocabulary usage, in particular for the use of articles in the present study (see also Acar et al., 2011). As in many previous studies, our novice writers were taught to use both the quotation marks and the wildcard and were successful in finding appropriate word combinations or lexico-grammatical patterns such as collocations or FSs (e.g., Conroy, 2010; Geiller, 2014; Geluso, 2013; Park, 2010). They also learned how to paraphrase by consulting Google. Learning gains in collocations and paraphrasing were marginal compared to gains in using articles, perhaps because Google consultation, in particular the use of quotation marks, was exceptionally useful for article usage. Another explanation might be that all but advanced EFL learners lack collocation and paraphrasing skills (Nesselhauf, 2005). For this reason, the GST workshop administered in a short period of time may not be enough to yield significant progress in the use of collocations or paraphrasing in L2 writing.

Research Question 2: Perceived Usage of GSTs versus the Bilingual Dictionary

The data for perceived usage of Google and the bilingual electronic dictionary on each post-test suggest that these learners used the two CALL tools differently for different writing components and tasks. While several previous studies have compared the effectiveness of using Google and other online resources (e.g., a corpus or dictionary), the present study reflected a real writing situation where the Korean EFL learners often look up words or expressions in an online bidirectional bilingual dictionary (see Park, 2010; Yoon, 2016). They were allowed to consult the online dictionary during the entire GST workshop to resolve any uncertainty while performing any in-class tasks.

On average, these novice EFL writers preferred Google consultation to the bilingual dictionary for completing error-correction tasks for both articles and collocations. However, they relied more on the bilingual dictionary when working on more productive tasks (i.e., L1-to-L2 translation or paraphrasing). A tentative conclusion based on this differential use of CALL tools might be that GSTs can be more useful for controlled writing tasks than productive ones because they offer immediate feedback. With controlled tasks, learners are presented with contextualized phrases or sentences that are grammatically or lexically incorrect, and are asked to revise them using such a tool, while in uncontrolled tasks the learners often have difficulty identifying problems in sentences. In this regard, GSTs can be best used as a quick corrector of writing errors once those errors are marked in the task (e.g., Chambers & O'Sullivan, 2004; O'Sullivan & Chambers, 2006).

Research Question 3: How L2 Writers Used Quotation Marks as a GST for the Revision Process

The Korean EFL learners used the quotation marks as a GST to revise their first drafts as homework in the procedure of process-oriented writing. It was their first encounter with this search technique and with how to use it for articles, which most of them found extremely challenging in their writing. Interestingly, it turned out that many of the students attempted to use the quotation marks not only for confirming and correcting their errors (not just articles), but also for revising whole expressions or enriching their discussions with new information. Park's (2012) study also found that the three ESL learners used Google consultation for various purposes such as a simple error checker, or as a usage finder and style checker. It should be noted that their experiences with computers, learning styles, and motivation might affect their

use of GSTs for revision as well. Alternatively, this could be simply due to the novelty factor, where learners tend to be interested when they are taught a new tool.

Another noteworthy point is that learners at intermediate-level proficiency and up often used the GSTs for confirmation and revision rather than error correction. This supports findings from Yoon (2016) that six Korean ESL graduate students as advanced learners were more likely to consult a variety of DDL-based materials to confirm their existing knowledge rather than to find new linguistic items. The lack of information about low-level learners notwithstanding, a tentative conclusion is that more proficient learners with sufficient grammar and vocabulary knowledge tend to confirm their existing knowledge of their self-generated expressions, which might not be the case for lower-level learners. In this regard, this DDL approach to self-editing skills can promote not only inductive learning (through self-correction) but also deductive learning (through self-confirmation) in the course of becoming advanced writers (for a detailed discussion on inductive and deductive approaches in DDL, see Flowerdew, 2015).

Research Question 4: Student Perceptions and Attitudes toward GSTs

The post-treatment questionnaire revealed that all 18 students were generally satisfied with the GST training, as also found in previous Google consultation research (e.g., Geluso, 2013; Park, 2010; Yoon, 2016). Regardless of proficiency levels, the students with a fair amount of computer literacy reported that the GST training successfully helped their writing and that they would keep using the techniques in their future writing. Most importantly, their comments on the use of GSTs as a language reference tool demonstrate their increased confidence with independent English writing. Unlike using articles and collocations, however, paraphrasing with Google was extremely difficult for most of the students except for a few advanced learners.

Perhaps paraphrasing, covering both article and collocation usage, increasingly challenged such novice EFL writers with comparatively limited knowledge of varying grammatical structures and words. That is likely why many of them used Naver relatively more for paraphrasing than for articles or collocations. It can therefore be concluded, albeit tentatively, that the effectiveness of Google consultation (and other DDL tools) depends on the language points and thus on proficiency in L2 writing, which has been cited as a critical factor in DDL (Franken, 2014; Yoon & Jo, 2014).

Conclusion

The present study examined the effectiveness of GSTs in the use of articles, collocations, and paraphrasing, in an attempt to improve L2 writing in a Korean EFL college academic context. A 4-day GST training workshop was offered with a focus on article use, collocation, and paraphrasing, which were challenging to the 18 novice writers with varying levels of English proficiency.

The quantitative and qualitative data provide insights into the two most important pedagogical implications for teaching GSTs in L2 writing courses. First, detailed guides to Google consultation are crucial for maximizing learning outcomes from formulating search phrases, using different Google subcorpora, and interpreting search results of frequency and contextual information. Second, it is important to integrate Google consultation into the curriculum of L2 writing instruction. In particular, it is strongly recommended to show how students can autonomously consult Google for their weak writing points as well as for other useful DDL tools (mainly corpora and associated tools) or dictionaries.

More research is needed to deepen our knowledge of the role of Google consultation in L2 writing. Two noteworthy venues for future research, which are also limitations of the present study, can be suggested. First, longitudinal studies are needed to examine if and how long the effectiveness of Google-informed L2 writing instruction lasts. It will be interesting to see if learners, as they become more familiar with GSTs, use them for other types of language learning. Second, as advocated in previous research, it would be worth examining any differential effectiveness of Google-informed L2 writing instruction in terms of varying proficiency levels, types of tasks (receptive, productive, or both), or learning styles (Boulton, 2015; Yoon

& Jo, 2014).

Apparently, the key to successful Google-informed L2 writing lies in how teachers introduce GSTs to students. Ironically, this so-called discovery language learning, which relies heavily on learner autonomy, requires a series of training sessions to help L2 learners become independent writers. A consensus in the field of DDL research is to use guided discovery or guided induction rather than fully autonomous discovery (e.g., Smart, 2014)—which the present study confirms. Teachers' meticulous guidance and vigilant individualized feedback are likely to be crucial in creating the conditions for successful self-directed learning with Google.

Notes

1. The class had 21 students in total, one of whom was not a Korean native speaker. At the outset of the class, we described the purpose and procedure of the study, and the students understood that their participation in the research was voluntary and there would be no penalties or loss of benefits even if they did not participate.
2. Figure 1 also shows that in *conduct a research*, *research* is usually a modifier, as in *conduct a research study* or *conduct a research interview*.
3. The three subcorpora of Google were deliberately chosen instead of searching the entire web, because (a) the class focused on academic writing on social issues and (b) Google as a whole seems to retrieve more erroneous search results (e.g., from personal blogs).
4. For the take-home assignment, the students made a separate copy (Draft 2), and they underlined any revised expressions for which they used the quotation marks in Google. Four types of Google searches were identified by comparing Draft 1 and Draft 2.
5. *Correction* is any word-level modification of grammatical or lexical errors, whereas *revision* is any phrase- or sentence-level modification of erroneous or valid expressions.

References

Acar, A., Geluso, J., & Shiki, T. (2011). How can search engines improve your writing? *CALL-EJ*, *12*(1), 1−10. Retrieved from http://callej.org/journal/12-1/Acar_2011.pdf

BNC Consortium. (2007). *The British national corpus, version 3* (BNC XML Edition). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from http://www.natcorp.ox.ac.uk/

Boulton, A. (2010). Data- driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534−572.

Boulton, A. (2013). Wanted: Large corpus, simple software. No timewasters. In A. Leńko-Szymańska (Ed.), *TaLC10: 10th International Conference on Teaching and Language Corpora. July 2012*. Warsaw, Poland: Warsaw University Press.

Boulton, A. (2015). Applying data-driven learning to the web. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 267−295). Amsterdam, Netherlands: John Benjamins.

Brezina, V. (2012). Use of Google Scholar in corpus-driven EAP research. *Journal of English for Academic Purposes*, *11*, 319−331.

Chambers, A., & O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, *16*(1), 158–172.

Chapelle, C. A. (2003). *English language learning and technology: Lectures on teaching and research in the age of information and communication*. Amsterdam, Netherlands: John Benjamins.

Chinnery, G. (2008). You've got some GALL: Google-assisted language learning. *Language Learning & Technology*, *12*, 3−11. Retrieved from http://llt.msu.edu/vol12num1/net

Chujo, K., Oghigian, K., & Akasegawa, S. (2015). A corpus and grammatical browsing system for remedial EFL learners. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 109–128). Amsterdam, Netherlands: John Benjamins.

comScore. (2015). *US desktop search engine rankings. comScoreq Search™*. Retrieved from http://www.comscore.com/Insights/Market-Rankings/comScore-Releases-November-2015-US-Desktop-Search-Engine-Rankings

Conroy, M. (2010). Internet tools for language learning: University students taking control of their writing. *Australasian Journal of Educational Technology*, *26*(6), 861−882.

Davies, M. (2008). *The corpus of contemporary American English (COCA): 520 million words, 1990-present*. Retrieved from http://corpus.byu.edu/coca/

Davies, M. (2013a). *Corpus of news on the web (NOW): 3+ billion words from 20 countries, updated every day*. Retrieved from http://corpus.byu.edu/now/

Davies, M. (2013b). *Corpus of global web-based English: 1.9 billion words from speakers in 20 countries*. Retrieved from http://corpus.byu.edu/glowbe/

Davies, M. (2013c). Google Scholar and COCA-Academic: Two very different approaches to examining academic English. *Journal of English for Academic Purposes*, *12*(3), 155−165.

Davies, M. (2015). *The Wikipedia corpus: 4.6 million articles, 1.9 billion words. Adapted from Wikipedia*. Retrieved from http://corpus.byu.edu/wiki/

Egbert, J. (2005). *CALL essentials: Principles and practice in CALL classrooms*. Alexandria, VA: TESOL Publications.

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* (pp. 15−36). Amsterdam, Netherlands: John Benjamins.

Franken, M. (2014). The nature and scope of student search strategies in using a web derived corpus for writing. *Language Learning Journal*, *42*(1), 85−102.

Geiller, L. (2014). How EFL students can use Google to correct their 'untreatable' written errors. *EuroCALL Review*, *22*(2), 26−45.

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.

Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Computer learner corpora, second language acquisition, and foreign language teaching*. Amsterdam, Netherlands: John Benjamins.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, *4*, 1−16.

Kern, R. (2006). Perspectives on technology in learning and teaching languages. *TESOL Quarterly*, *40*(1), 183−210.

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics*, *29*(3), 333−347.

Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *Modern Language Journal*, *93*(1), 61−78.

Milton, J. (2006). Resource-rich web-based feedback: Helping learners become independent writers. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 123−137). Cambridge, UK: Cambridge University Press.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam, Netherlands: John Benjamins.

O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, *15*(1), 49–68.

Panah, E., Yunus, M. M., & Embi, M. A. (2013). Google-informed pattern-hunting and pattern-defining: Implication for language pedagogy. *Asian Social Science*, *9*(3), 229−238.

Park, K. (2010). *Using a specialized corpus and Google custom search engine (CSE) for enhancing L2 teaching and learning*. Retrieved from http://www.personal.psu.edu/xxl13/teaching/au10/588/park_2010.pdf

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge. UK: Cambridge University Press.

Robb, T. (2003). Google as a corpus tool? *TESL-EJ*, *7*(2), 1–4. Retrieved from http://www.tesl-ej.org/wordpress/issues/volume7/ej26/ej26int/

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.

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.

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

Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System*, *29*, 91–102.

Wu, S., Franken, M., & Witten, I. H. (2009). Refining the use of the web (and the web search) as a language teaching and learning resource. *Computer Assisted Language Learning*, *22*(3), 83–102.

Yoon, C. (2016). Concordancers and dictionaries as problem-solving tools for ESL academic writing. *Language Learning & Technology*, *20*(1), 209–229. Retrieved from http://llt.msu.edu/issues/february2016/yoon.pdf

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2. *Journal of Second Language Writing*, *13*(4), 257–283.

Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology*, *18*(1), 96–117. Retrieved from http://llt.msu.edu/issues/february2014/yoonjo.pdf

Appendix A. Sample Questions from Practice Exercises

The following search results came from using Google Books.

GST Session 1: Articles

Part A: Error Correction

For example, copying other people's ideas or expressions without source citations is an academic dishonesty.
My first answer(s): academic dishonesty

Search phrases	Search Results (Hit numbers / Context)
1. an academic dishonesty	1. NA (Not Applied)
2. academic dishonesty	2. NA
3. "an academic dishonesty"	3. 8 / ~ statement, policy, problem
4. "academic dishonesty"	4. 18,880 / academic dishonesty

My final answer(s): academic dishonesty

Part B: Korean-to-English Translation

[Korean sentence provided]
My first answer(s): The second most important way is to help students study.

Search phrases	Search Results (Hit numbers / Context)
1. second most	1. NA
2. "a second most"	2. 35,600 / into a second. Most ~
3. "the second most"	3. 379,000 / the second most

My final answer(s): The second most important way is to help students study.

GST Session 2: Collocations

Part A: Error Correction

I will have research on English learning.
My first answer(s): carry out research

Search phrases	Search Results (Hit numbers / Context)
1. will have research	1. NA
2. "will have research"	2. 8,060 / ~ procedures, assistants, interests
3. "will * research"	3. 552,000 / do, undertake, conduct

My final answer(s): conduct research, carry out research

Part B: Korean-to-English Translation

[Korean sentence provided]
My first answer(s): College students search much information in Google.

Search phrases	Search Results (Hit numbers / Context)
1. "college students"	1. 602,000 / every , a , his ~ /~ learning, data
2. "search * Google"	2. 7 / through ~ (Google News: 31,100,000 / about, on, with)
3. "search on Google"	3. 327,000 / search on Google

My final answer(s): College students search on Google very much.

GST Session 3: Paraphrasing

I still need to improve grammar skills as well as my writing skills.
My first answer(s): It is necessary to improve not only writing skills but also grammar skills.

Search phrases	Search Results (Hit numbers / Context)
1. improve grammar skills	1. NA
2. "to * grammar skills"	2. 54,200 / learn, develop, improve, practice, advance, polish + their
3. "need improvement"	3. 185,000 / need improvement
4. "need an improvement"	4. 5,410 / need an improvement, need an improvement factor

My final answer(s): It is necessary to improve not only writing skills but also grammar skills.
I still should develop not only my writing skills but also grammar skills.
My grammar and writing skills need improvement.

Appendix B. Sample Questions and Answers from Pre-test and Post-test 1

Pre-Test 1: Articles

Directions: To complete Parts A and B

- Give your first answer, your final answer, and searched words/expressions
- If necessary, consult Naver online dictionary only
- Confirm or revise your answer

Part A: Identify and underline one unnatural expression and correct it appropriately

Without any source, the writing cannot be regarded as a true information.

- Your first answer: true information
- Your final answer: true information
- Searched words/expressions: any / a true information

Part B: Translate each Korean sentence into English

[Korean sentence provided]

- Your first answer: *Korean government should try to reduce the air pollution in China.*
- Your final answer: *Korean government should try to reduce the air pollution in China.*
- Searched words/expressions: *the air pollution*

Post-Test 1: Articles

Directions: To complete Parts A and B

- Give your first answer, your final answer, and searched words/expressions
- Use both Naver online dictionary and the three corpora including Google Books, Google Scholar, and Google News if necessary
- Confirm or revise your answer by using **the quotation marks** when you use any Google subcorpora

Part A: Identify and underline one unnatural expression and correct it appropriately.

For example, copying other people's ideas or expressions without citing sources is an academic dishonesty.

- Your first answer: *is academic dishonesty*
- Your final answer: *is academic dishonesty*
- Searched words/expressions: "*is an academic dishonesty"* / *"is academic dishonesty"*

Part B: Translate each Korean sentence into English

[Korean sentence provided]

- Your first answer: *We all have to try hard to stop the global warming.*
- Your final answer: *We all have to try hard to stop global warming.*
- Searched words/expressions: *"stop global warming"* / *"stop the global warming"*

Appendix C. Sample Questions and Answers from Pre-test and Post-test 2

Pre-Test 2: Collocations

Directions: To complete Parts A and B

- Give your first answer, your final answer, and searched words/expressions
- If necessary, consult Naver online dictionary only
- Confirm or revise your answer

Part A: Read each sentence with one underlined somewhat unnatural expression. Change or delete one or two words to make each sentence sound more natural. Add, delete, or change one or two words in each underlined expression.

I hope that when I die, people remember me as a good contributor of pedagogy.

- Your first answer: *a good contributor for*
- Your final answer: *a good contributor for*
- Searched words/expressions: *pedagogy*, *contributor for*, *contributor of*

Part B: Translate each Korean sentence into English

[Korean sentence provided]

- Your first answer: *I would like to spend vacation reading many books.*
- Your final answer: *I would like to spend vacation in reading many books.*
- Searched words/expressions: *spend vacation*, *spend on*, *spend in*

Post-Test 2: Collocations

Directions: To complete Parts A and B

- Give your first answer, your final answer, and searched words/expressions
- Use both Naver online dictionary and the three corpora including Google Books, Google Scholar, and Google News if necessary
- Confirm or revise your answer with and without **the quotation marks** when you use any Google subcorpora
- Use *** (wild card)** if necessary

Part A: Read each sentence with one underlined somewhat unnatural expression and revise it to make each sentence sound more natural. Add, delete, or change one or two words in each underlined expression.

In the future, I would love to participate on the voluntary work of helping the poor.

- Your first answer: *participate in*
- Your final answer: *participate in*
- Searched words/expressions: *"participate in", "participate on"*

Part B: Translate each Korean sentence into English

[Korean sentence provided]

- Your first answer: *University students in Korea search for information in Naver.*
- Your final answer: *University students in Korea search for information mainly using Naver.*
- Searched words/expressions: *"search for", "search for * information", "search on", "search in"*

Appendix D. Sample Questions and Answers from Pre-test and Post-test 3

Pre-Test 3: Paraphrasing

Directions: Read each sentence carefully and paraphrase it accordingly.

- Give your first answer, your final answer, and searched words/expressions
- If necessary, consult Naver online dictionary only
- Confirm or revise your answer

The belief that the quality of education cannot exceed the quality of teachers indicates how important the role of individual teachers is in education.

- Your first answer: *People believe that the quality of education cannot exceed ability of teachers. It shows that teachers are important in education.*
- Your final answer: *People believe that the quality of teaching cannot exceed the ability of teachers. It shows that teachers are important in education.*
- Searched words/expressions: *quality, ability*

When I go to college I will have different opportunities to do volunteer work during my free time.

- Your first answer: *I will be able to do more volunteer activities when I am a college student.*
- Your final answer: *I will be able to do more volunteer activities when I become a college student.*
- Searched words/expressions: *volunteer work*

Post-Test 3: Paraphrasing

Directions: Read each sentence carefully and paraphrase it accordingly.

- Give your first answer, your final answer, and searched words/expressions
- Use both Naver online dictionary and the three corpora including Google Books, Google Scholar, and Google News if necessary
- Confirm or revise your answer with and without **the quotation marks** when you use any Google subcorpora
- Use *** (wild card)** if necessary

Studies have shown that around 2050 India will have the highest growth of economy because of their number of population, followed by China.

- Your first answer: *According to research, India will grow economically fast than any other countries around 2050, because it has a huge population, followed by China.*
- Your final answer: *According to research, the economy of India will grow fast than any other countries around 2050, because it has a huge population, followed by China.*
- Searched words/expressions: *"to the research", "economy of", "a * population"*

When I study abroad, I would like to work for other people in need.

- Your first answer: *I willing to help other people in trouble when studying abroad.*
- Your final answer: *I am willing to support other people in trouble when studying overseas.*
- Searched words/expressions: *work for, abroad, "willing to"*

Appendix E. A Coding Sheet for Pre-test and Post-test 3

Category	Questions	Points
Grammar/Vocabulary Appropriateness	Is the paraphrased sentence grammatically and lexically appropriate?	
	• Inappropriate	0
	• Appropriate	1
Maintenance of the Same Meaning	How much does the paraphrased sentence maintain the same meaning of the original sentence?	
	• Unsuccessfully	0
	• Somewhat	1
	• Successfully	2
Paraphrasing Types	Which linguistic part of the original sentence was mainly paraphrased?	
	• Grammatical structure	1
	• Vocabulary	2
	• Both grammatical structure and vocabulary	3

About the Authors

Sumi Han received her PhD in applied linguistics from Northern Arizona University. She is currently an assistant professor at Hallym University and a member of Digital Humanities Collaboratory at Hallym. Her

research interests include corpus linguistics, computational linguistics, and language testing.

E-mail: sumihan@hallym.ac.kr

Jeong-Ah Shin received her PhD in educational psychology from the University of Illinois at Urbana-Champaign. She is an assistant professor in the Department of English at Dongguk University in Korea. Her fields of research center on second language learning and processing by employing psycholinguistic and corpus linguistic methods.

E-mail: jashin@dongguk.edu, jashin@gmail.com