

Applying Corpus-Based Grammar Instruction: Its Retention Effect on Students' Grammatical Awareness and Writing Accuracy

Tsegaye Mekasha¹, Dawit Amogne²

Email: tsegaye.mekasha@kiot.edu.et

¹ PhD candidate in TEFL, Department of English Language and Literature, Wollo University, Wollo, Ethiopia

² Addis Ababa University, Addis Ababa, Ethiopia
Dessie, Amhara Region, Ethiopia

ARTICLE INFO

Article history:

Received: 21/06/2025

Revised: 18/07/2025

Accepted: 05/08/2025

Available online: 20/08/2025

Keywords:

Corpus-based
grammar instruction;
Grammatical awareness;
Writing accuracy.

ABSTRACT

This study examined the retention effect of corpus-based grammar instruction on students' grammatical awareness and writing accuracy and explored their perceptions toward the instruction. The study employed mixed-method research and a quasi-experimental research design. Two intact classes, 86 students, were selected using a convenient sampling technique from first-year students in 2024. The experimental group students ($n=42$) were taught by corpus-based grammar instruction and the comparison group students ($n=44$) were taught conventionally. Moreover, data were gathered through tests and interviews. Descriptive and inferential statistics were computed to analyze quantitative data whereas thematic analysis to qualitative data. The delayed grammar and writing post-test results indicated that corpus-based instruction had a significant retention effect on students' grammatical awareness ($\eta^2p = 0.491, p < .05$) and writing accuracy ($\eta^2p = 0.378, p < 0.05$). Overall, corpus-based grammar instruction had a statistically significant retention effect ($\eta^2p = 0.576, p < 0.05$). Finally, the qualitative data proved that students who received corpus-based grammar instruction had a positive attitude toward the instruction. Therefore, EFL teachers should apply corpus-based grammar instruction to enhance students' grammatical awareness and writing accuracy. Further research should also explore the retention effects of the instruction and learners' perception across different language skills.

© 2025 L'Geneus. All rights reserved.

1. Introduction

Grammar plays a paramount role in EFL classrooms. However, the way it is taught is still conventional. It always starts with a mechanical explanation of the rule before moving on to exercises and activities (Hedge, 2000; Richards and Rogers, 2014). This kind of rule presentation in language learning and teaching has significant downsides- promoting rote memorization of grammar rules without contextual understanding (Ellis, 2002; Larsen-Freeman, 2003) that results in learners' inability to apply rules meaningfully in authentic communication situations (Schmidt, 2001; Nunan, 2004), and contributing to decreased learner engagement and motivation during the lesson.

The approach also neglects natural language use patterns, leading to compartmentalized language learning experiences (Skehan, 1998). Therefore, it must be altered because grammar is recognized as a fundamental and indispensable aspect of language teaching (Burgess and Etherington, 2002), and it has to establish itself in the proper place in language teaching and learning (Richards & Rogers, 2001). Thus, to enhance students' grammatical ability and writing skills, teachers should apply more integrated and communicative approaches that connect grammar instruction with authentic language use. As a result, this marks a paradigm

shift in how language teachers view the use of technology as an approach to teaching grammar, and the need for innovative approaches to grammar instruction that align with technological advancements (Larsen-Freeman & Anderson, 2011; Brown, 2016).

Among the different technology-based approaches to teaching language skills (including grammar), the corpus-based approach is the most potent, a topic of study since the late 1980s, with Tim John's work on data-driven learning (DDL) being one of the first to explore corpora for teaching purposes (Gavioli & Aston, 2001; Braun, 2007). Since then, researchers have studied the pedagogical implications of using corpora in language teaching (O'Keeffe *et al.*, 2007; Flowerdew, 2009; Römer, 2011; Leech, 2013).

The corpus is the computerized storage of text collections (McEnery & Wilson, 2001; O'Keeffe *et al.*, 2007; McEnery & Hardie, 2012; Timmis, 2015; Friginal, 2018), which are of natural texts that represent a variety or genre of a language (Sinclair, 1991; McEnery & Wilson, 2001), which can be either written or spoken, or both. Granger (2003) and Flowerdew (2012) also define corpora as systematic collections of learner language data comparing usage with native or expert usage and electronic collections of authentic texts produced by foreign or second language learners for different purposes.

As a huge collection of texts, the corpus has multiple roles and purposes in ELT, particularly in grammar classes. One of the most significant benefits of the corpus is that it allows students to explain language processes by themselves. It provides systematic access to naturally occurring language, encouraging autonomous learning (Bernardini, 2004; Braun, 2005), and offers hands-on experience, adopting an inductive approach. Thus, using corpus in ELT, typically in grammar, promotes a more autonomous and engaging learning experience (Johns, 2002; O'Keeffe *et al.*, 2007).

Grammar teaching, using corpora is, in fact, a current phenomenon that has been started since the late 1980s, has grown immensely, and has revolutionized English language studies with its authentic appeal (Conrad, 2000). Using the corpus in grammar teaching allows students to focus on the specific areas of grammar that they struggle with (Conrad, 2000; O'Keeffe *et al.*, 2007). This also helps to bridge the gap between classroom grammar teaching and real-world English language use. Overall, corpus-based grammar instruction using corpus has opened up new avenues for ELT that were previously unavailable, targeting the provision of a ready-to-use resource of natural or authentic texts (Reppen, 2010).

Therefore, it is believed that corpus-based grammar instruction highly affects grammar teaching in many ways due to its potential to provide learners with rich data on authentic and accurate language use instead of invented usages (McCarthy & McCarten, 2022). It provides learners with register-specific descriptions, enabling them to appropriately use the English language (i.e., grammar) in spoken and written communication (Conrad, 2000). This resulted in much corpus incorporation into grammar teaching (Johansson, 2009).

As research findings indicate, this incorporation would greatly benefit English language learners and directly affect their grammatical ability (Lin, 2016). Likewise, Sinclair (2013) advocates that teachers should, whenever possible, provide learners with real examples of language use, and for that reason, a corpus would be highly beneficial for both learners and teachers. Sinclair states that access to these patterns is considerably assisted by using corpus tools since learners can analyze real and more extended samples of English language, facilitating their understanding of how words co-occur and how these language units convey meaning in context.

As mentioned in some references, in ELT, teachers and students can apply corpus-based grammar instruction in two different ways: directly by using the online corpora platform and indirectly with corpus-informed instructional tools (Braun, 2005; O'Keeffe *et al.*, 2007; McEnery & Xiao, 2010; Leech, 2013). The former, also known as DDL, can be used directly in ELT classrooms (Romer, 2011; Timmis, 2015) since the corpus presents language patterns that

would be perfectly relevant in grammar classes. For instance, in learning relative clauses, learners can be exposed to many examples of sentences with relative clauses from a corpus, and they search for patterns, recognize, analyze, and manipulate them producing similar patterns based on those language data. Here, teachers facilitate and guide students to achieve the desired goals of the lesson. On the other hand, the latter refers to employing teacher-made English language teaching materials, dictionaries, reference grammar books, course books, etc., that use corpora as linguistic sources from which examples of language patterns are drawn for immediate actual classroom use (McEnery & Xiao, 2010; Leech, 2013; Meunier & Reppen, 2015).

However, some scholars argue that corpus-based grammar instruction may not be practical for different reasons. For instance, Lin and Lee (2015) reported some challenges, including technical difficulties, issues in designing and conducting activities such as concordance analysis, and increased teacher workload. McEnery and Xiao (2010) also mention the learners' level and experience, the knowledge and skills required for corpus analysis, and online access to use the corpus as challenges.

Besides these arguments, corpus-based grammar instruction has theoretical foundations. It relies on the usage-based theory that claims language emerges from patterns and regularities in language use and exposure (Robinson & Ellis, 2008) and aims to ground language structure in the usage event, that is, the actual instances of language (Kemmer & Barlow, 2000). Usage-based approaches to language acquisition hold that language constructions emerge because of experiences that language users have had (Ellis *et al.*, 2012). Likewise, CBGI relies on analyzing and practicing grammar patterns through actual contexts used in different registers.

The input and noticing hypotheses are also the basis for corpus-based approach. The noticing hypothesis suggests that learners cannot advance their language abilities without consciously noticing them. Accordingly, the noticing theory proposes that learners consciously notice language features and patterns to internalize them (Schmidt, 1990). Corpus-based instruction facilitates noticing by bringing learners' attention to language patterns, frequencies, and variations. Through corpus-based activities, learners are encouraged to actively engage with the language and notice its recurring patterns.

Similarly, CBGI employs authentic resources to make the teaching and learning of grammar more accurate and understandable for learners. In contrast, the input hypothesis guides its material selection and presentation from corpora to make it meaningful and understandable for ELT classrooms. In this regard, Krashen (1982) proposed that language acquisition takes place when students are exposed to comprehensible input that is just above their current level of understanding. In such instances, corpus data can provide real and diverse language inputs, allowing learners to notice natural language patterns and structures in various circumstances. Moreover, the approach relies on the Output Hypothesis through integrating authentic language input with possible outcomes for meaningful language production. Swain's (1993) Output Hypothesis stresses that producing language helps learners identify gaps in their knowledge, which improves their learning. Similarly, CBA facilitates this process by delivering learners with real-world language data from corpora, allowing them to detect patterns like collocations and grammatical structures and apply them to their output- speaking and writing.

Apart from the above-and-often-cited theories supporting corpora, corpus linguistics is the most noticeably related and important arena. In fact, corpus linguistics is the empirical study of language using computer-assisted techniques to analyze large databases of naturally occurring language (Conrad, 2000), on which corpus-based grammar instruction extensively relies.

Considering these theoretical supports, what has triggered us to conduct a study on applying CBA in EFL classrooms is learners' poor grammatical awareness. Our observation is congruent with the various research findings. Researchers have revealed that students' grammatical awareness is poor. For instance, studies conducted in the Arabic-speaking community revealed that one of the most critical issues in learning ESL by Arab students is mastering English grammar (Azar, 2007; Fareh, 2010; Khan, 2011). These studies concluded that learners found grammar so complicated because there was too limited opportunity for them to practice English outside the classroom. Fareh (2010) identifies activities centered on a teacher rather than on the students; focus on rote learning rather than on developing skills; inappropriate textbooks and other teaching materials; and insufficient exposure to English as major obstacles to proper mastery of English grammar. Marzulina *et al.*, (2019), also conducted a study on student-teachers in Indonesia and found that the student-teachers had a poor ability to explain grammatical errors.

Moreover, studies confirm that students' writing skills have faced multifaceted problems due to a lack of grammar and vocabulary. According to Abiy's (2013) investigation into high school students' writing skills and English language proficiency as indicators of their performance in English language writing, EFL students greatly demanded grammar and vocabulary. Moreover, Alamrew (2005) and Haregewain (2008) conducted studies related to writing skills. They found that students are not able to write a single grammatically correct sentence at the secondary and tertiary education levels which was later confirmed by Asres (2014) that the problem is getting worse from time to time. By and large, in Ethiopia, many writing teachers in EFL classrooms are seen teaching writing and spending much time correcting students' grammatical errors (Dawit, 2014). Consequently, it is possible to say that good knowledge of grammar leads to better writing, which offers confidence and credibility.

Owing to its significant role in ELT, considering the Ethiopian higher education context in which English is taught as a foreign language, it is important to notice a CBGI's role in enhancing learners' grammatical awareness and writing accuracy. This gives the impression that corpus-based grammar instruction has become increasingly popular and is recognized as an effective instruction to enhance learners' grammatical awareness and writing skills. However, the previous studies on corpus-based instruction investigated the effect on grammar or writing in particular and learners' attitudes towards the instruction. In addition, little attention has been given to examining the retention effect of CBGI on both students' grammatical awareness and writing accuracy simultaneously. Moreover, the worsened and sophisticated grammatical errors observed in students' writing and the need for innovative approaches to grammar instruction in this era of digitalization to improve these problems calls great attention. Therefore, the study examined the retention effect of corpus-based grammar instruction on students' grammatical awareness and writing accuracy. To this end, the following research questions were formulated. Does corpus-based grammar instruction have retention effect on students' grammatical awareness? What is the retention effect of corpus-based grammar instruction on students' writing accuracy? What is the combined retention effect of corpus-based instruction on students; grammatical awareness and writing accuracy? How do students perceive learning grammar and writing through corpus-based instruction?

2. Method

2.1 Research design

A pragmatic approach to the research paradigm and a quasi-experimental, non-equivalent control group design were employed to meet the research objectives. These involved two intact groups: one experimental group receiving a pretest, an intervention in the form of experimental treatment, a post-test, and a delayed post-test; the other, the control group, was subjected to only the pretest, post-test, and delayed post-test. This research design

was chosen as two naturally existing intact classes were taken to examine the effect of the independent variable(s)- the conventional approach and the corpus-based grammar instruction- over the dependent variables, which were the effects of corpus-based grammar instruction on students' grammatical awareness and writing accuracy, and their view toward the instruction. Moreover, the mixed-methods research approach with an explanatory-sequential design was used. In the data collection and analysis procedure, the quantitative data were collected and analyzed first, followed by the qualitative data.

2.2 Participants, samples, and sampling technique of the study

This study was conducted at Wollo University, on first-year undergraduate students placed at Kombolcha Institute of Technology in the 2024 academic year. Due to the quasi-experimental nature of this study, only two intact sections were conveniently chosen as the experimental and control groups. The study comprised 86 students ($n = 42$, experimental; and $n = 44$, control). A pretest was administered to both to determine the homogeneity of the participants. Moreover, Levene's test was computed to scientifically measure groups' homogeneity. Consequently, Levene's test results showed that the variance in the scores for the grammar and writing pretests were $p = 0.392$, and $p = 0.373$ ($p \geq 0.05$), respectively, suggesting that there was no significant difference in variances between the groups and that they were homogenous.

2.3 Instruments

We used quantitative and qualitative data-gathering instruments. Grammar and writing tests were used to gather quantitative data, while reflective journals and interviews were employed to gather data qualitatively. Tests were used as the main tools for data collection. Accordingly, different grammar and writing tests were employed: a pre-test, a post-test, and a delayed post-test for various purposes. The grammar tests were adapted from TOEFL preparation materials (i.e., simple past and present perfect tense, conditionals and relative clauses), modified, and reviewed by TEFL experts. The revision helped to ensure the reliability and validity of the tests. Moreover, the reliability of the grammar test was acceptable ($\alpha = 0.703$). The ICC for the writing test was (0.85) acceptable since Cronbach's alpha, and the interclass correlation coefficient value over 0.70 is acceptable (Muijs, 2022). These tests were administered to measure the participants' grammatical awareness and writing accuracy, observe their progress in grammar learning, and, most fundamentally evaluate the retention effect of the corpus-based grammar instruction.

In addition, reflective journals and semi-structured interviews were conducted. The reflective journals were distributed and collected throughout the intervention, while interviews were held with four randomly selected students from the experimental group at the end of the treatment. We used these instruments to gather qualitative data to answer research question number three related to students' perceptions towards the instruction.

2.4 Methods of data analysis and procedures

We utilized quantitative and qualitative methods of data analysis. The quantitative data were analyzed using SPSS version 25.0, with descriptive statistics, such as percentages, mean scores, and standard deviations. These statistics would offer a good understanding of the central tendencies of the data and the comparative analysis across multiple dependent variables. Moreover, a MANOVA was used since there were two dependent variables: grammatical awareness and writing accuracy, and independent variables: a corpus-based approach to teaching grammar (experimental group) and the conventional method (control group). It was used to observe the statistically significant effects of the independent variables on the dependent variables. Additionally, assumption tests were computed to ensure that the data had no problems with normal distribution, multicollinearity, outliers, and homogeneity variance-covariance matrices.

Furthermore, thematic analysis was incorporated to analyze the qualitative data from reflective journals and interviews. The data from the reflective journals were categorized into themes and then analyzed thematically. Likewise, data gathered using interviews was recorded, transcribed verbatim, and then analyzed thematically allowing for a detailed exploration and an in-depth understanding of the data. To sum up, initially, quantitative data were gathered and analyzed followed by qualitative data analysis.

2.5 The Intervention

Students in the experimental and comparison groups received instruction on the same grammar and writing lessons for an equivalent period. The grammatical concepts covered during the experimental phase included simple past tense, present perfect tense, conditional sentences, and relative clauses. These items were selected based on the preliminary investigation conducted prior to the study through grammar and writing tests. Then, students were taught each grammatical item, with the expectation of writing sentences and paragraphs on given topics. The following intervention procedures were implemented. These procedures were adapted from Ma et al. (2021) and Reppen (2010).

Phase 1: The Preparation Stage: This stage comprised three main tasks and lasted for seven hours. The first task was a pretest that assessed learners' existing knowledge of grammar through grammar and writing tests. This was done to collect baseline data regarding the students' grammatical awareness and their level of writing accuracy. The second task incorporated introducing the grammar items that the intervention would cover, setting ground rules, and agreeing on the schedule and importance of the intervention. The third task was exploring COCA with model examples.

Phase 2: The Intervention Stage: students were provided with sentences or texts extracted from COCA that contained similar grammatical structures. Working in pairs and groups, students identified the target structures, patterns, and language usage variations. Students were supposed to conduct concordance analysis, pattern recognition, error analysis, sentence correction, compare and contrast, and finally apply the knowledge of grammar they learned in writing different sentences and paragraphs.

Sample grammar lesson from COCA online interface:

Activity 1

1. The following sentences are taken from the COCA interface. Read each sentence and discuss in groups what the sentences have in common. Identify the relative pronouns and the relative clauses used in each sentence.
2. Jalen **John, who** transferred from Aloha to Lakeridge this year, also was a member of that FBU.
3. **Sarah, whom** I trust as a truthful person, says she saw Fred in the classroom near.
4. Our teachers are not to become intimately involved with **a student whom** they are teaching.
5. In addition, I wanted to point out **the people who** fidget and move around constantly... I'm guilty of many ticks and neurotic behavior.

6. Now, behind me is St. Patrick's **Cathedral, which** in some ways has been getting ready for this for the last three years.

Figure1: concordance lines and KWIC for relative pronouns 'whom' and 'which'

SEARCH	FREQUENCY	CONTEXT	ANALYZE TEXT
1 2013 ACAD CollegeSoud	22.9%	included a sample size of 6,828 students of which reported their country of origin as China. When looking at	
2 2019 MAG The Verge		has a matte finish, not that like these days , which acted to minimize glare. The display has a thin bezel on	
3 2015 TV Castle		mean not getting into the "right" elementary school , which affects the school , which affects college, and, oh,	
4 2015 ACAD ReadingImprovement		refers to students' self-perceptions and that which allows students (a) attempt different activities, (b	
5 2012 WEB accesstoinight.org		place, The 37 wings to Awakening (bodhi-pakshya-dhamma) -- which are the Buddha 's very heart that he gave to us	
6 2014 NEWS OrangeCR		small commercial drones should be treated as model aircraft , which are regulated by 1981 guidelines. The FAA has appealed the	
7 1991 MAG Horticulture		the leaf surface instead of beading up or rolling off , which can be a problem with waxy-leaved plants such as cabbage and	
8 2009 MAG MotherEarth		a few days of shade after they are set out , which can be easily provided by placing an old sheet over the row	

1. Now, login into your COCA account and search for the following relative pronouns as indicated in the above figures (who, whose, where, and that)

Phase 3: Final Stage: it was the stage in which learners' progress was assessed by providing a grammar and writing post-tests. Moreover, interviews were conducted to gather data on students' perceptions toward the instruction. Finally, delayed post-tests were administered three weeks after the intervention was completed.

3. Results and Discussion

These sections present results of students' grammar and writing post and delayed posttests.

3.1 The Retention Effects of CBGI on Students' Grammatical Awareness and Writing Accuracy

Table 1: Comparison of the two groups' post and delayed posttests

Descriptive Statistics						
Variables	Groups	N	Posttest		Delayed posttest	
			Mean	Std. Deviation	Mean	Std. Deviation
Grammatical Awareness	Experimental	42	29.6429	3.71432	30.7143	2.77871
	Control	44	23.2500	3.84662	25.0227	3.06889
Writing Accuracy	Experimental	42	14.393	1.6876	14.9048	1.79171
	Control	44	10.170	1.2847	11.8636	2.13052

As demonstrated in Table 1 above, the mean scores of the control group (conventional method) on the grammar posttest ($M = 23.25$) are significantly lower than those of the experimental group (corpus-based grammar instruction, $M = 29.64$). This indicates that the corpus-based grammar instruction intervention positively affected students' grammatical awareness. Similarly, the mean score on the grammar delayed posttest for the experimental group ($M = 30.71$) is significantly higher than the mean score for the control group ($M = 25.02$). The mean difference between the two groups is $M = 5.69$, which indicates that the experimental group performed better than the control group on the grammar-delayed posttest. The data depicts that corpus-based grammar instruction had a significant retention effect on the students' grammatical awareness.

Regarding students' writing accuracy, the descriptive statistics show significant differences in writing post-test results between the experimental and control groups. The experimental group had a higher mean writing post-test score ($M = 14.39$), with a standard deviation ($SD = 1.687$), compared to the control group's mean ($M = 10.17$) and standard deviation ($SD = 1.284$). This shows that corpus-based grammar instruction increased students' writing accuracy more effectively.

Likewise, students who received corpus-based grammar instruction ($M = 14.90$) outperformed those taught with the conventional methods ($M = 11.86$) on the delayed writing post-test. The mean difference between the two groups is ($M = 3.04$), which indicates that the experimental group performed better than the control group on the writing-delayed posttest. Although the descriptive statistics in Table 1 revealed that corpus-based grammar instruction had a retention effect on students' writing accuracy, multivariate analysis should further support it to determine if the impact is significant. Therefore, the level of significance is demonstrated in the following tables:

Table 2: Multivariate Analysis Result for Instructions on the two groups' delayed posttests

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Squared	Eta Observed Power ^b
Instructions Pillai's Trace	.576	56.329 ^a	2.000	83.000	.000	.576	1.000

Table 2 presents the results of a multivariate analysis of variance with two dependent variables (grammar and writing) and independent variables (corpus-based grammar instruction and conventional method). The results indicate a significant multivariate effect for instruction types (Pillai's Trace = 0.576, $F(2, 83) = 56.329$, $p = .000$). The Pillai's Trace value for instruction is 0.576, which indicates a large effect size. This means the instruction accounts for 57.6% of the variance in the dependent variables, and it indicates that CBGI had a significant impact on long-term students' grammatical awareness and writing accuracy.

Table 3: Parameter estimates of the two groups' delayed posttests

Dependent Variable	Parameter	B	Std. Error	T	Sig.	95% Confidence Interval		Partial Squared	Eta
						Lower Bound	Upper Bound		
Grammar Delayed posttest	[Experimental]	5.692	.632	9.002	.000	4.434	6.949	.491	
	[Control]	0 ^b	.	.	.	.	.	.	
Writing Delayed Posttest	[Experimental]	3.041	.426	7.147	.000	2.195	3.887	.378	
	[Control]	0 ^b	.	.	.	.	.	.	

Table 3 shows the results of two separate regression analyses, with grammar-delayed post-test scores and writing-delayed post-test scores as the dependent variables, are presented, respectively. For the grammar delayed posttest, the results indicate that the CBGI indicator variable ($\beta = 5.692$, $t(85) = 9.002$, $p < .05$) is a significant predictor of the grammar delayed posttest score. Similarly, the partial eta squared value ($\eta^2 p = 0.491$) indicates that the effect of group accounts for 49.1% of the variation in grammar-delayed post-test scores. This suggests that students who received CBGI scored significantly higher on the grammar-delayed posttest than students who received the conventional method. Similarly, for the delayed writing posttest, the results indicate that the CBGI indicator variable ($\beta = 3.041$, $t(85) = 7.147$, $p < .05$) is also a significant predictor of the delayed writing posttest score. In the same way, the partial eta squared value ($\eta^2 p = 0.378$) indicates that the effect of the group accounts for 37.8% of the variation in delayed post-test scores. The data revealed that students who were instructed with CBGI significantly outperformed their counterparts on the delayed posttest. Therefore, corpus-based grammar instruction had long-term, significant effects on students' grammatical awareness and writing accuracy.

3.2 Qualitative Data Results and Analysis

3.2.1 Data from the students' reflective journals and interviewees

This section attempts to summarize the important findings obtained from the students' reflective journals and interviewees in the study. The main purpose of these tools was to explore students' perceptions of corpus-based instruction applied to learning grammar in EFL classes. Data obtained from the reflective journals and interviewees were organized into key themes for analysis as presented below.

3.2.2 The effectiveness of CBI in improving grammatical skills and awareness:

Data from students' reflective journals indicated that CBGI helped them understand and internalize complex grammatical structures from the concordance lines. Most students appreciated the depth of understanding they gained through the grammatical structures, examples, and patterns provided by COCA. The following excerpts were taken from students' reflective journals:

S2: *"COCA helps me to know how to learn grammatical structures by input words or phrases and teaches me to expand my understanding about English grammar."*

S8: *"Before I used COCA I learned only by teachers. My understanding was limited but now by using it I learned more and expand my understanding about grammar."*

S14: *"Yes, COCA is helpful because it has many grammatical structures to show different examples and explanations on any grammar."*

These excerpts depicted the effectiveness of COCA in enhancing learners' grammatical awareness. Moreover, the interviewees' transcripts also substantiated this data revealing that the students found the instruction helpful in learning grammar. They replied that they had ample examples from diverse contexts for a single grammar structure, which they could not get from their course modules.

3.2.3 Improved writing:

Data obtained through reflective journals and interviewees showed that COCA helped students improve their writing ability. To consolidate it, some excerpts from the reflective journals are presented below.

S3: *"If we know the grammatical structures, we will be able to improve our writing skills, so COCA has multiple purposes."*

S6: *"COCA is helpful because the grammar contexts that we get from it are a collection of expressions that are helpful for our writing skill development."*

S11: *"Writing is not easy for many students and for me, it is difficult to write even sentences but now I learn and read a lot on COCA and understand how to write error free paragraphs."*

Further, data gathered through interviews confirmed that the exposure to and understanding of different grammatical structures enabled them to improve their writing.

3.2.4 Motivation and engagement:

Regarding students' level of motivation and engagement during the intervention, students evidenced in their reflective journals and interviews that they felt exceptionally motivated and inspired. They reported that the hands-on activities allowed them to observe and identify the practical application of grammatical structures (i.e., simple past, present perfect, conditional sentences, and contexts for relative clauses) in real-life examples, which made the learning process engaging; as a result, they liked the instruction

To sum up, students expressed how the interactive nature of COCA motivated them and enhanced their engagement with grammar learning. Most students reported enjoying the hands-on experience, though some initially faced technical challenges, such as account setup and navigating the platform. However, these obstacles diminished as they became more familiar with COCA. The interviewees collectively agreed that accessing examples of grammar

in authentic contexts such as simple past, present perfect, and relative clauses made learning both enjoyable and enriching. Ultimately, students valued the practical, context-rich learning experience provided by CBGI and viewed it as a transformative approach to mastering English grammar.

Discussion

The main purpose of this study was to examine the retention effect of corpus-based grammar instruction on students' grammatical awareness and writing accuracy. Consequently, the study confirmed that corpus-based grammar instruction had statistically significant retention effect on students' grammatical awareness and writing accuracy. Furthermore, the experimental group students showed a positive attitude towards the instruction.

Corpus-based grammar instruction had significant retention effect on students' grammatical awareness. This result is consistent with the findings of previous studies. To give a few illustrations, Charles (2003), O'Keeffe et al., (2007), and Gardner and Davies (2014), stated that CBGI helped learners in error identification, correction, and exposure to diverse language patterns, and enhanced learners' understanding of grammar rules and structures, respectively. Similarly, Rapti (2010) has found that corpus-based instruction improved students' grammar knowledge, and they preferred to access corpora. With the same token, Elsherbini and Ali (2017) investigated the effects of corpus-based language learning on Iranian EFL learners' grammatical knowledge by analyzing the learners' writing samples. The results indicated noticeable differences in the EFL writing quality between the two groups. In addition, there were significant differences, with the Concordance group gaining more grammatical skills than the Thesaurus group.

Moreover, this study aligned with the Alshahrani et al., (2021) study that demonstrated students utilizing corpus tools, specifically in grammar instruction, showed significant improvements in grammar knowledge and writing skills compared to those taught with traditional methods. With a specific retention effect of the instruction, Lin and Lee (2015) and Girgin (2019) corpus contextualized materials assisted students in learning and increased retention of the grammar learned.

In addition, Kogan et al., (2018) also reported that corpus-based instruction improved students' writing skills. The authentic contexts of grammatical elements stored in the database (corpus) helped EFL learners construct effective writing. Yoon and Hirvela (2004) examined the use of corpora in an ESL setting by studying intermediate and advanced-proficiency-level ESL students' attitudes toward using corpora in L2 writing instruction. They proved that students' writing skills were improved.

Besides, Amare (2021) reported that corpus instruction significantly improved students' writing skills and engagement, similar to the current study's finding. To make a clear connection with the study as mentioned earlier findings, the delayed grammar and writing posttest results of this study indicated that corpus-based instruction had a significant retention effect on students' grammatical awareness ($\eta^2p = 0.491, p < .05$) and writing accuracy ($\eta^2p = 0.378; p < .05$) with a large and moderate effect size, respectively. Therefore, corpus-based grammar instruction had a significant combined retention effect ($\eta^2p = 0.576, p < .05$) on students' grammatical awareness and writing accuracy. Further, results from the qualitative data proved that students who received corpus-based grammar instruction had a positive attitude toward the instruction and were highly motivated to learn grammar using the COCA platform.

4. Conclusion

Corpus-based grammar instruction has increasingly influenced students' grammatical awareness, and writing accuracy, with its immense authentic contexts for different

grammatical elements from various registers (Gaskell & Cobb, 2004). This study proved that the instruction had a long-term effect on undergraduate first-year students' grammatical awareness and writing accuracy. More specifically, the study results showed that students who instructed with CBGI exhibited improved grammatical awareness and enhanced writing accuracy longitudinally. The instruction helped the students perform better in grammar and writing tests than those taught through the conventional approach in EFL grammar class over time. This would confirm that the CBGI significantly improves EFL students' grammatical knowledge and practical applications (Nesi & Gardner, 2012; Boulton, 2017). Further, we considered the experimental group students' positive attitudes and motivations toward the instruction as a genuine impact CBGI exhibited, which underscored its motivational implications for teaching grammar to EFL students. Thus, the findings have various implications for the field of English as a Foreign Language. First, results indicated that CBGI is crucial to integrate corpus tools into EFL classes and it has indispensable roles in grammar teaching. Second, curriculum designer should emphasize the importance of corpus-based grammar instructions, equipping English language teachers with the skills to implement the instruction effectively. The findings also implied that students improved their grammatical knowledge and simultaneously enhanced their writing accuracy through multiple exposures from different contextual language usages (i.e., grammar in context) that would benefit them in using the language for real-world written and spoken communications, where understanding and applying language structures accurately is crucial. Additionally, the study was conducted at Wollo University with undergraduate first-year students. Therefore, further research could examine the retention effect of CBGI across diverse learner populations and contexts. Investigating implementation variations, such as using different corpus types, could provide deeper insights into best practices for maximizing the effectiveness of CBGI. Lastly, qualitative studies that capture student experiences and feedback could offer valuable perspectives on the most engaging and impactful ways to incorporate CBGI into English as a foreign language learning and teaching.

References

- Abiy, Y. (2007). Effects of teacher mediation on students' concepts and approaches to reading. VDM Verlag. ISBN-13: 9783639323467
- Alamirew, G. (2005). *Study on the Perception of Writing, Writing Instruction, and Students' Writing Performance*. Unpublished Ph.D. Thesis, Addis Ababa University.
- Alshahrani, A., Alharthi, A., & Alzahrani, A. (2021). The impact of corpus-based instruction on EFL learners' grammar knowledge and writing skills. *Journal of Language Teaching and Research*, 12(5), 789-799.
- Amare T. & Mulugeta T. (2021). Effects of using corpus-based instructional mediation on EFL Students' academic writing skills improvement. *Theory and Practice of Second Language Acquisition*, 7(2), 133-153.
- Asres, N. (2014). *The effectiveness of indirect focused and indirect unfocused corrective feedback in improving the accuracy of first-year university students' writing*. Unpublished Ph.D. Thesis, Addis Ababa University.
- Azar, B. (2007). Grammar-based teaching: A practitioner's perspective. *TESL-EJ*, 11(2), 1-12.
- Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In *How to Use Corpora in Language Teaching*, J. M. Sinclair (ed.), 15-36. Amsterdam: John Benjamins.
- Boulton, A. (2017). *Corpora in language teaching and learning*. *Lang. Teach*, 50, 483-506. <https://doi.org/10.1017/S0261444817000167>
- Braun, S. (2007). Integrating corpus work into secondary education: from data-driven learning to needs-driven corpora. *ReCALL*, 19(3), p. 307-328. <https://doi.org/10.1017/S0958344007000535>
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47-64. <https://doi.org/10.1017/S0958344005000510>
- Brown, H. (2007). Principles of language learning and teaching. Pearson Education.
- Brown, H. and Frazier, S. (2001) Teaching by Principles: An Interactive Approach to Language Pedagogy (2nd ed.): H. Douglas Brown. *TESOL Quarterly*. 35. <https://doi.org/10.2307/3587655>.
- Burgess, J., & Etherington, S. (2002). Focus on Grammatical Form: Explicit or Implicit? *System Journal*, 30, pp. 433-458. [https://doi.org/10.1016/S0346-251X\(02\)00048-9](https://doi.org/10.1016/S0346-251X(02)00048-9)

- Charles, M. (2003). 'This mystery...': a corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, 2(4), 313-326.
- Conrad, S. (2000). Will corpus linguistic revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34, 548-560. <https://doi.org/10.2307/3587743>
- Conrad, S. (2010). What can a corpus tell us about grammar? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 227-240). Routledge.
- Cullen, R. (2008). Teaching grammar as a liberating force. *ELT journal*, 62(3), 221-230. <https://doi.org/10.1093/elt/ccm042>
- Dawit, A. (2014). The effect of communicative grammar on journalism students' writing skills. *Journal of Media and Communication Studies*, 6(9), 144-149. DOI: [10.5897/JMCS2014.0399](https://doi.org/10.5897/JMCS2014.0399)
- Ellis, N., O'Donnell, M., & Römer, U. (2012). Usage-based language: Investigating the latent structures that underpin acquisition. *Currents in Language Learning*, 1, 25-51.
- Ellis, R. (2002). Grammar teaching-practice or consciousness-raising. *Methodology in language teaching*. Cambridge University Press.
- Elsherbini, S., & Ali, A. (2017). The Effects of Corpus-based Activities on EFL University Students' Grammar and Vocabulary and Their Attitudes toward Corpus. *Journal of Research in Curriculum Instruction and Educational Technology*. 3. 133-161. <https://doi.org/10.21608/jrciet.2017.24482>.
- Fareh, S. (2010). Challenges of teaching English in the Arab world: Why can't EFL programs deliver as expected? *Procedia – Social and Behavioral Sciences*, 2(2), 3600-3604. <https://doi.org/10.1016/j.sbspro.2010.03.559>
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: a critical evaluation. *International Journal of Corpus Linguistics*, 14, 393-417. <https://doi.org/10.1002/9781444315783.ch19>
- Flowerdew, L. (2012). *Corpora and Language Education*. Basingstoke: Palgrave Macmillan.
- Friginal, E. (2018). *Corpus Linguistics for English Teachers: Tools, Online Resources, and Classroom Activities* (1st ed.). Routledge. <https://doi.org/10.4324/9781315649054>
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327. <https://doi.org/10.1093/applin/amt015>
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors?. *System*, 32(3), 301-319. <https://doi.org/10.1016/j.system.2004.04.001>
- Gavioli, L. & Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT Journal*, 55(3), pp. 238-246. <https://doi.org/10.1093/elt/55.3.238>
- Girgin, U. (2019). Perceptions of Turkish EFL student teachers towards learning phrasal-prepositional verbs through corpus-based materials. *Language Teaching and Educational Research*, 2(1), 1-19. <https://doi.org/10.35207/leter.52673>
- Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly*, 37(3), 538-546. <https://doi.org/10.2307/3588404>
- Haregewain, A. (2008). *The Effect of Communicative Grammar on the Grammatical Accuracy of Students' Academic Writing: An Integrated Approach to TEFL*. Ph.D. Thesis, Addis Ababa University.
- Hedge, T.** (2000). *Teaching and learning in the language classroom*. OUP.
- Johansson, S. (2009). Some thoughts on corpora and second language acquisition. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 33-44). Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.33.05joh>
- Johns, T. (2002). Data-driven learning: The perpetual challenge. In *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000*, B. Kettemann & G. Marko (ed.), 107-117. Amsterdam: Rodopi. <https://doi.org/10.1017/S027226310522020X>
- Kemmer, S., & Barlow, M.** (2000). *Usage-based models of language*. CSLI Publications.
- Khan, I. (2011). Learning difficulties in English diagnosis and pedagogy in Saudi Arabia. *Educational research*, Vol. 2, Number.7, pp.1248- 1257.
- Kogan, M., Yaroshevich, A., & Ni, O. (2018). Corpus-based teaching of German compound nouns and lexical bundles for improving academic writing skills. *Revue de linguistique et de didactique des langues*, 58, 1-24. <https://doi.org/10.4000/lidil.5438>
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press Inc.
- Larsen-Freeman, D. (2003). *Teaching language: From grammar to grammaring*. Boston, MA: Thomson/Heinle.
- Larsen-Freeman, D., & Anderson, M. (2011). *Techniques and principles in language teaching* (3rd ed.). New York, NY: Oxford University Press.
- Leech, G. (2013). Teaching and language corpora: a convergence. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora*. London & New York: Routledge.
- Lin, M. (2016). Effects of Corpus-Aided Language Learning in the EFL Grammar Classroom: A Case Study of Students' Learning Attitudes and Teachers' Perceptions in Taiwan. *TESOL Quarterly*, 50, 871-893. <https://doi.org/10.1177/21582440211029936>
- Lin, M., & Lee, J. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, 69(3), 264-274. <https://doi.org/10.1093/ELT/CCV010>

- Ma, Q., Tang, J., & Lin, S. (2021). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer-Assisted Language Learning*, 1–30. <https://doi.org/10.1080/09588221.2021.1895225>
- Marzban, A., Farahani, M., & Hassani, M. (2019). The role of corpus-based instruction in enhancing engagement and effectiveness in grammar learning. *Linguistics and Education*, 54, 100-110. <https://doi.org/10.1016/j.linged.2019.100752>
- Marzulina, L., Mukminin, A., Erlina, D., Astrid, A., Ajriyah, N., Holandiyah, M., & Habibi, A. (2019). The Grammatical Awareness of Student Teachers: The Case of an English Education Study Program in Indonesia. *Universal Journal of Educational Research* 7(9): 1847-1859.
- McCarthy, M., & McCarten, J. (2022). Writing corpus-informed materials. In J. Norton & H. Buchanan (Eds.), *The Routledge handbook of materials development for language teaching* (pp. 170–184). Routledge.
- McEnery, T. & Wilson, A. (2001) *Corpus Linguistics: An Introduction*. 2nd ed., Edinburgh University Press, Edinburgh.
- McEnery, T. & Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, p. 364-380). London & New York: Routledge.
- McEnery, T. & Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- Meunier, F. & Reppen, R. (2015). Corpus versus non-corpus-informed pedagogical materials: Grammar as focus. In D. Biber & R. Reppen (ed.). *The Cambridge handbook of English Corpus Linguistics* (p. 498-514). Cambridge: Cambridge University Press.
- Moore, M. (2021). Grammatical concepts and metalinguistic awareness in first-year college writers: A study of reading journals. *Journal of College Reading and Learning*, 51(3), 178-202. DOI: [10.1080/10790195.2020.1867670](https://doi.org/10.1080/10790195.2020.1867670)
- Muijs, D. (2022). Doing quantitative research in education with IBM SPSS statistics. Sage: New Delhi. <https://doi.org/10.18276/ejsm.2016.20-01>
- Nesi, H., & Gardner, S. (2012). *Genres across the Disciplines: Student Writing in Higher Education*. Cambridge University Press. <https://doi.org/10.1017/9781009030199>
- Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511497650>
- Rapti, N. (2010). A study of classroom concordancing in the Greek context : data-driven grammar teaching and adolescent EFL learners.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- Richards, J., & Rogers, T. (2001). *Approaches and methods in language teaching*. New York: Cambridge University Press
- Richards, J., & Rodgers, T. (2014) *Approaches and Methods in Language Teaching*. Cambridge University Press, Cambridge, England.
- Robinson, P., & Ellis, N. (2008). *Handbook of cognitive linguistics and second language acquisition*. London: Routledge.
- Römer, U. (2011). Corpus research applications in second language acquisition. *Annual Review of Applied linguistics*, 31, p.205-225. <https://doi.org/10.1017/S0267190511000055>
- Schmidt, R. (1990). Consciousness, learning and interlanguage pragmatics. *Paper presented at the Meeting of the World Congress of Applied Linguistics sponsored by the International Association of Applied Linguistics*
- Schmidt, R., (2001). Attention. In: Robinson, P. (Ed.), *Cognition and Second Language Instruction*. Cambridge University Press, Cambridge, pp. 3–32.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford, UK: Oxford University Press.
- Sinclair, J. (2013). Corpus evidence in language description. In. A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Orgs.). *Teaching and language corpora*. London & New York: Routledge.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, 17(1), 38-62. <https://doi.org/10.1093/applin/17.1.38>
- Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian modern language review*, 50(1), 158-164.
- Timmis, I. (2015). *Corpus linguistics for ELT: Research and Practice*. London. New York: Routledge, Taylor & Francis group. <https://doi.org/10.4324/9781315715537>
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13, 257-283. <https://doi.org/10.1016/j.jslw.2004.06.002>