Gist Education and Learning Research Journal. ISSN 1692-5777.
No. 29 (July - December, 2024). pp. 123-149.

Using data-driven learning approach to enhance EFL learners' academic speaking skills

Uso del enfoque de aprendizaje basado en datos para mejorar las habilidades de expresión académica en inglés de los estudiantes de EFL

Getasew Chanie[1]
Yinager Teklesellassie[2]
Kassie Shifere[3]
Bahir Dar University,Ethiopia

[1] getasewcha12@gmail.com
[2] yinagertekle61@gmail.com
[3] kassieshifere@gmail.com

Abstract

The application of corpus to language instruction encouraged the use of data-driven learning (DDL). It is assisted by computer technology, and uses authentic language data as the basis for language instruction. Previously conducted studies have paid little attention to the use of corpus tools in speaking instruction. Thus, this study aimed to examine the effects of data-driven learning approach on EFL learners' speaking skills development. A quasi-experimental research design that employed an interrupted time series design with single group participants was used. The participants were fourth-year EFL major undergraduate students at Mekdela Amba University, Ethiopia. In the intervention, which lasted for six weeks, the participants were taught target language features for speaking via data-driven learning approach. Tests, questionnaires, and the students' reflective journals were used to collect the data. ANOVA and a one-sample t-test were used to analyze the quantitative data, while a qualitative analysis was employed for the students' reflective journals. The findings indicated that the data-driven learning approach improved the speaking skills of EFL students. Additionally, participants felt that data-driven learning was beneficial to their speaking skill development and had positive attitudes on the utilization of DDL approach in speaking instruction. Finally, considering the results, it was recommended that data-driven learning approach ought to be included into EFL speaking instruction.

***Keywords*:** data-driven learning, speaking skills development, authentic language data, computer-assisted approach, speaking instruction

Resumen

La aplicación del corpus en la enseñanza de idiomas ha fomentado el uso del aprendizaje basado en datos (Data-Driven Learning, DDL). Este enfoque, asistido por tecnología informática, emplea datos auténticos del lenguaje como base para la instrucción lingüística. Sin embargo, estudios previos han prestado poca atención al uso de herramientas de corpus en la enseñanza de la expresión oral. Por ello, este estudio tuvo como objetivo examinar los efectos del enfoque de aprendizaje basado en datos en el desarrollo de las habilidades orales de los estudiantes de inglés como lengua extranjera (EFL). Se empleó un diseño de investigación cuasi-experimental con una serie temporal interrumpida y un solo grupo de participantes. Los participantes fueron estudiantes de último año de la licenciatura en inglés como lengua extranjera en la Universidad de Mekdela Amba, Etiopía. Durante la intervención, que tuvo una duración de seis semanas, se enseñaron características lingüísticas específicas para la expresión oral mediante el enfoque de aprendizaje basado en datos. Para la recolección de datos se utilizaron pruebas, cuestionarios y diarios de reflexión de los estudiantes. El análisis de datos cuantitativos se realizó mediante ANOVA y una prueba t de una muestra, mientras que los diarios de reflexión de los estudiantes fueron analizados cualitativamente. Los resultados indicaron que el enfoque de aprendizaje basado en datos mejoró las habilidades de expresión oral de los estudiantes de EFL. Además, los participantes percibieron que este enfoque fue beneficioso para el desarrollo de sus habilidades orales y expresaron actitudes positivas hacia su uso en la enseñanza de la expresión oral. Finalmente, en función de los resultados, se recomendó la inclusión del aprendizaje basado en datos en la instrucción del habla en inglés como lengua extranjera.

Palabras clave: aprendizaje basado en datos, desarrollo de habilidades orales, datos auténticos del lenguaje, enfoque asistido por computadora, enseñanza de la expresión oral

Resumo

A aplicação de corpus no ensino de idiomas tem incentivado o uso da aprendizagem baseada em dados (Data-Driven Learning, DDL). Essa abordagem, assistida por tecnologia computacional, utiliza dados autênticos da linguagem como base para a instrução linguística. No entanto, estudos anteriores prestaram pouca atenção ao uso de ferramentas de corpus no ensino da expressão oral. Portanto, este estudo teve como objetivo examinar os efeitos da abordagem de aprendizagem baseada em dados no desenvolvimento das habilidades orais dos alunos de inglês como língua estrangeira (EFL). Foi utilizado um desenho de pesquisa quase-experimental com uma série temporal interrompida e um único grupo de participantes. Os participantes eram estudantes do último ano da graduação em inglês como língua estrangeira na Universidade de Mekdela Amba, Etiópia. Durante a intervenção, que teve duração de seis semanas, foram ensinadas características linguísticas específicas para a expressão oral por meio da abordagem de aprendizagem baseada em dados. Para a coleta de dados, foram utilizados testes, questionários e diários reflexivos dos alunos. A análise dos dados quantitativos foi realizada por meio de ANOVA e um teste t de uma amostra, enquanto os diários reflexivos dos alunos foram analisados qualitativamente. Os resultados indicaram que a abordagem de aprendizagem baseada em dados melhorou as habilidades de expressão oral dos alunos de EFL. Além disso, os participantes perceberam que essa abordagem foi benéfica para o desenvolvimento de suas habilidades orais e expressaram atitudes positivas em relação ao seu uso no ensino da expressão oral. Finalmente, com base nos resultados, recomendou-se a inclusão da aprendizagem baseada em dados no ensino da expressão oral em inglês como língua estrangeira.

Palavras-chave: aprendizagem baseada em dados, desenvolvimento de habilidades orais, dados autênticos da linguagem, abordagem assistida por computador, ensino da expressão oral

Introduction

Utilizing computer technology in language learning and teaching contexts involves an innovative application of methods and tools that are appropriate for language instruction. Currently, computer technology is widely acknowledged as an essential educational tool in various teaching and learning contexts. This is especially true regarding teaching English language, as computer technology provides several opportunities to improve pedagogy and content (Alqahtani, 2019). Using this computer technology in the classroom ensures that students understand English better by giving them quick access to a range of real language samples (Ludeling & Kyto, 2008). One of the products of this computer technology is the corpus which is regarded as an important tool for English language teaching (Romer, 2010).

According to Reppen (2010), a corpus is defined as "a large, principled collection of naturally occurring written or spoken texts stored electronically" (p.10). The use of corpus in the classroom promoted the application of data-driven learning (DDL) to the instruction of various language skills (Braun, 2007). DDL is an approach of language learning where students use authentic language materials, such corpora, as a basis for their language learning and practice. For teaching English purposes, DDL provides students access to a wide variety of structured and focused written and spoken texts. With the help of digital technologies, learners can access vast databases of texts pertinent to their needs and numerous examples of target elements in context.

Applying DDL in EFL classrooms

According to Romer (2011), there are two ways that DDL can be used in the classroom: direct application and indirect application. By using computer corpus data to examine and internalize the target language, learners can employ corpora directly in an inductive manner. On the other hand, indirect applications are deductive in nature and include students using second-hand corpus data with the help of paper-based examples prepared by teachers. Low-proficiency learners are flooded with the large amount of confusing and unfamiliar language used in direct DDL (hands-on DDL). When learners with limited language skills are directly exposed to authentic language use, the complex and unfamiliar language may hinder their comprehension of the text. Therefore, as teachers work to adapt corpus data to match students' needs and levels of expertise, the indirect approach may help alleviate the concerns raised above (Boulton, 2010). Accordingly, in this study, the researchers used hands-off DDL, in which learners were exposed to concordance outputs and activities in the form of printed handouts.

DDL in speaking instruction

Through exposure to authentic language in DDL instruction, learners acquaint themselves with natural and contextually suitable expressions (Boulton, 2017). A corpus presents valuable resources for language learners aiming to enhance their speaking skills, as it provides numerous examples and language usage patterns. Likewise, instruction based on data-driven learning (DDL) can effectively enhance language learners' speaking skills by offering them genuine language data.

Speaking instruction can benefit greatly from data-driven learning, especially when utilizing a rich corpus such as Michigan Corpus of Academic Spoken English (MICASE). Spoken language corpora show actual, authentic, natural speech that occurs in a variety of settings. According to Baker et al. (2006), many English language learners are interested primarily in learning the language for practical purposes. Thus, it is imperative that students be exposed to as much authentic speech as possible. Similarly, Thornbury (2007) argued that students' speaking abilities improve when exposed to authentic speaking contexts. To expose learners to natural speech patterns and vocabulary used in specific contexts, DDL utilizes real-life language from the corpus. Students can experience features of spoken languages such as small talk, discourse markers, formulaic sequences, vague language, ellipsis, and hedges through DDL-based approaches activities.

Teachers can use existing corpora such as MICASE or compile a corpus based on classroom interactions, activities, or findings from studies such as interviews (O'Keeffe, McCarthy, & Carter, 2007) for speaking instruction. For teachers interested in corpus data, recording classroom interactions is a great way to begin creating a corpus. The procedure may prove to be useful for preparing activities, monitoring the development of learners, and tracking their professional development. Teachers can critically assess their instruction and classroom practices by compiling a teaching corpus (O'Keeffe et al., 2007).

When using corpora for classroom speaking instruction, teachers need to consider the learners' English level and the skills they need to carry out tasks with them. In addition, the proposed corpus activities must match the learners' needs and background knowledge. It may appear at first that corpus data are limited to students at the upper intermediate or advanced levels, but this is not true (O'Keeffe et al., 2007). The teacher can prepare activities by manipulating and organizing the data according to the students' proficiency level and the objectives established.

Using corpus-based materials becomes more important in countries where English is taught as a foreign language because learners in those countries have limited access to the target language (Esayas, 2019). In line with this idea, Yilmaz and Koban Koc (2020) reported that in countries where English is taught as a foreign language, students can produce target language items more quickly and seem more natural by utilizing corpus-based materials. By providing authentic language examples for

language learners and aiding language instruction, this approach allows students to learn how language works through access to authentic language data that are stored electronically (Braun, 2007).

Nevertheless, corpora are not commonly used as a means of language instruction in EFL contexts in Ethiopia (Lakew et al., 2021); consequently, learners have insufficient exposure to authentic English. The classroom material serves as the source of linguistic input for language instruction, and learners are only given examples of a target structure that the teacher has made up. Therefore, adapting speaking instruction to real-world communication may be challenging if it relies too heavily on textbook English or is taught in controlled environments. A study by Adem and Berkessa (2022) confirmed that actual speaking practices in the classroom did not align with the principles of communicative language instruction, which include using and practicing authentic language.

Despite its importance, the applications of corpus-based research have remained limited in spoken learner corpus research (De Cock, 2010). Studies that have previously been conducted have paid little attention to using corpus tools in speaking skills instruction. As mentioned, there is clearly an empirical gap; studies on corpora have failed to consider speaking skills. Many of them have attempted to conduct studies by combining corpus with grammar (Johns, 1991; Boulton, 2009), collocation (Vyatkina, 2016), prepositions (Boontam & Phoocharoensil, 2018), writing (Levchenko, 2017; Birhan et al., 2021; Yoon & Hirevela; 2004, Boulton, 2010) and vocabulary (Geluso &Yamaguchi, 2014). A local study by Lakew et al (2021) aimed to examine the effect of corpus-informed spoken grammar instruction on EFL learners' oral language production, and the findings of the study revealed that corpus-informed spoken grammar instruction enhanced EFL learners' oral language production.

Regarding the researchers' reading, few studies have focused on the effect of the DDL approach on learners' speaking skills (Geluso & Yamaguch, 2014; Sahin Kizil & Savran, 2018). For this reason, studies on the application of corpus into speaking instruction need to be conducted. In the same vein, De Cock (2010) suggested that further must be done on spoken learner corpora. In addition, Cobb and Boulton (2015) highlighted the necessity of conducting research that integrates corpus techniques into speaking skills instruction in an EFL context and suggested that this area of study is a gap to be filled by future research. Therefore, the objective of the current study was to address this gap by employing DDL in speaking activities for EFL students. Consequently, this research aimed to answer the following questions:

1. Does implementing the DDL approach affect undergraduate EFL learners' academic speaking skills?

2. What are undergraduate EFL learners' perceptions of using the DDL approach in EFL speaking skills instruction in an academic context?

Methodology

Research Design

The study employed quasi-experimental research with an interrupted time series design in single group participants. An interrupted time series design involves studying one group, obtaining multiple pre-test measures for a period of time, administering an intervention, and then measuring outcomes (or post-tests) several times. In other words, this design involves repeated observations before and after the intervention and allows the researcher to observe behavior across time (Bordens & Abbott, 2011). In the present study, both quantitative and qualitative data collection and analysis techniques were employed. Thus, the study used the mixed methods approach, which helps the researchers use quantitative and qualitative data to better understanding the research problem.

Research setting and participants

The current study participants were fourth-year undergraduate EFL students at Mekdela Amba University in the academic year of 2023. They were registered for the course 'Seminar on Selected Topics'. A total of 17 participants were included in the study. The sample consisted of 11 males and six females. They ranged in age from 23-26 years old. The participants were from diverse cultural, socioeconomic, and L1 backgrounds. They had exposure to English only through formal education. They used English to study courses at undergraduate level, and they could comprehend and carry out oral and written activities in English despite noticeable difficulties. According to the Common European Framework of Reference for Languages (CEFR): Learning, Teaching, and Assessment (Council of Europe, 2020), the students' level of proficiency can fit into B2 (upper intermediate). The intact group was directly taken for the purpose of the research because the study employed a quasi-experimental study which relies on intact group randomization (Creswell, 2012).

Data collection instruments

Tests

In this study, oral presentation tests were the major data collection tools. Each participant was asked to make four consecutive pre-intervention oral presentations and four post-intervention oral presentations on academic issues. The pre-intervention oral presentations were aimed to examine the students' current level of oral proficiency. Similarly, post-intervention oral presentations were administered to determine the

effect of the intervention on the dependent variable. The oral presentations were informative and argumentative. The time allotted for each oral presentation was seven minutes. The topics of the oral presentations include: 'reasons for university students' drop out' and 'living in countryside versus living in urban areas', 'factors for students' failure in university entrance examination,' and 'controlling university students' way of life'. It was supposed that the contents of these topics of the oral presentations were familiar to students at university level. The contents of the pre- and post-intervention presentations were related but not identical topics. The interval between each pre-test is one week, which is deemed sufficient to establish a baseline measurement and assess any natural fluctuations or trends in the dependent variable. Similarly, the interval between the subsequent post-tests was one week, considered appropriate for tracking immediate changes. Moreover, there was a 6-week gap between the pre and post-intervention oral presentation test because the DDL intervention took 6 weeks, and the first post-intervention was administered immediately after the intervention. Prior to the intervention, the oral presentation tests were piloted to a pilot group to identify any discrepancies in the difficulty level and make possible improvements. Then, the reliability of the pre-test and post-test results was assessed using test-retest reliability. Accordingly, the Cronbach's alpha test of reliability results (for the pre-test, $\alpha = .901$, and for the post-test, $\alpha = .861$) showed that the tests were reliable.

The students' academic oral presentations were audio recorded during the test administration, and their scores were assessed using rubrics that focus on speaking performance. A rubric based on the descriptors of the speaking band of the International English Language Testing System (IELTS) (British Council, 2024) was used to assess the students' speaking performance. The speaking band descriptors include fluency and coherence, grammatical range and accuracy, lexical resources, and pronunciation. Two senior EFL university lecturers were given the recorded oral presentations of the students to score. Based on the speaking band descriptors, the two raters marked the students' oral presentations independently. Orientation was given to raters regarding the speaking band descriptors to help them develop a shared understanding of the scale. Additionally, inter-rater reliability was computed using Pearson's correlations (Pearson's, r), and the result was 0.75, indicating that it was reliable.

Post-instruction questionnaire

This study used a post-intervention perception questionnaire to gather learners' perceptions of the DDL approach in speaking skills instruction. The perception questionnaire was adapted from Yoon and Hirvela (2004), who assessed ESL student attitudes toward corpus use in L2 writing. The questionnaire for perception consists of 18 items that deal with the use of DDL in learning speaking, and the difficulties in the use of the DDL approach. The majority of the items (16 of 18) were positively worded. All the questionnaire items were closed-ended questions which could yield

quantitative data. After the questionnaire items were constructed, it was reviewed by two scholars who were specialized in Teaching English as a Foreign Language (TEFL) for content validity. The reviewers basically agreed on the contents of most of the questionnaire items. Their degree of agreement was estimated to be 78%. Among 18 items, improvements were made on four of the questionnaire items. According to their feedback, two items were found unnecessary, left out, and replaced with relevant ones. The other two items were revised in a way that was simple to understand. In addition, the questionnaire's internal reliability was checked using the Cronbach's alpha test, and the result ($\alpha = .861$) showed that the perception questionnaire was reliable. The questionnaire items were presented as Likert scales, and participants were asked to indicate their levels of agreement on a 5-point Likert scale, where 1 represents strongly disagree and 5 represents strongly agree.

Students' reflective journals

Reflective journals were used to collect qualitative data about students' feelings, thoughts, and experiences on DDL instruction. The students' reflective journals were collected throughout the intervention, which lasted for six weeks, to obtain data on students' reactions while working on corpus-based DDL tasks. The participants (N=17) were requested to keep regular records of their daily experiences regarding the DDL intervention. This helped to understand the participants' perceptions of the DDL approach in their own descriptions. In addition, students' reflective journals provided a means to identify the hindering and contributing factors of the intervention process. Accordingly, a checklist containing six questions was designed to guide students' reflections on the effectiveness of the DDL intervention, major problems encountered during intervention, and ways to improve the intervention.

Data Collection Procedure

The data collection procedure started with selecting the conveniently available university (Mekdela Amba University) for the study and obtaining permission from the university through a formal letter approved by Bahir Dar University. This was followed by receiving participants' consent for participation. Once participants were given necessary orientation, pre-intervention oral presentations were administered to determine the learners' speaking skills/oral presentation skills before the intervention. Then, the intervention was given for six weeks. After the completion of the intervention, post-intervention oral presentations were administered to determine the effects of the DDL intervention on the learners' oral presentation skills. In addition, the study participants completed a post-intervention perception questionnaire immediately after the intervention. To substantiate the perception questionnaire, qualitative data were

collected via students' reflective journals. Learners recorded their daily experiences of learning speaking through the DDL approach. The English version was used for both the questionnaire and the reflective journals because the participants had different L1 backgrounds.

Corpus material design and intervention

The corpus selection was the first step that had been performed before the actual intervention began. A specialized corpus that focuses mainly on the academic genre is necessary to study academic spoken English. Accordingly, the Michigan Corpus of Academic Spoken English (MICASE, 2002) was selected for the study under investigation. This corpus contains a total of 152 speech events, containing examples of virtually every kind of speech event that occurs on a university campus. This wide variety of speeches was recorded and transcribed. It is a corpus composed entirely of academic spoken English, providing the best possible snapshot of the language used in a university context. The corpus was used as a source for two types of information: expression frequency and expression context. Frequency of the expressions was used to select common expressions as models for the students. The corpus also provided contextualized, authentic examples of the expressions used.

After selecting the corpus, the next step was identifying target expressions using the course material and other instructional materials. We collected a list of the formulaic academic expressions for greeting audiences, introducing oneself, introducing the topic, illustrating points, sequencing ideas, expressing opinions, signalling conclusions, summarizing, and thanking, which are commonly used in different sections of an academic oral presentation. Based on the lower cut-off limit for an expression to be considered frequent, which is 10 occurrences/million words, according to the guidelines proposed by Bardovi-Harlig et al. (2015) and Biber et al. (1999), the following frequency counts of the formulaic expressions were taken from the MICASE.

Table 1. Frequency count of formulaic sequences from MICASE

Focus	**Expressions (Occurrences pmw in MICASE)**
Greeting audience	Good morning ladies and gentlemen (14), Good afternoon (17)
Introducing oneself	My name is (18), Let me introduce myself (11), I would like to introduce myself (10)
Announcing the topic	I'm going to talk about (12+), I want to talk about (10)
Illustrating points	To illustrate this point (15), Let me give you an example (12), for example (381)
Sequencing ideas	First/ second/ then/ next I will explain about (11), let's now take a look at (10)
Expressing opinion	Personally, I think (11), I believe that (27)
Signalling conclusion	To conclude (16), as a final point (10+)
Summarizing a point	To summarize (12), Let me briefly summarize (11)
Thanking audience	Thank you (455), Thank you for your attention (10+)

After identifying the target expressions, contextualized examples from the MICASE corpus were extracted. Accordingly, examples of each of the target expressions were selected. The examples were in the form of a concordance line containing the target language feature in context. Moreover, the students were provided with extracts of model oral presentations. For example, the following concordance outputs and sample extracts, which indicate topic introduction expressions, were taken from the corpus.

Concordance line for 'I want to talk about'

Transcript ID: (click to view)	Left context	Match	Right context	View context
LES425SU093	inal as well. so it's kind of a combination of the two. are there any questions...? okay, then, what	i want to talk about	is um, a little bit about the processes of rivers, and understanding and understanding more about wh	view
LES355SU009	ts classical form, although this is not the immediate form that would underlie the Romance form that	i want to talk about	. uh this is the verb that meant to speak or to talk. alright originally you could tell fables to te	view

Sample extract for the topic introduction expression: 'what I want to talk about'

> *Okay, then,* ***what I want to talk about is*** *um, a little bit about the processes of rivers, and understanding more about what the purpose of a river is. Therefore, rivers are basically they're great, levellers that transport material from high in the landscape, and they wash it out to lower in the landscape. So they're moving this material from upward, down, to lower grounds and water basins and then...*

Transcript ID: LES425SU093

Once the language samples were selected and adapted for use in the classroom, the next step was to develop classroom materials and activities. Lessons that provide learners with input and production activities were designed. The lesson included both production and focused-noticing activities.

For the actual intervention, two kinds of activities were designed: consciousness-raising activities and production activities. First, based on the inputs from the corpus, consciousness-raising and noticing activities were designed. These activities contained concordance evidence and model oral presentation extracts from the MICASE corpus. The consciousness-raising activities were designed to make learners aware of the target formulaic expressions found in the authentic language use, and learners were asked to identify and analyse target formulaic sequences used in the spoken transcripts taken from the MICASE corpus. Moreover, they were told to notice the language features presented in the concordance lines and extracts of model oral presentations. For example, learners worked in groups of three to identify how the speakers in the model extracts used the formulaic expressions in the different stages of the oral presentations. Moreover, they were ordered to underline the formulaic expressions that the speaker uses in the different stages of the oral presentations. In another focused noticing activity, learners were given model oral presentation excerpts and concordance lines, which contained highlighted and bolded target formulaic expressions, engaged in noticing the target formulaic expressions to observe how native speakers use these target formulaic expressions in speaking. The consciousness-raising and focused noticing activities generally provided learners with an explicit knowledge of the target language features. Besides, these corpus-based activities could help learners understand essential language to improve their speaking skills/ oral presentations.

In the production activities, speaking activities were designed so that learners could have an opportunity to use knowledge of the target language features or declarative knowledge in communication through repeated practice. This section of the DDL lesson took relatively longer time. For example, in the first section of the DDL lesson, learners worked on practicing how to greet the audience; in the second section, they were asked to announce the oral presentation topics using appropriate formulaic expressions, and so on. For the first time, learners consciously retrieve formulaic sequences to produce speech or to make oral presentation because this is a rehearsal stage in which learners get the opportunity to practice presentation with their new formulaic sequences. However, through repeated exposure and use, learners became more familiar and gradually developed the ability to retrieve and produce them automatically. This leads them gradually to acquire procedural knowledge, possibly forming implicit knowledge.

The corpus-based teaching of target formulaic expressions in speaking instruction improved the sub skills of speaking. When learners are exposed to a wide repertoire of formulaic sequences through corpus-based noticing and production activities, the

formulaic sequences may help improve fluency in speaking because repeated exposure to formulaic sequences allows speakers to retrieve and produce language chunks more quickly, facilitating speech fluency (Wood, 2009). In the current intervention, learners were exposed to formulaic expressions via model oral presentation excerpts from the MICASE corpus, and they had repeated oral presentations using the target formulaic expressions, and this made students' academic presentations more fluent.

Moreover, corpus-based teaching of formulaic sequences introduces learners to a wide range of vocabulary items and other fixed expressions, increasing learners' vocabulary diversity. Studies also confirmed that corpus-based teaching of formulaic sequences positively affects the improvement of lexical resources, which is the main component of the speaking performance. Learners who were treated with DDL approach improved their lexical resources in spoken English (Pan, 2024; Tosun & Sofu, 2023).

Using corpus-informed spoken materials also helps learners to observe how grammar is used in context, allowing learners to observe and internalize correct grammar patterns. When learners are exposed to corpus-based formulaic sequences, they can access different grammatical structures that they may not be able to produce accurately on their own. Moreover, using prefabricated chunks of language reduced the rate of making mistakes that might occur when producing language word by word. Regarding this, Wood (2009) found that ESL learners who used formulaic sequences exhibited increased fluency and a wider range of grammatical structures.

Data analysis methods

In this study, quantitative and qualitative data analysis techniques were employed to analyse the data obtained from the research participants. The one-way repeated-measures ANOVA was employed to determine whether there were any differences in the academic oral presentation scores of the students before and after the intervention. The repeated measures ANOVA used include descriptive statistics (mean scores of the participants' in the four pre-intervention oral presentations and four post-intervention oral presentations), and tests of within-subjects effects (assessing the statistical significance of the effect of the DDL approach across different time periods within the same participants). Additionally, a corpus-based analysis of participants' academic oral presentations was used to determine the frequencies of using the target formulaic sequences before and after the intervention. To do this, oral presentations were digitally recorded and then transcribed into word and text files for use in the data analysis. Then, the transcripts were scanned and converted into an electronic file, and then the Laurence Anthony's Ant Conc (Anthony, 2024) version 4.3.0 software was applied to obtain the frequencies of using the target formulaic sequences from the learners' oral presentations. Moreover, one-sample t-tests and descriptive statistics

were used to analyze the post-intervention perception questionnaire data. This statistical technique compares the sample's observed mean score to the population mean to analyze data from a single sample group. The statistical package for the social sciences (SPSS) version 25 was used to conduct the study's descriptive and inferential statistical operations. Furthermore, in this study, thematic analysis was used to analyze the qualitative data that was gathered from student-reflective journals. This analysis involved identifying and reporting significant themes from the qualitative data. The steps followed in the thematic analysis include reading the students' reflective journals repeatedly to make sense of the data, taking significant quotes from the data, developing themes, and finally analysing and narrating the data.

Findings and discussion

Test results

Each participant delivered eight oral presentation tests (four pre-intervention presentations and four post-intervention presentations). These oral presentation tests were used to examine the effect of the data-driven learning approach on the academic oral presentations of EFL learners. The oral presentation test results analysed via one-way repeated measures ANOVA are presented.

The descriptive statistics

Table 2. Students' test scores before and after the intervention

	Mean	Std Devation	N
Pre-test 1	37.65	2.714	17
Pre-test 2	38.41	3.022	17
Pre-test 3	40.65	2.999	17
Pre-test 4	41.82	2.651	17
Post-test 1	46.12	2.956	17
Post-test 2	48.88	3.018	17
Post-test 3	51.29	2.640	17
Post-test 4	52.41	2.265	17

The descriptive statistics result in Table 2 shows the potential differences in the students' test scores for academic speaking skills before and after the intervention. Accordingly, the mean scores of the students' speaking test results before the intervention were (Pre-test-1, M = 37.65; Pre-test-2, M = 38.41; Pre-test-3, M = 40.65,

and Pre-test-4, M = 41.82). Likewise, the mean scores of the students' speaking skills performance post-test scores were (Post-test-1, M = 46.12; Post-test -2, M = 48.88; Post-test-3, M = 51.29, and Post-test-4, M= 52.41). From these test results, one can understand that the students' mean scores in the pre-tests were slightly increased from Pre-test 1 to Pre-test 4. Similarly, there is a slight increment in the students' mean scores from Post-test 1 to Post-test 4. However, there is a significant difference in the mean scores of the pre-tests and post-tests. In other words, the students' mean scores in the post-tests were higher than those of the pre-tests. It implies that the students' academic speaking skills have been improved after the intervention.

Tests of within-subjects effects

The test of within-subject effects assesses the statistical significance of the effect of the independent variable across different situations or time periods within the same subjects. In this study, the performance of the sample students on tests revealed a notable mean effect of data-driven instruction on their academic oral presentations ($F (3.908, 62.534) = 166.537$; $p < 0.05$). The obtained p-value signifies a statistically significant effect of data-driven instruction on the students' academic oral presentations. Additionally, the partial eta squared (ηp2) indicates the magnitude of the effect on the sample students' oral language production. As shown in Table 3, the intervention yielded an effect size of .912, signifying a large effect size greater than the usual cut-off point of .14.

Table 3. Tests of within-subjects effects

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Test	Sphericity Assumed	4019.228	7	574.175	166.537	.000	.912
	Greenhouse-Geisser	4019.228	3.908	1028.360	166.537	.000	.912
	Huynh-Feldt	4019.228	5.330	754.144	166.537	.000	.912
	Lower-bound	4019.228	1.000	4019.228	166.537	.000	.912
Error(Test)	Sphericity Assumed	386.147	112	3.448			
	Greenhouse-Geisser	386.147	62.534	6.175			
	Huynh-Feldt	386.147	85.272	4.528			
	Lower-bound	386.147	16.000	24.134			

Corpus analysis results

In addition to the above quantitative analysis, a corpus-based analysis of participants' academic oral presentations was used to determine the frequencies of using the target formulaic sequences.

This method enabled the researchers to note the effect of the DDL intervention on the participants' oral presentations by comparing their use of the target formulaic sequences before and after the implementation of the DDL intervention. Table 4 below presents the frequency of using the formulaic sequences.

Table 4. Frequency of using formulaic expressions in pre-tests and post-tests

Focus	**Formulaic sequences**	**Pre-tests**	**Post-tests**
Greeting audience	Good morning	6	14
	Good afternoon	12	18
Introducing oneself	My name is	10	10
	Let me introduce myself	5	7
	I would like to introduce myself	2	3
Announcing the topic	I am going to talk about	2	11
	I want to talk about	8	14
Illustrating points	To illustrate this point	-	3
	Let me give you an example	4	6
	For example	23	25
Opinion	Personally, I think	6	16
	I believe that	4	3
Conclusion	To conclude	3	17
	As a final point	-	4
Summary	To summarize	5	13
	Let me briefly summarize	1	9
Thanking	Thank you	14	27
	Thank you very much for your attention	8	29

Each study participant delivered four consecutive pre-intervention and four post-intervention oral presentations. This means that we have 136 oral presentations from the 17 participants. The above-mentioned frequencies of formulaic sequences are obtained by analysing the oral presentations delivered by the research participants. Based on the data, the participants used a total of 113 formulaic sequences in the pre-tests and 228 formulaic sequences in the post-tests. Formulaic sequences used for greeting, illustrating a point, and thanking the audience are the most used. The number of formulaic sequences in the post-tests is much greater than the formulaic sequences used in the pre-tests. This indicates that the quality of oral presentations after the intervention was better, as learning formulaic expressions contributed to improving students' speaking, particularly oral presentation (Dickinson, 2019).

Post-intervention questionnaire results

Table 5. Participants' perceptions of the DDL-based intervention

Items	Statements	Mean	SD
1	The DDL instruction was interesting	3.71	.686
2	The presentation of the DDL tasks and activities were good	3.59	.507
3	The DDL instruction helped you improve your oral presentation skills	3.88	.600
4	The DDL tasks were not difficult to complete	3.88	.485
5	If you attend a speaking lesson, you should always use the DDL units	3.53	.624
6	The concordance-based activities were not relevant to the course seminar on selected topics	2.47	.514
7	The concordance inputs helped you discover new patterns	3.59	.507
8	New vocabulary and formulaic expressions were taught in the DDL instruction.	3.65	.702
9	The DDL instruction helped you learn how to greet audiences in your oral presentation	3.59	.507
10	The DDL instructionw helped you learn how to introduce yourself and your topic of presentation	3.59	.507
11	The DDL units helped you learn how to express the purpose of your presentation	3.59	.507
12	The instruction in DDL helped you learn how to outline your points of presentation.	3.41	.507
13	The units helped you learn how to illustrate a point in a presentation	3.35	.493
14	The DDL based instructions were easy to follow	2.71	.470
15	The language points selected in the DDL instruction were useful	3.35	.493
16	The concordance lines were easy to understand	2.65	.493
17	The concordance lines were well chosen (e.g. vocabulary not too difficult)	2.35	.606
18	Noticing of language features in concordance lines was an interesting way of learning English	3.00	.500

According to Table 5, four of the items (items 6, 14, 16, and 17) have mean scores that are less than 3, and most of the items (items 14 out of 18) have mean scores that are greater than the average (i.e., 3). The mean score of item 6 is below the average, but it doesn't mean it should be interpreted negatively. Most participants did not agree with the item, as indicated by the statement, "The units are not relevant to the course seminar on selected topics." Consequently, this item suggests a positive interpretation, and the researchers see a lower mean score as positive. The lower mean scores of the other items (items 14, 16, and 17) are taken negatively, indicating that participants do not think the concordance lines and instructions are easy. This suggests that the participants faced difficulty comprehending the input presented in the concordance lines due to the difficulty of the vocabulary.

According to the responses of the participants of the study, the implementation of DDL activities in the classroom helped the students learn formulaic expressions that helped them use various sections of their academic oral presentations. In addition, most of the learners responded that the corpus helped them improve their academic oral presentation skills.

Generally, by examining the intervention and questionnaire findings, it was observed that students benefited from DDL-based activities and held positive attitudes towards using the corpus in the classroom. Thus, instructional materials for speaking instruction should be corpus-informed to include real-life practices and to teach learners actual language use. The findings of the study also align with the findings of Sahin Kizil and Savran (2018). These authors conducted a small-scale study on the integration of corpus into speaking instruction, specifically explored students' perceptions towards the instruction, and found that students benefited from concordance-based learning activities and held positive attitudes towards using it in learning speaking.

To determine whether the mean scores of the Likert scale were significantly greater than the expected mean, the researchers computed a one-sample t-test. The descriptive statistics reveal that the majority of the sample students' responses have mean scores that are above the average point, but they do not indicate whether these mean scores are significantly greater than the average value. Table 6 below presents a summary of the t-test results.

Table 6. The descriptive statistics of the students' questionnaire

One-Sample Statistics				
Perception score	N	Mean	Std. Deviation	Std. Error Mean
	17	3.3465	.48600	.11787

Table 6 depicts the mean and standard deviation scores of the participants were 3.3465 and 0.48600 respectively. The participants' mean score (i.e., 3.3465) was slightly greater than the population mean (i.e.,3.00). However, a one-sample t-test was carried out because the descriptive statistics results by themselves were insufficient to determine whether there was a statistically significant difference between the sample mean and the hypothesized mean, and the results are shown below.

Table 7. The one-sample t-test results for the student questionnaire

	Test Value = 3					
	T	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Perception score	2.939	16	.010	.34647	.0966	.5964

As shown in Table 7, the perception score (t (16) = 2.939, p = 0.010, MD = 0.34647) was greater than the hypothesized mean score, and the difference was statistically significant (p = 0.010). Thus, it can be concluded that the participants' perceptions of learning speaking through the DDL-based approach were positive.

Data from student-reflective journals

The students' reflective journal results showed that students' academic oral presentation skills improved after implementing the data-driven instruction. The students reported that the method is helpful for their oral presentations. For instance, Student 4 said:

> Compared to the strategies I utilized previously, the DDL approach is beneficial for my academic oral presentation because it gives me the chance to observe crucial linguistic features for my presentation. I developed my presentation with the assistance of the sample oral presentation extracts that were given in class. The linguistic inputs in the form of concordance lines helped me enhance my presentation by providing me with significant expressions.

Furthermore, students in the reflective journal reported that they had problems in their academic presentations before the current intervention. They stated that their problems were due to a shortage of formulaic sequences used in different sections of oral presentations. To this end, the students reported that they liked the DDL approach because the data-driven instruction helped them to ease their oral presentation problems. One of the students put his view this way:

> Compared to the previous teachers' methods of teaching speaking skills through oral presentations, I felt comfortable with the data-driven instruction because this instruction provided me with authentic input and guided me to improve my academic oral presentation skills. I am now better at my oral presentation.

As the students stated in the student-reflective journal, the inputs given to them in the form of concordance lines and model extracts were helpful to know more about formulaic sequences that are used to make academic oral presentations. The students mainly stated that the noticing and consciousness-raising activities used in the corpus-driven instruction helped them identify target linguistic features used to improve their academic oral presentation skills. In the same vein, the learners stated that the concordance helped them improve their understanding of formulaic sequences, and they suggested that it was useful approach for learning formulaic sequences.

Regarding the difficulties faced during DDL-based instruction, one of the learners mentioned the following:

> At the very beginning of the intervention, I faced many challenges. For example, the unfamiliar vocabulary and the cut-off sentences in the concordance output challenged me to understand the target language feature. However, the teacher's mediation helped me to understand and enjoy the instruction.

In summary, the data obtained from the students' reflective journal showed that despite some difficulties, the DDL unit helped enhance students' academic oral presentation skills.

Discussion

This study aimed to examine the effect of data-driven instruction on EFL learners' speaking skills. The research findings confirmed that data-driven instruction significantly improved EFL learners' speaking skills. As indicated in the data analysis section, there was a significant difference between the mean scores of the pre-test and post-test results. The mean scores from the pre- and post-tests verified that the mean differences were statistically significant. The post-test scores were greater than the pre-test scores, implying that the intervention significantly affected the participants' academic oral presentations. Moreover, the tests of within subjects effects result ($F(3.908, 62.534) = 166.537$; $p < 0.05$) indicated that the DDL intervention brought a significant effect on learners' academic speaking skills.

The results obtained from this study agree with those of previous investigations. For example, a study conducted by Sahin Kizil and Savran (2018) proved that corpus-based activities were beneficial for EFL learners' speaking skills development and that there was a positive attitude toward web-based concordancing. In the same vein,

Pan (2024) confirmed that the DDL approach significantly improved the vocabulary production of EFL learners in their spoken English. The effectiveness of corpus-based instruction for improving pragmatic and speech act knowledge was also demonstrated by Sabzalipour et al. (2017) and Bardovi-Harlig et al. (2017), the latter of which revealed that both direct corpus searches and teacher-developed materials were effective.

Regarding the perception of participants, the descriptive statistics result showed that the participants' perception mean score (3.34) was greater than the hypothesized population mean score (3.00). Moreover, the one sample t-test result, (t (16) = 2.939, p = 0.010, MD = 0.34647) confirmed that participants had favourable perception towards learning speaking through DDL approach. In support of this, results obtained from reflective journals revealed that the DDL approach helped learners with their academic speaking skills. Participants noted that the corpus data provided for learners as input, and the concordance-based activities helped them to enhance their academic speaking. Participant clearly stated that the DDL intervention effectively taught the target language features. They noted that their academic speaking has improved and they became more aware of formulaic sequences for speaking. Moreover, they were pleased to be exposed directly to authentic real-life language or corpus data.

These findings obtained from the perception questionnaire were also consistent with previous studies. The results of the study by Geluso and Yamaguchi (2014); students believe DDL to be a useful and effective tool in the classroom for speaking skills instruction. As a result, these authors concluded that for speaking skill, which is an area in language in which learners have various types of difficulty, providing students with corpus-informed materials can improve their oral communication. In their studies, Lakew et al. (2021) and Birhan et al. (2021) also reported that students exhibiting significant achievements and showing a favourable attitude toward the instruction.

In addition to the speaking tests, learners in their reflective journals indicated that their lexical resources, which are basic speaking components, improved due to the paper-based DDL instruction. This, in turn, helped them develop their speaking skills. This finding is also consistent with the study of Lay and Yavuz (2020), who found that, at the B2 level, paper-based DDL instruction is more effective than conventional instruction using academic lexical bundles.

Collectively, the results of this study and previous studies indicate that data-driven learning can help EFL learners improve their speaking abilities, including in Ethiopia.

Conclusion and pedagogical implications

This study examines the effect of incorporating corpus-based DDL instruction into teaching academic speaking, particularly academic oral presentations. Teacher-prepared paper-based DDL instruction helps learners improve their academic oral presentation skills. The clear pedagogical implication of this study is that a careful combination of teacher-developed DDL materials for speaking instruction helps learners notice that formulaic sequences would be ideal. Moreover, teachers can supplement classroom teaching materials through the use of corpora containing language samples for speaking instruction. This study also shows that a corpus-informed approach has an important impact on learning spoken features of a language and micro skills of speaking.

The findings of this research also carry different implications. First, the findings suggest that a computer-aided DDL approach is beneficial for discovering formulaic sequences that can be used for speaking in natural contexts. Second, the findings indicate that English language teachers should consider empirical language data when preparing academic speaking materials to meet students' academic speaking needs instead of relying on teacher-made examples; teachers should depend on authentic examples and linguistic resources content.

Limitations

Despite the effective attempts to conduct this study, it was not without limitations. Activities and tasks for the intervention material were designed depending on an academic spoken English corpus (MICASE) because the problem under investigation is academic speaking English. However, taking speech transcripts did not contribute to improving participants' pronunciation, which is an important component of speaking. If an audio corpus had supplemented the preparation of activities and tasks, participants could have observed native speakers' pronunciations. Moreover, the application of DDL approach can be better if the combination of direct DDL and indirect DDL is applied. However, logistic constraints prevented the researchers from applying the direct approach. For this reason, the preparation of DDL activities becomes time-consuming and tiresome.

References

Adem, H., & Berkessa, M. (2022). A case study of EFL teachers' practice of teaching speaking skills vis-à-vis the principles of Communicative Language Teaching (CLT). *Cogent Education*, *9*(1), 2087458

Alqahtani, A. (2019). The use of technology in English language teaching. *Frontiers in Education Technology*, *2*(3), 168-180.

Anthony, L. (2024). AntConc (MacOS 10/11, 4.1.3 Version) [Computer Software]. Waseda University. https://www.laurenceanthony.net/software/antconc/

Baker, P., Hardie, A. and McEnery, T. (2006.) *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press Ltd.

Bardovi-Harlig, K., Mossman, S., & Vellenga, H. E. (2015). Developing corpus-based materials to teach pragmatic routines. *TESOL Journal*, 6(3), 499-526.

Bardovi-Harlig, K., Mossman, S., & Su, Y. (2017). The effect of corpus-based instruction on pragmatic routines.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* London, UK: Longman.

Birhan, A., Teka, M., & Asrade, N. (2021). Effects of using Corpus-Based instructional mediation on EFL Students' academic writing skills improvement. *Theory and Practice of Second Language Acquisition*, *7*(2), 133-153.

Boontam, P and Phoocharoensil, S. (2018). Effectiveness of English preposition learning through Data-Driven Learning (DDL): *The Southeast Asian Journal of English Language Studies – Vol 24(3), 125 – 141.*

Bordens, K., & Abbott, B. (2011). Research design and Methods: A process approach. Indiana: McGraw-Hill.

Boulton, A. (2009). Testing the limits of data-driven learning: language proficiency and training. In *Re-CALL, 21*, 37–54. URL: https://doi.org/10.1017/S0958344009000068

Boulton, A. (2010). Learning outcomes from corpus consultation. In M. Moreno Jaen, F; Serrano Valverde & M, Calzada Perez (Eds.), Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching, (1st ed., 129-144). Equinox.

Boulton, A. (2017). Data-driven learning and language pedagogy. In S. Thorne & S. May (eds.), Language, Education, and Technology: Encyclopedia of Language and Education. New York: Springer. DOI 10.1007/978-3-319-02328-1_15-1

Braun, S. (2007). Integrating corpus work into secondary education: from data-driven learning to needs-driven corpora. *ReCALL, 19, 3*, 307-328.

British Council. (2024). IELTS Speaking Band Descriptors. Retrieved from https://takeielts.britishcouncil.org/sites/default/files/ielts_speaking_band_descriptors.pdf

Cobb, T. & Boulton, A. (2015). Classroom applications of corpus analysis. In: Biber, D. & Reppen, R. (eds.): *The Cambridge handbook of English corpus linguistics*. Cambridge: CUP, 478–497.

Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment-Companion volume. Council of Europe Publishing. www.coe.int/lang-cefr

Creswell, O. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th Edn.). Pearson.

De Cock, S. (2010). Spoken learner corpora and EFL teaching. In M. C. Campoy, B. Bellés-Fortuño and M. L. Gea-Valor (Eds.), *Corpus-based Approaches to English Language Teaching* (pp. 123-137). London: Continuum.

Dickinson, P. (2019). Improving second language academic presentations with formulaic sequences, Journal of Niigata University of International and Information Studies, 15, 25-36.

Esayas, T. (2019). The Practice of Teaching Speaking Skills: the case of three Secondary Schools in Gedeo Zone, Ethiopia. IOSR Journal of Humanities and Social Science (IOSR-JHSS).vol. 24 no. 03, 2019, pp40-47.

François, J. & Albakry, M. (2021). Effect of formulaic sequences on fluency of English learners in standardized speaking tests. Language Learning & Technology, 25(2), 26–41. http://hdl.handle.net/10125/73429

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL, 26*(2), 225–242. http://doi.org/10.1017/S0958344014000044

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

Lakew, A., Teshome, S., & Negede, D. (2021). Introducing corpus-informed spoken grammar instruction in EFL classrooms: the Ethiopian experience. *Corpus Pragmatics*, *5*(4), 487-510.

Lay, K. J., & Yavuz, M. A. (2020). Data-driven learning of academic lexical bundles below the C1 level. *Language Learning & Technology*, 24(3), 176–193. http://hdl.handle.net/10125/44741

Levchenko, V. (2017). Use of Corpus-based classroom activities in developing academic awareness in doctoral students. *The New Educational Review*, *48*, 28-40. DOI: 10.15804/tner.2017.48.2.02

Ludeling A.& Kyto M. (2008). *Corpus Linguistic*, An International Handbook. Walter deGruyter GmbH & Co. KG, 10785 Berlin, Germany

O'Keeffe, A., McCarthy M., & Carter R. (2007). From corpus to classroom. Cambridge: Cambridge University Press.

Pan, Z. (2024). EFL Learners' Use of Data-driven Learning and their Attitudes in the Improvement of English-speaking Skills. *International Journal of Learning, Teaching and Educational Research*, *23*(8).

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

Romer, U. (2010). Using general and specialized corpora in English language teaching: Past, Present, and Future. In M.C. Campoy, B. Belles-Fortuno & M. L. Gea-Valor (Eds.), *Corpus-based Approaches to English Language Teaching* (18-38). London: Continuum

Romer, U. (2011). Corpus research applications in second language teaching. Annual Review of Applied Linguistics, 31, 205–225. https://doi.org/10.1017/S0267190511000055

Sabzalipour, B., Koosha, M., & Afghari, A. (2017). Investigation the effect of colloconstructural corpus-based instruction on pragmalinquistic knowledge of request speech act: Evidence from Iranian EFL students. *International Journal of Applied Linguistics and English Literature*, *6*(7), 103-109.

Sahin Kizil, A., & Savran, Z. (2018). The integration of corpus into EFL speaking instruction: A study of learner perceptions. *International Online Journal of Education and Teaching (IOJET),* 5(2), 376-389. http://iojet.org/index.php/IOJET/article/view/368/241

Thornbury, S., 2007. *How to teach speaking*. Harlow: Pearson Education Limited.

Tosun S. & Sofu H. (2023). The Effectiveness Data-Driven Vocabulary Learning: Hands-on Concordancing through a Pedagogical Corpus. Journal of Language and Education, 9(3), 177-191.https://doi.org/10.17323/jle.2023.12426

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. Language Learning & Technology, 20(3), 159–179. Retrieved from http://llt.msu.edu/issues/october2016/vyatkina.pdf

Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. Canadian Journal of Applied Linguistics, 12(1), 39-57.

Yilmaz, N., & Koban Koc, D. (2020). Developing pragmatic comprehension and production: Corpus-based teaching of formulaic sequences in an EFL setting. *Journal of Language and Linguistic Studies, 16*(1), 474-488. Doi: 10.17263/jlls.712880

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

Authors

Getasew Chanie is a lecturer at Mekdela Amba University, Ethiopia. He is a Ph.D. candidate at Bahir Dar University. His area of research interest includes using corpus-based instruction in grammar instruction, introducing corpus-informed materials in writing instruction, improving learners' pragmatic competence through corpus-based instruction and corpus-based instruction of formulaic sequences to improve learners' speaking skills.

ORCID: https://orcid.org/0009-0005-2585-2247

Yinager Teklesellassie (Ph.D.) is an associate professor and serves as an instructor and researcher at Bahir Dar University, Ethiopia. He has been teaching various courses to students in undergraduate and postgraduate programs. He has also been advising and supervising MA and Ph.D. students.

ORCID: https://orcid.org/0000-0001-8053-474

Kassie Shifere (Ph.D.) is an associate professor and serves as an instructor and researcher at Bahir Dar University, Ethiopia. He has been teaching various courses to students in undergraduate and postgraduate programs. He has also been advising and supervising MA and Ph.D. students.

ORCID: https://orcid.org/0000-0001-6035-0596