



## **Lexical Complexity in ELT Students' Essays across Genres and over the Course of Time**

**Nalan Bayraktar Balkir**

Canakkale Onsekiz Mart University

**Handan Celik**

Trabzon University

**Gokhan Cepni**

Trabzon University

### **ABSTRACT**

*This study investigated lexical complexity in English language teaching (ELT) students' essays across genres and over the course of time. The essays included descriptive, compare and contrast, advantage and disadvantage, and argumentative genres. Employing a longitudinal, quasi-experimental and repeated measures design in a year-long writing course in the ELT department of a northeastern Turkish university, we analyzed lexical density, lexical sophistication and lexical variation in a corpus of 229 essays via the web-based lexical complexity analyzer developed by Ai and Lu (2010). The scores of all measures were analyzed using SPSS for further examination through descriptive and inferential statistics. Of the three indices, only the lexical density index was found to have a significant difference across genres, with the highest score in the descriptive essay. The findings related to lexical complexity development over time revealed that a statistically significant change was only captured in the lexical variation index. These findings might contribute to the language acquisition research as our study encompasses the examination of all three lexical complexity indices across four different genres and over time. In this vein, some considerable implications for future writing research and for writing instruction in foreign/second language education are suggested.*

### **INTRODUCTION**

Writing as a tool and means for formal and informal communication has always had importance since cave drawings. The significance of writing increases even more in today's world as it connects people for various purposes ranging from the need to write emails to formal business letters across various sectors, regions, and languages. As Ortega (2015) stated, "writing becomes formally and structurally more complex only because emergent and skilled writers are challenged and seek to challenge themselves with the creation of meanings that are, conceptually and socially, also increasingly more complex" (p. 83). In this vein, with an increasing need to teach English language skills, in particular writing skills, to students and professionals of various disciplines, the role played by English language teachers who are competent in their own writing skills before they teach them becomes more obvious.

However, due to the difficulty and challenge inherent in 'writing' (see O'Leary & Steinkrauss, 2022), there has been a large amount of growth in the empirical second and foreign

language (L2-FL) writing research (Manchon & de Haan, 2008). This has further resulted in greater emphasis on the importance of cultivating English language teachers' writing skills and performance in both paragraph and essay writing (see Arslan, 2013; Cheung, 2011; Meza, Rodriguez, & Caviedes, 2021; Reichelt, 2019) which may stress out the writer, i.e. FL writer or pre-service English teachers, due to the demanding nature of idea generation and organization and expression of them in a limited time (C. Lee, Wong, Cheung, & F. S. L. Lee, 2009). However, owing to the concerns regarding the lack of reliability in writing assessment (Wu, Steinkrauss, & Lowie, 2023) that is often based on holistic or analytic rating scales (Wolfe-Quintero, Inagaki, & Kim, 1998), there is now an increasing interest among writing researchers in using corpora of various sources i.e., learner-generated in formal language classrooms (see Ha, 2022; Kalantari & Gholami, 2017; Saricaoglu, Bilki, & Plakans, 2021; Zare, Bagheri, Sadighi, Rassaei, & Goel, 2021; Zhan, Sun, & Zhang, 2021) or learner-generated in language immersion programs (Schnur & Rubio, 2021) to measure language development in written modality (Hsieh & Liou, 2008; Smirnova, 2017) in a progressive manner. This is in line with the efforts in second language acquisition research derived from developmental index studies which aim to assess and describe learners' developmental levels through fluency, accuracy, and complexity indices serving as objective, precise, and comparable measures (Wolfe-Quintero, Inagaki, & Kim, 1998).

Attempts to describe pre-service foreign language teacher education program learners' developmental levels in written modality seem to hold as much importance as the efforts to teach them how to teach writing skills in a foreign language. As Hirvela and Belcher (2007) state, "there are many among us whose roots are in teacher education and whose daily work involves helping pre-service and in-service teachers become familiar with the theories, research, and practices that form the praxis of L2 writing instruction" (p. 125). Being among them, we, as the authors of this paper, are curious about how complex the lexis is and how far lexical complexity in the essays written by pre-service English language teachers (interchangeably ELT students) varies across genres over the course of the Writing course they take in their junior year in the English language teaching program in a state university in Türkiye.

## **THEORETICAL FRAMEWORK**

### **Lexical complexity**

People with a limited vocabulary tend to have a major constraint in communicating their messages both in spoken and written discourse as the "words are the basic building blocks of language, the units of meaning, from which larger structures such as sentences, paragraphs and whole texts are formed" (Read, 2001, p.1). As seen, vocabulary forms the core of communication among the users of any language. Therefore, it becomes essential to build an ever-developing body of knowledge and use of vocabulary to communicate effectively. In this respect, vocabulary building or lexical development is a key area of language learning as L2-FL learners are expected to expand the capacity and range of their vocabulary gradually to become increasingly competent in the use of the language they are studying. In this regard, efforts to assess and improve learners' lexical capacity become crucial. This closely relates to researching lexical complexity in language acquisition indicating the range and degree of sophistication of FL learners' productive vocabulary (Wolfe-Quintero, Inagaki, & Kim, 1998).

To understand where lexical complexity fits in its broader context, it might be useful to refer to Bulten and Housen's (2012) work where they describe complexity along with fluency and accuracy "as a basic descriptor of L2 performance and as an indicator of L2 proficiency" (p. 22) and define it as

a property or quality of a phenomenon or entity in terms of (1) the number and the nature of the discrete components that the entity consists of, and (2) the number and the nature of the relationships between the constituent components. (p. 22)

It is seen that language complexity is quite an elusive term and a multi-level and multi-dimensional construct; nevertheless, within the scope of this study, *lexical complexity* could be regarded as one of the two major components of linguistic complexity, along with grammatical (syntactic) complexity (Bulté & Housen, 2012). As said, lexical complexity is often conceptualized as a multidimensional feature of language use with three main sub-constructs: lexical density, lexical variation, and lexical sophistication (Lu, 2012). *Lexical density (LD)* is a measure of the ratio of content words such as nouns, verbs, adjectives and often adverbs (as opposed to function or grammatical words, i.e., pronouns, conjunctions, prepositions, determiners, etc.) to the total number of words in a text. *Lexical variation (LV)*, also known as lexical diversity or range, refers to the number of different words in a text. *Lexical sophistication (LS)* is the number of infrequent or rare words in a text (Thornbury, 2002). These three measures of lexical complexity are the main indicators of learners' developmental level of lexical competence in FL-L2 spoken or written discourse. Lexical complexity measures, i.e. the size, variety, and quality of a learner's vocabulary, are regarded as good predictors of writing quality (see Schnur & Rubio, 2021) and target language proficiency. From a theoretical perspective, as an L2-FL learner becomes more proficient, they are expected to gain more lexical richness in their language production (Spring & Johnson, 2022).

### **Research on lexical complexity and L2-FL writing**

Lexical complexity is an important construct in L2-FL research and teaching and strongly linked to learners' ability in spoken and written communication (Lu, 2012). In the literature to date, indices of lexical complexity in L2-FL writing have been investigated in a handful of experimental studies on writing performance or quality (e.g., Gao & Min, 2021; Gregori-Signes & Clavel-Arroitia, 2015; Kessler, Polio, Xu, & Hao, 2020; Seidinedjad & Nafissi, 2018; Uzun, 2019; Yıldız & Yeşilyurt, 2017). To exemplify, oral prewriting discussions and text chat were observed to result in increased lexical complexity and syntactic richness in Chinese university learners' essays (Kessler, Polio, Xu, & Hao, 2020). Besides, using creative thinking techniques improved lexical complexity in Iranian university learners' essays written under both timed and untimed conditions (Seidinedjad & Nafissi, 2018). In a Turkish university setting, while genre-based instruction and genre-focused feedback improved the mastery of the literary analysis essay, lexical complexity did not change throughout the intervention (Uzun, 2019). In another study conducted with Turkish EFL learners in higher education, the use of a prewriting technique showed that lexical variation remained unchanged during the study (Author1, 2016). As indicated by this brief review, the interventions have yielded varying results in terms of improvement in different dimensions of lexical complexity.

In addition to the aforementioned experimental studies, some other studies have been implemented to track the changes in lexical complexity indices in learner essays. One such study examined the lexical and syntactic features of essays written by learners of two different proficiency groups enrolled in an English for Academic Purposes program at a British university (Mazgutova & Kormos, 2015). The learners from a lower and a higher proficiency level were required to write two argumentative essays, one at the beginning and one at the end of the course. The findings revealed that lexical diversity increased in both groups' essays. In a similar study, syntactic and lexical complexity indices were examined in argumentative essays written by two groups (B2 and C1 level) of EFL learners at a Turkish university setting (Saricaoglu & Atak, 2022). The findings displayed significant variation among the two

proficiency levels in several syntactic structures, whereas no significant difference was observed between the groups regarding lexical complexity. In a cross-sectional comparative study, Nasser and Thompson (2021) examined lexical density and diversity differences in the dissertation abstracts of English L1 academic writers and L2 (both EFL and ESL) peers. They found that the EFL students' texts were the least lexically dense and diverse compared to the ESL and L1 group, which implied a lexical proficiency gap between the EFL and the other groups. Thus, the authors concluded that learners in EFL academic settings might be explicitly instructed on how to write more lexically dense and diverse abstracts. However, when it comes to the examination of lexical complexity development across different genres in learner-generated texts, as we do in the current study, a limited number of studies are reported. For instance, Ansarin, Karafkan and Hadidi (2021) investigated the likely role of writing task type on lexical diversity and sophistication and also the relationship between these two dimensions in the narrative, argumentative, and descriptive texts of upper-intermediate Iranian EFL learners. They found that lexical diversity displayed significant variation across the three writing tasks with scores from highest to lowest in the narrative, argumentative, and descriptive texts. Besides, lexical sophistication indices were higher in the argumentative texts than in the narrative and descriptive genres. Also, a significant positive relationship between lexical sophistication and lexical diversity was manifest in all three genres. In another study conducted by Yoon and Polio (2017), ESL learners' narrative and argumentative essays were examined for linguistic development over time and across genres during a 4-month long semester at a U.S. university. The findings resemble those of Ansarin et al. (2021) in that the students had a wider variety of lexis in the narrative texts than in the argumentative, which might be a result of learners' overuse of formulaic phrases that are prevalent in argumentative essays. On the other hand, they used more lexically sophisticated words in argumentative essays compared to narrative essays like in the previous study mentioned above (i.e., Ansarin et al., 2021). The authors attribute this to the fact that argumentative essays involve a higher proportion of nominalization which is realized through derivational suffixes. In a Turkish university setting, Yıldız and Yeşilyurt (2017) explored the impact of task complexity and rhetorical mode of writing on lexical complexity on EFL learners' written production. Descriptive essays were reported to have higher lexical complexity including all three components compared to narrative essays.

As already stated above, lexical complexity is a multidimensional construct involving three main sub-constructs (i.e., LD, LV, and LS). However, most studies implemented within this area have not examined all of these three metrics. Based on his research synthesis and quantitative meta-analysis of task complexity and its influence on CALF (i.e., syntactic complexity, accuracy, linguistic complexity, and fluency) of written L2 production, Johnson (2017) concludes that LD and LS have been less frequently investigated compared to LV, and few studies have investigated more than one component of lexical complexity. Therefore, for a more balanced and holistic examination of lexical complexity, all three measures should be included in the scope of the investigation.

As seen, there seems to be a stronger need for a thorough and balanced examination of lexical complexity in a wider range of genres. For this reason, in this study, we primarily investigated lexical complexity with its three main dimensions in learner-generated descriptive, compare and contrast, advantage and disadvantage, and argumentative essays in a year-long Writing skills course in an ELT program at a state university in north-eastern Türkiye. Besides, we also examined if and to what extent lexical complexity varied in all these genres and across time.

With these in mind, we sought answers to the following research questions.

1. What is the level of lexical complexity in pre-service English language teachers' essays of descriptive, comparative and contrast, advantage and disadvantage and argumentative genres?
2. Does the level of lexical complexity vary across genres and time? If yes, to what extent?

## **METHODOLOGY**

### **Research Purpose and Design**

We employed a longitudinal exploratory approach and a quasi-experimental repeated measures design (Nunan & Bailey, 2009) for data collection to explore the level of lexical complexity in the essays generated by pre-service English language teachers. We also examined if and to what extent lexical complexity varies across genres and over time in the essays. To address the research gap stated above, we examined lexical complexity level via LD, LS and LV indices in essays of four different genres, namely descriptive ( $n=36$ ), compare and contrast ( $n=42$ ), advantage and disadvantage ( $n=34$ ), and argumentative ( $n=42$ ) written by junior ELT students ( $N=44$ ) in a year-long Writing skills course. The total number of essays was 229. Besides, as the essays were generated throughout the academic year, we aimed to capture if the level of lexical complexity changes from the very beginning to the end. Therefore, to shed more light on the likely change(s) and growth in the level of lexical complexity, we adopted a pre-test and post-test design which included assigning a descriptive essay about an indispensable person in the students' lives (pre-) and another descriptive essay about an important person for the students (post-). We intentionally selected the descriptive genre as we knew it is the most common and almost always the first genre learners study and practice in Writing courses. At the end of the course, we repeated the same task which we assigned as the pre-test. The number of essays we collected as a pre-task was 29, while it was 46 in the post-test.

### **Setting and Participants**

The study was conducted at a state university in a northeastern city of Türkiye. The Writing course (Advanced Writing) which is offered in the junior year for two hours in both fall and spring terms (28 weeks in total) in the English language teaching program constituted the main context. The PSTs have been through preparatory language education where they took a writing course to learn the basic principles of writing and how to write various types of paragraphs. The Writing course syllabus covers different types of essay genres such as description, comparison and contrast, advantage and disadvantage, and argumentative.

### **Data Collection**

The data for this study consists of a corpus of essays ( $N=229$ ) in four genres written throughout the course. To determine the level of lexical complexity in the essays before the instruction, we asked the students to write an essay in a descriptive genre ( $n=29$ ). We repeated the same task at the end of the entire instruction and collected the same essay ( $n=46$ ) as a post-test. The essays in between the pre-tests and post-tests were generated throughout the course. The essays written in descriptive ( $n=36$ ) and compare and contrast genres ( $n=42$ ) were collected in the fall semester, while others generated in advantage and disadvantage ( $n=34$ ) and argumentative ( $n=42$ ) genres were collected in the spring term. All essays were written using word-processing software and accepted through Turnitin so as to prevent plagiarism and

replicas among the classmates. As pre-test and post-test, they were asked to write a descriptive essay within the limit of a class hour which was 45 minutes.

### Data analysis

The essays were analyzed through a natural language processing tool named the single-mode web-based lexical complexity analyzer (LCA) developed by Ai and Lu (2010) and Lu (2012). Via this tool, we examined lexical complexity through lexical density (LD), lexical sophistication (LS), and lexical variation (LV) indices. For LD, there is only one metric in the LCA, and there are multiple metrics for LS and LV. For LS, we chose the lexical sophistication-1 (LS1) metric, and for LV the corrected type-token ratio (CTTR) metric. CTTR was reported to be one of the three transformations of the original TTR, along with RTTR and the D measure, which indicated the strongest effect in the previous research (Lu, 2012).

After analyzing the essays through LCA, we entered the scores of all measures (LD, LS, and LV) into SPSS for further examination through descriptive and inferential statistics. We had the PSTs' essays in four genres, which meant that we had multiple measures to examine lexical complexity development throughout the study. Therefore, to explore if and how lexical complexity changed in the PSTs' essays and if the changes across four genres were statistically significant, we performed inferential statistics. Before we did so, we checked the normality score in all measures across four genres.

**Table 1.** Shapiro-Wilk Test of Normality p Values for Three Indicators of Lexical Complexity across Four Genres

Measures	Pre-test	Descriptive	Compare - Contrast	Advantage - Disadvantage	Argumentative	Post-test
LD	.161	.002	.028	.221	.206	.223
LS	.669	.097	.004	.351	.707	.000
LV	.838	.162	.316	.946	.839	.583

As the table shows, the test showed that except the LD values of descriptive and compare and contrast essays, and post-test descriptive essay's LS Shapiro-Wilk p values ( $p < .05$ ), the rest indicated that the data in all these measures across all genres were normally distributed which required parametric tests for further inferential statistics. For those three measures mentioned above, as Field (2018) suggested we "look[ed] at the shape of distribution visually, interpret[ed] the value of the skewness and kurtosis statistics" (p. 393) which indicated normal distribution. Therefore, to have a deeper and closer examination of lexical complexity in the PSTs' essays over the course of the study and across genres, we continued with parametric tests and ran paired samples t-test to see if there was any statistically significant difference in the students' essays from the very beginning to the end, and one-way repeated measures of ANOVA (Field, 2018) for lexical complexity in the essays of four genres throughout the course.

## RESULTS

### Lexical complexity in learner-generated essays in four genres and over the course of time

In this section, we first present the results of lexical complexity indices, namely LD, LS and LV gained from the LCA (Ai & Lu, 2010; Lu, 2012) in the four genres.

**Table 2.** Lexical Complexity across Genres

Measures	Genre 1 (Descriptive)		Genre 2 (Compare - Contrast)		Genre 3 (Advantage - Disadvantage)		Genre 4 (Argumentative)	
	M	SD	M	SD	M	SD	M	SD
LD	.53	.24	.48	.22	.48	.60	.51	.28
LS	.26	.48	.23	.48	.42	.99	.24	.71
LV	6.4	.63	6.6	1.22	6.7	.78	6.8	.87

As the table presents, despite fluctuations, LD was quite similar across genres. It was the highest in descriptive essays. LS was almost the same in descriptive, compare and contrast, and argumentative essays, while it was the highest in advantage and disadvantage essays. On the other hand, LV displayed a steady increase across all genres.

To further examine if any of the lexical complexity indices changed statistically significantly across the genres, we ran one-way repeated measures of the ANOVA test. Mauchly's test revealed that the variances of differences were roughly equal and the sphericity assumption was met ( $p=.046>.05$ ) indicating that LD displayed a statistically significant difference across genres ( $F(3, 84)=29.279, p=>.05$ ).

Therefore, using the Bonferroni test, we performed post hoc comparisons which indicated that LD index in the descriptive essay ( $M=.53, SD=.24$ ) was statistically significantly different from the level of LD in the compare and contrast genre ( $M=.48, SD=.22$ ), and the level of LD in the advantage and disadvantage genre ( $M=.48, SD=.60$ ). Besides, the comparison revealed a statistically significant difference between the comparison and contrast ( $M=.48, SD=.22$ ) and the argumentative genre ( $M=.51, SD=.28$ ). Additionally, LD was found to be significantly different between the advantage and disadvantage ( $M=.48, SD=.60$ ) and the argumentative genres ( $M=.51, SD=.28$ ).

Additionally, we also examined if LS changed significantly across the genres. Mauchly's test revealed that the variances of differences were not equal and the sphericity assumption was not met ( $p=.005>.05$ ). Therefore, a Greenhouse-Geisser correction was applied ( $\epsilon = 0.722$ ) indicating that LS did not display statistically significant difference across the genres ( $F(2.166, 60.648)=2.102, p=>.05$ ). Lastly, the variances of difference were equal and the sphericity assumption was met ( $p=.148>.05$ ). However, no statistically significant change was found between the LV across genres ( $F(3, 84)=2.865, p=>.05$ ).

Besides, we also present the lexical complexity measures in pre- and post-test essays (see Table 3) to address if lexical complexity changed over the course of time as the students were engaged in the Writing skills course and learning and writing essays of different genres.

**Table 3.** Lexical Complexity over the Course of Time in the Writing Skills Course

Measures	Pre-test		Post-test	
	M	SD	M	SD
LD	.44	.27	.45	.26
LS	.25	.53	.27	.63
LV	5,3	.55	6,7	.63

As the table shows, LD and LS were found to be quite similar in both tests, while LV displayed a noticeable increase. To reveal if any of these changes were statistically significant, we ran a paired samples t-test. The test revealed no statistically significant difference in the LD indices in the pre-test ( $M=.44$ ,  $SD=.27$ ) and the post-test ( $M=.45$ ,  $SD=.26$ ),  $t(25)=-1.552$ ,  $p=.133$ . The LS indices of both tests also did not display any statistically significant difference between the pre-test ( $M=.25$ ,  $SD=.53$ ) and the post-test ( $M=.27$ ,  $SD=.63$ ),  $t(25)=-1.341$ ,  $p=.192$ . However, there was a statistically significant difference in the LV measure from the pre-test ( $M=.5359$ ,  $SD=.55$ ) and the post-test ( $M=.6783$ ,  $SD=.63$ ),  $t(25)=-13.231$ ,  $p=.000$ . These indicate that the density and the size of vocabulary in the essays did not differ significantly, while the diversity did.

## DISCUSSION and CONCLUSION

This study aimed to investigate lexical complexity indices in written discourse across four genres and over the course of time by analyzing essays written by the PSTs throughout an academic year. The initial focus of the study was to examine lexical complexity differences in terms of LD, LS and LV indices in four genres. The findings pointed out a statistically significant difference across genres with respect to LD. It scored the highest in the descriptive essay. The argumentative essay gained the second highest score, while compare-contrast and advantage-disadvantage essays had the same scores at the lowest level. This finding is similar to that of Yıldız and Yeşilyurt (2017), who reported that descriptive essays scored higher compared to narratives. They conclude that the reason for this finding is related to the fact that learners were required to write about more personal topics in descriptive compositions compared to other genres. This could also be explained by the advantage of the topic familiarity inherent in the descriptive essay prompt. With regard to this concept, Yang and Kim (2017) stated that the essays with a more familiar topic yielded a significantly higher lexical complexity score than those with a less familiar topic. Since the learners in our study were required to write about an important person in their life, they might have made use of a higher proportion of content words depicting the various properties of that person as opposed to function words. In another study, while the descriptive genre was not involved within the scope of the investigation, LD was reported to be higher in argumentative essays compared to expository ones (Heng, Pu, & Liu, 2023). As noticed, there seems to emerge a shortage of research examining LD, as a sub-construct of lexical complexity, in the descriptive genre in an effort to make a comparison to other genres. As such, it appears to be difficult to arrive at conclusive results regarding the predictive role of LD. Nonetheless, it is reported that rhetorical task or genre has an effect on LD, with higher levels in expository and argumentative essays than narratives (Ravid, 2004, as cited in Yang, 2014), yet excluding the case of descriptive essays from the conclusion made.



As for the findings concerning LS, another sub-construct of lexical complexity, while the level of LS was almost the same in the descriptive, compare and contrast, and argumentative essays, it was the highest in advantage and disadvantage essays. However, LS did not display a statistically significant difference across the genres. While it is not that possible to find a study which sought LS levels in multiple genres as we did in our study, Ansarin et al. (2021) found that LS gave the highest scores in the argumentative essays followed by narrative and descriptive ones. Likewise, LS in argumentative essays scored higher than in the narratives in Yoon and Polio (2017). Yıldız and Yeşilyurt (2017), on the other hand, concluded that descriptive texts were more lexically sophisticated than the narratives.

With reference to the LV measure, in spite of an observable and steady increase from the first genre (descriptive) to the last one (argumentative), no statistically significant change was found between the LV across genres. Despite the lack of a significant difference in LV in our study, narrative essays were reported to have higher LV or diversity than the argumentative ones in the two studies (Ansarin et al., 2021; Yoon & Polio, 2017). However, Yıldız and Yeşilyurt (2017) found that the descriptive essays were more lexically diverse than the narrative texts.

Finally, we assessed lexical complexity measures to capture the time impact on the learners' lexical development. The findings revealed that LD and LS indices did not indicate a significant difference throughout the study. However, a statistically significant change in the LV measure was detected over time. These findings indicate that learners' lexical development in terms of the density and the size of vocabulary reached a plateau, while their use of diverse vocabulary resulted in growth. The results concerning density and sophistication did not match the theoretical expectation that as learners become more proficient, they are expected to gain more lexical richness in their L2 language production (Spring & Johnson, 2022). On the other hand, it could be concluded that their capacity in using a more diverse range of vocabulary was enhanced. This finding is congruent with that of Crossley and McNamara (2012) in that as L2 learners become more proficient in language production, their lexical diversity level also increases. Mazgutova and Kormos (2015) also found that the lower proficiency group in their study improved significantly in terms of lexical diversity over the program. However, Kalantari and Gholami (2017) reported that only the index of LS was found to have improved in the study implemented with Iranian EFL learners. One explanation for the inconsistency of the research findings they offer is that LD, LS and LV are distinct sub-constructs of lexical complexity which evolve in their own developmental paths. This justification sounds reasonable for us as well to explain the nonparallel development in all three measures over the course of time.

The improvement in the LV index in our study might relate to the interaction of various factors. For instance, the learners' vocabulary size might have increased as a result of both incidental and intentional learning of lexis throughout their study in the ELT program. In addition, they could have been exposed to a variety of lexical items in their autonomous use of digital and online sources of multimedia in English. In other words, exposure to English at both non-academic and academic levels could explain the progress in their use of a range of diverse words.

## **IMPLICATIONS and LIMITATIONS**

The results obtained from this study carry some important implications with respect to future research and pedagogical practices. First of all, this study might contribute to the growth of literature on language acquisition or developmental studies within the realm of L2 writing research. More specifically, one main contribution of this study is to have examined all three main sub-constructs of lexical complexity. Indeed, according to Johnson (2017), studies have

mostly investigated LV or lexical diversity as a metric of lexical complexity. On the other hand, LD and LS have been less frequently examined. For this reason, we suggest that further research be implemented by the inclusion of all lexical complexity sub-constructs through alternating metrics for a more balanced and holistic examination of the construct. Another contribution of this study is that we measured lexical complexity indices across four different genres. To the best of our knowledge, most studies reviewed here have examined lexical complexity indices in one or two genres, most frequently in the argumentative genre (e.g., Mazgutova & Kormos, 2015; Saricaoglu & Atak, 2022, Yoon & Polio, 2017). In contrast, the descriptive genre has rarely been studied in this scope. To arrive at more conclusive results about the predictive role of genres in lexical complexity, researchers should direct their focus on other genres or rhetorical modes as well. The results of such research, then, can inform L2 writing pedagogy in delivering more fine-tuned instruction with respect to teaching peculiar lexical and syntactic features of different genres. Finally, it is recommended that other indicators of language development like accuracy, fluency, and syntactic complexity be also examined along with lexical complexity in further research so as to capture a more holistic picture of linguistic development in language learning.

With regard to the pedagogical implications arising from this study, we might first suggest that L2 writing teachers should place much more emphasis on the teaching of a variety of genres in that writers need to have a command of “different structures to communicate different functions in different genres” (Yoon & Polio, 2017, p. 291). Secondly, in order for learners to gradually improve their vocabulary size and capacity of using more diverse and sophisticated lexical items in their written or spoken production, opportunities for both explicit and implicit vocabulary instruction should be maximized in L2 learning environments. In this vein, we would like to emphasize the importance of deploying effective vocabulary-building or activation techniques, especially in the prewriting stage so that learners can transfer that knowledge into productive vocabulary in writing. Thirdly, teachers need to track learners’ progress in language acquisition or evaluate the effects of the program or interventions they implement. To complement the traditional means of writing assessment and evaluation, teachers ought to be introduced to and trained in using natural language processing (NLP) tools which provide practical, reliable, and objective results that inform them about essay quality, learners’ proficiency levels, or how much progress they have made in language acquisition over time.

One main limitation of this study is that we cannot ascertain if the order of teaching the four genres affected the findings belonging to each genre or if the findings were able to reflect peculiar features of a specific genre irrespective of time effect on language acquisition. There remains a question about whether the findings could have been different if the order of genres had been changed for counterbalancing the time impact. Owing to the lack of a control group in the research setting, we could not adopt a true experimental research design to counterbalance the time impact regarding the sequence of genres. In addition, the findings obtained from this study are limited to the selected lexical complexity metrics and the natural language processing tool.

## REFERENCES

- Ai, H., & Lu, X. (2010, June). A web-based system for automatic measurement of lexical complexity. Paper presented at the *27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10)*. Amherst, MA. June (pp. 8-12).
- Ansarin, A., Karafkan, M. A., & Hadidi, Y. (2021). The effects of task type on Iranian EFL learners’ use of lexical diversity and sophistication. *Applied Research on English Language*, 10(4), 39-70.

- Arslan, R. Ş. (2013). An integrated approach to enhancing prospective English language teachers' writing skills. *Journal of Language and Linguistic Studies*, 9(2), 1-17.
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA* (pp. 21-46). UK: John Benjamins.
- Cheung, Y. L. (2011). Teacher training for effective writing instruction: Recent trends and future directions. *Procedia Social and Behavioral Sciences*, 15, 531-534.
- Crossley, S.A., & McNamara, D.S. (2012). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35, 115-135. <https://doi.org/10.1111/j.1467-9817.2010.01449.x>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). London, England: SAGE Publications.
- Gao, J., & Min, S. (2021). A comparative study of the effects of L1 and L2 prewriting discussions on L2 writing performance. *System*, 103, 102654.
- Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia - Social and Behavioral Sciences*, 198, 546-556. <https://doi.org/10.1016/j.sbspro.2015.07.477>
- Ha, M. J. (2022). Syntactic complexity in EFL writing: Within-genre topic and writing quality. *Computer Assisted Language Learning Electronic Journal (CALL-EJ)*, 23(1), 187-205.
- Han, J., & Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. *Journal of Second Language Writing*, 40, 44-59. <https://doi.org/10.1016/j.jslw.2018.03.001>
- Heng, R., Pu, L., & Liu, X. (2023). The effects of genre on the lexical richness of argumentative and expository writing by Chinese EFL learners. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1082228>
- Hirvela, A., & Belcher, D. (2007). Writing scholars as teacher educators: Exploring writing teacher education. *Journal of Second Language Writing*, 16, 125-128. <http://dx.doi.org/10.1016/j.jslw.2007.08.001>
- Hsieh, W. M., & Liou, H. C. (2008). A case study of corpus-informed online academic writing for EFL graduate students. *CALICO Journal*, 26(1), 28-47.
- Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing*, 37, 13-38.
- Kalantari, R., & Gholami, J. (2017). Lexical complexity development from Dynamic Systems Theory perspective: Lexical density, diversity, and sophistication. *International Journal of Instruction*, 10(4), 1-18. <https://doi.org/10.12973/iji.2017.1041a>
- Kessler, M., Polio, C., Xu, C., & Hao, X. (2020). The effects of oral discussion and text chat on L2 Chinese writing. *Foreign Language Annals*, 53(4), 666-685. <https://doi.org/10.1111/flan.12491>
- Lee, C., Wong, K. C. K., Cheung, W. K., & Lee, F. S. L. (2009). Web-based essay critiquing system and EFL students' writing: a quantitative and qualitative investigation. *Computer Assisted Language Learning*, 22(1), 57-72. <https://doi.org/10.1080/09588220802613807>
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96 (2), 190-208. [https://doi.org/10.1111/j.1540-4781.2011.01232\\_1.x](https://doi.org/10.1111/j.1540-4781.2011.01232_1.x)
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85-104. <https://doi.org/10.1191/0265532202lt221oa>

- Manchon, R. M., & de Haan, R. (2008). Writing in foreign language contexts: An introduction. *Journal of Second Language Writing*, 17(1), <https://doi.org/10.1016/j.jslw.2007.08.002>
- Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, 29, 3-15. <https://doi.org/10.1016/j.jslw.2015.06.004>
- Meza, A., Rodriguez, I., & Caviedes, L. (2021). Fostering EFL preservice teachers' academic writing skills through reflective learning. *Profile Issues in Teachers' Professional Development*, 23(1), 89-106. <https://doi.org/10.15446/profile.v23n1.85145>
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, 100511. <https://doi.org/10.1016/j.asw.2020.100511>
- O'Leary, J. A., & Steinkrauss, R. (2022). Syntactic and lexical complexity in L2 English academic writing: Development and competition. *Ampersand*, 100096. <https://doi.org/10.1016/j.amper.2022.100096>
- Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, 29, 82-84. <http://dx.doi.org/10.1016/j.jslw.2015.06.008>
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Reichelt, M. (2019). Preparing teachers to teach writing in various English as a foreign language contexts. In L. Seloni, & S. Henderson Lee (Eds.), *Second language writing instruction in global contexts: English language teacher preparation and development* (pp. 288-304). Bristol: Multilingual Matters.
- Saricaoglu, A., & Atak, N. (2022). Syntactic complexity and lexical complexity in argumentative writing: Variation by proficiency. *Novitas- ROYAL (Research on Youth and Language)*, 16(1), 56-73.
- Saricaoglu, A., Bilki, Z., & Plakans, L. (2021). Syntactic complexity in learner-generated research paper introductions: Rhetorical functions and level of move/step realization. *Journal of English for Academic Purposes*, 53, 101037. <https://doi.org/10.1016/j.jeap.2021.101037>
- Schnur, E., & Rubio, F. (2021). Lexical complexity, writing proficiency, and task effects in Spanish Dual Language Immersion. *Language Learning and Technology*, 25(1), 53-72.
- Seidinejad, L., & Nafissi, Z. (2018). Developing lexical complexity in EFL students' essays via creative thinking techniques. *Pertanika Journal of Social Sciences and Humanities*, 26(3), 1697-1712.
- Smirnova, E. A. (2017). Using corpora in EFL classrooms: The case study of IELTS preparation. *RELC Journal*, 48(3), 302-310. <https://doi.org/10.1177/0033688216684280>
- Spring, R., & Johnson, M. (2022). The possibility of improving automated calculation of measures of lexical richness for EFL writing: A comparison of the LCA, NLTK and SpaCy tools. *System*, 106, 102770. <https://doi.org/10.1016/j.system.2022.102770>
- Thornbury, S. (2002). *How to teach vocabulary*. Pearson Education Limited.
- Uzun, K. (2019). *Genre-based instruction and genre-focused feedback: A multiperspective study on writing performance and the psychology of writing*. [Unpublished Doctoral Dissertation], Çanakkale Onsekiz Mart University. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawaii Press.

- Wu, M. Y., Steinkrauss, R., & Lowie, W. (2023). The reliability of single task assessment in longitudinal L2 writing research. *Journal of Second Language Writing*, 59, 100950. <https://doi.org/10.1016/j.jslw.2022.100950>
- Yang, W. (2014). *Mapping the relationships among the cognitive complexity of independent writing tasks, L2 writing quality, and complexity, accuracy and fluency of L2 writing*. [Unpublished Doctoral Dissertation], Georgia State University. Retrieved from [http://scholarworks.gsu.edu/alesl\\_diss/29/](http://scholarworks.gsu.edu/alesl_diss/29/)
- Yang, W. & Kim, Y. (2020). The effect of topic familiarity on the complexity, accuracy, and fluency of second language writing. *Applied Linguistics Review*, 11(1), 79-108. <https://doi.org/10.1515/applirev-2017-0017>
- Yıldız, M., & Yeşilyurt, S. (2017). Effects of task planning and rhetorical mode of writing on lexical complexity, syntactic complexity, and overall writing quality of EFL writers' task performance. *Journal of Language and Linguistic Studies*, 13(2), 440-464.
- Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, 51(2), 275-301. <https://doi.org/10.1002/tesq.296>
- Zare, M., Bagheri, M. S., Sadighi, F., Rassaei, E., & Goel, P. (2021). An investigation of the linguistic complexity of IELTS writing topics based on the levels of discourse representation and the degree of meaning coding. *Cogent Education*, 8(1), Article 1868235 <https://doi.org/10.1080/2331186X.2020.1868235>
- Zhan, J., Sun, Q., & Zhang, L. J. (2021). Effects of manipulating writing task complexity on learners' performance in completing vocabulary and syntactic tasks. *Language Teaching Research* <https://doi.org/10.1177/13621688211024360>

**Nalan Bayraktar Balkir** is an assistant professor at the School of Foreign Languages at Çanakkale Onsekiz Mart University. Her main research interests include the psychology of language learning, teaching vocabulary and academic writing, global education, second language acquisition, and teaching Turkish as an FL.

Email: [nbayraktar@comu.edu.tr](mailto:nbayraktar@comu.edu.tr)

**Handan Celik** is an assistant professor at Trabzon University, Fatih Faculty of Education, Department of Foreign Languages Teaching, English Language Teaching Division. She currently teaches at undergraduate level. Her current research interests include pre- and in-service English language teachers' education, English for Specific Purposes, and teaching and researching language.

Email: [handanchelik@gmail.com](mailto:handanchelik@gmail.com)

**Gokhan Cepni** received his Ph.D. in English Language Teaching from Çukurova University in 2022. He is currently an Assistant Professor of English at the ELT Department, Faculty of Education, Trabzon University. His interest areas cover EFL, EFL Writing, Applied Linguistics and English Language Teaching.

Email: [gokhancepni@gmail.com](mailto:gokhancepni@gmail.com)