Tarbiat Modares
University

The Effect of Topic on EFL Writing by Vietnamese Tertiary Students: Insights from Combining a Lexical Richness Analysis with Student Self-Reports

Bao Trang Thi Nguyen*[1] iD , *Jonathan Newton[2]* iD , &
Tran Ngoc Quynh Phuong[3] iD

Abstract
Lexical richness in written language production has been examined in different contexts and in relation to different genres but research into the effects of within-genre topics on lexical richness is still limited and takes little account of learners' perspectives. The present study therefore sought to understand the effects of topics on lexical richness (lexical density, lexical diversity and lexical sophistication) in written essays by Vietnamese EFL learners. Data were collected from sixty-four English-major university students who each wrote four essays on four different topics across two genres (persuasive and descriptive) in their normal classes as progress tests. Ten students were also interviewed after they had completed all the essays to understand their perceptions of the writing topics. Lextutor's VocabProfile was used to obtain measures of lexical density, lexical diversity and lexical sophistication. The results showed mixed effects of writing topics on these three lexical measures. Learner perspectives provided insights into how cognitive and affective individual factors contributed to the complexity of topic influence. Pedagogical implications are discussed in regard to how to enhance productive lexical use in writing through task selection and instructional planning.

[1] Corresponding Author: Faculty of English, University of Foreign Languages, Hue University, 57 Nguyen Khoa Chiem, Hue, Vietnam; *Email: ntbtrang@hueuni.edu.vn*;
ORCID ID: https://orcid.org/0000-0002-5285-7255
[2] School of School of Linguistics and Applied Language Studies, Victoria University of Wellington, VZ 418, Von Zedlitz Building, 26 / 28 Kelburn Parade, Wellington, 6012, New Zealand,
Email: jonathan.newton@vuw.ac.nz; ORCID ID: https://orcid.org/0000-0001-7264-6829
[3] Faculty of English, University of Foreign Languages, Hue University, 57 Nguyen Khoa Chiem, Hue, Vietnam; *Email: quynhphuong47ltk@gmail.com*, ORCID ID: https://orcid.org/0000-0003-0825-7866

1. Introduction

As González (2017, p.1) points out, "writing converts our thoughts and ideas into text, a process that requires the purposeful choice and use of words". Given the important role of vocabulary in writing (Derakhshan & Karimain Shirejini, 2020; research has investigated the lexical richness of writing by L2 learners using measures such as lexical density, lexical diversity and lexical sophistication (Crossley, 2020; Read, 2000). One question this line of research addresses is how lexical richness correlates with measures of writing quality (e.g., Crossley, 2020; Gebril & Plakans, 2016; Gregori-Signes & Clavel-Arroitia, 2015; Higginbotham & Reid, 2019; Maamuujav, 2021; Zheng, 2016) and with learner proficiency (e.g., Crossley et al., 2011; Olinghouse & Wilson, 2013). Results have been mixed, thus calling for a more nuanced understanding of the multi-faceted nature of lexical richness in writing, as well as more research in different contexts and with different writing tasks and task conditions. As Read (2000, p.199) notes, "the choice of task has an effect on the kind of vocabulary elicited".

Although recent research has looked at how learners use vocabulary in different writing task genres (e.g., Frear & Bitchner, 2015; Li, 2000; Park, 2013), understanding of the effects of topic on lexical richness in L2 writing is limited and has not investigated learner perspectives. The present research seeks to address this gap. Firstly, it examined the effect of topic on the lexical density, lexical diversity and lexical sophistication in English as a foreign language (EFL) writing by Vietnamese EFL university learners, an underexplored group in research on vocabulary in written language production. Secondly, it explored the learners' perspectives on their writing on different topics and their lexical choices. Understanding learners' lexical choices "could be a first step in the development of a viable aid for learners evaluating their own writing" (Goodfellow et al., 2002, p.133).

2. Literature Review

2.1. Lexical richness: Lexical density, Lexical diversity And Lexical Sophistication

Lexical density refers to the ratio of lexical words (content words) to the total number of words in a given text (Ure, 1971). Lexical words are generally nouns,

adjectives, verbs, and adverbs whereas grammatical items or functional words include prepositions, articles, determiners, conjunctions, etc. If a text has more content words than another text, it usually contains more information, i.e., is denser (Laufer & Nation, 1995, p. 309). In effect, then, lexical density "is a measure of information in any passage of text, according to how *tightly* the lexical items have been *packed* into the grammatical structure" (Halliday & Martin, 1993, p. 76, italics added). Research has shown that lexical density is an indicator of writing proficiency (Kim, 2014), although some studies have failed to find a relationship between lexical density and writing quality (e.g., Lu, 2012; Nasseri & Thompson, 2021). Lexical density is also a key feature of academic writing (Biber & Gray, 2016).

Lexical diversity is another dimension of lexical richness which indicates how varied learners' use of lexis is (Malvern, et al., 2004). It is described as lexical "range" or "variety" of vocabulary in use in a text (McCarthy & Jarvis 2007, p. 459). While lexical density refers to the density of information in a text, lexical diversity describes "non-repetitive"/unique words in use (Nasseri & Thompson, 2021, p.2).

Lexical diversity is traditionally defined as the ratio of different word types (T) over the total number of tokens (T) of a given text, or type-token ratio (TTR) (Templin, 1957). For example, the sentence *"She thinks Facebook is a good way of communication, so she uses Facebook quite often"* has 15 words (tokens) but 13 types (13 different words) because 'Facebook' and 'she' has been used twice. The greater the TTR, the richer one's lexical resource supposedly is. Although TTR has been used widely, its reliability is influenced by text length (Malvern et al., 2004; Park, 2013). Alternative measures of lexical diversity have been developed such as Mean Segmental TTR (MSTTR) (Johnson, 1944), Corrected TTR (CTTR) (Carrol, 1964), Root TTR (RTTR; Guiraud, 1960, as cited in Lu, 2012), logarithmic TTR (LogTTR) (Herdan, 1964) or D/ vocD (Malvern et al., 2004). However, these measures have also been criticized for various weaknesses and so currently "there is no consensus among researchers concerning a single best measure" (Lu, 2012, p.194). In the present study, we use the traditional measure of TTR as used in previous research (e.g., Park, 2013; Ryoo, 2018).

Lexical sophistication is defined as "the proportion of relatively unusual or advanced words in the learner's text" (Read, 2000, p. 203). Laufer and Nation (1995) introduced the Lexical Frequency Profile (LFP) to gauge the size of productive vocabulary of L2 learners and measure their language proficiency based on their language production (written/spoken). LFP is calculated as the proportion of higher and lower frequency words by reference to "established lexical frequency lists" (Ryoo, 2018, p. 64). In essence, LFP takes learners' written texts as "raw input" and derives their LFPs in measures of different "frequency bands" or word frequency lists (Meara, 2005). LFP consists of four-word frequency lists: i) the first 1000 most frequent words (K1), ii) the second 1000 frequent words) (K2), iii) the academic word list (AWL) (Coxhead, 2000) and iv) words that are not on these lists (off the list words) (OLW). According to some researchers (e.g., Lu, 2012; Ryoo, 2018), a sophisticated or complex word is one which does not belong to the K1/K2 category, and lexical sophistication is measured by means of the proportion of AWL words and OWL words out of the total number of words in a written text. Ryoo (2018, p.69) argues that "LFP is a valid measure of lexical richness because it relates the words in texts to their frequency in a corpus."

2.2 Topics and Lexical Richness

Task-related factors that influence the lexical features of L2 writing include topics, prompts, and genres. Studies have found that different writing task types or genres influence lexical diversity and sophistication in writing (Li, 2000; Park, 2013; Sadeghi & Dilmaghani, 2013). For example, Li (2000) found that interactive email writing was more likely to lead to higher lexical diversity in English as a second language (ESL) learners' writing performance than other types of email writing. Sadeghi and Dilmaghani (2013) found that Iranian EFL learners tended to use more complex and diverse lexical items in comparative than argumentative essays. Park (2013) found Korean EFL university students used more diverse words in narrative than argumentative texts, though the latter task elicited greater mean word length values.

Kormos (2011) investigated the effects of task prompts on the linguistic features of the English L2 narrative texts written by Hungarian school students at an upper-

intermediate level of English proficiency. She found that the presence or absence of predetermined content affected students' writing performance, particularly on measures of lexical diversity and sophistication. Providing content led to writing that was more complex, as measured by frequency of abstract words.

The topic per se of the writing task might incline students to use words and structures in different ways (e.g., Lee et al., 2021; Yoon, 2017). Research on the effect of topics on productive lexical use has largely focused on syntactic complexity (e.g., Atak & Saricaoglu, 2021; Yang et al., 2015; Yoon, 2017). For example, Yoon (2017) found more complex language use in persuasive genre essays on the topic of part-time jobs than on the topic of banning smoking. Yang et al. (2015) found topics that differed in causal reasoning elicited different levels of syntactic complexity. Atak and Saricaoglu (2021) found that impersonal topics led to greater syntactic complexity while topics with greater cognitive demand elicited greater complexity.

These findings suggest the potential impact of topics on the linguistic dimensions of written scripts. Indeed, "if a learner does not know much about a topic, then it is likely that their knowledge of the vocabulary related to that topic is also not well known" (Coxhead, 2018, p.1). Yet, research on the potential impact of topics on lexical richness in written performance is still limited, with only a handful of studies on how topics impact upon students' EFL lexical richness in essay writing.

Some research has explored the impact of topics on the measure of lexical diversity (Yu, 2010) or lexical sophistication of vocabulary in use in writing (Guo et al., 2013), or both (Ryoo, 2018). These studies have shown mixed results of topic effects on different measures of lexical use. For example, Yu (2010) examined how topic and proficiency affected lexical diversity measured by D value in Michigan English Language Assessment Battery (MELAB) written performance. He found that topics that were impersonal and familiar induced greater lexical diversity than personal unfamiliar topics and lexical diversity predicts text quality. Though not focusing on the effects of topics, Zheng (2016) found that familiar topics elicited greater lexical diversity and sophistication. Overall, these studies show how

different topics induce different patterns of lexical use in students' written language production.

He and Shi (2012) researched the use of vocabulary in writing tests by ESL students in a Canadian college. They found an effect of task topics, that is, topical knowledge aided the development of ideas and a writing task that involved specific topical knowledge elicited fewer academic words than a general topic. Ryoo (2018) used VocabProfile to analyse lexical diversity and lexical sophistication in writing by Korean EFL university students. In Ryoo's (2018) study, lexical diversity was measured by the TTR index, and lexical sophistication by proportions of academic and uncommon words. The findings revealed a clear impact of task topics on those two lexical measures. Students employed significantly more diverse words and academic words when they wrote about the topic of climate change than about the topic of leadership.

In his meta-analysis of studies on task complexity and the complexity of L2 writing, Johnson (2017) noted that there are few studies on lexical sophistication and even fewer on lexical density, and that, furthermore, most of these studies measure lexical richness using only one dimension. Another gap we see is that research on this topic has not combined lexical richness analysis with any analysis of learners' perspectives on how they write in response to different topics. To address these gaps, the present study sought to investigate the impact of topics on the lexical richness of persuasive and descriptive essays written by Vietnamese EFL university students. Lexical richness was determined using measures of lexical density, lexical diversity, and lexical sophistication. Follow-up interviews with selected learners provided data on how the learners perceived the topics they wrote on. The study seeks to answer the following research questions:

1. How do different topics affect lexical richness (lexical density, lexical diversity, and lexical sophistication) in students' writing?

2. What are students' perspectives on how they write on different topics and the relationship between these topics and their vocabulary use?

3. Methodology

3.1. Data Collection

3.1.1 Participants

Sixty-four Vietnamese third-year English-major students from the three writing classes taught by the first author at a university in central Vietnam participated in this research. They were informed of the research and provided consent for the data (de-identified) to be used for this study. Most of the 64 participating students were female (55), which reflects the dominance of female students in English courses in Vietnam. The students were all aged 20 and in the same year level and writing course. While they all had passed a prior writing course and had achieved at least the target learning outcome of B1 level on the Common European Framework of Reference (CEFR), the first author observed in her teaching of the students that their essay writing proficiency levels varied from B1 to C1.

3.1.2. Writing tasks

For the writing tasks the students each handwrote two persuasive essays and two descriptive essays, each on a different topic as summarized below (See Appendix 1 for all full writing tasks):

Persuasive essay 1: Express a viewpoint on the impact of Facebook (henceforth the Facebook task)

Persuasive essay 2: Express a viewpoint on the role of school/family in individual development (henceforth the school task).

Descriptive essay 1: Describe a university (henceforth the University task)

Descriptive essay 2: Describe a supermarket in students' city/town (henceforth the Supermarket task)

There was a one-week interval between each essay. The learners wrote each of the essays as progress tests. They were required to write about 250 words per essay in 45 minutes in their normal class hours. No reference to any resources was allowed during their writing. The specified word length of 250 words was selected because texts should be at least 200 words to provide reliable results in terms of

vocabulary use (Laufer & Nation, 1995). Since the potential effect of topics might compound the effects of genres, the effects of topics were only examined within the same genre.

Quite a few studies on L2 writing complexity (e.g., Bestgen, 2017; Chen & Liu, 2022; Kim et al., 2018) have drawn their raw data from existing corpora of written English (e.g., International Corpus of Learner English (ICLE), Writing Corpus of English Learners, the Crossley Written Lexical Proficiency Corpus). We chose not to use existing corpora for the current study as our research required essays to be written on different topics by the same group of learners and under the same conditions, and for us to be able to interview the learners about their writing experiences. To this end, our research responds to the call by Larsen-Freeman (2014) for research that employs authentic classroom data which need to be analyzed in their own right.

3.1.3. Data Set

The data set consisted of 256 essays, 64 on each of the four tasks. On average students wrote about 308 words per persuasive essay and 280 words per descriptive essay (see Table 1). The results revealed that students wrote a similar number of words on both topics for each genre. The results of paired-sample t-tests indicated there was no statistically significant difference between the average length of essays on the Facebook and school topics, t (63) = .023, p = .982, or between the average length of essays on the Supermarket and University topics, t (63) = 1.575, p =.120.

The written scripts were analysed to explore the effects of topics on lexical richness of Vietnamese EFL writing performance of the same writing type.

Table 1
Topic And Text Length

	Topic	Tokens			
		Min	Max	Mean	Std. Deviation
Persuasive essays	Facebook	210	574	308.73	65.41
	School	211	552	308.61	60.30
	University	182	461	277.59	65.60
Descriptive essays	Supermarket	187	541	288.62	68.50

3.1.4. Student interviews

Ten of the students (eight females and two males) were interviewed individually in Vietnamese on a voluntary basis. They were asked about how they perceived the writing tasks and their experience of writing the essays. Three interviews were face to face, four were via phone calls and three via Zalo chats due to logistic issues. Each verbal interview was audio recorded with students' permission and lasted approximately 30 minutes.

3.2. Data Analysis

3.2.1 Analysis of lexical richness

Hand-written texts that were unreadable were not used. The collected handwritten essays were de-identified, typed and saved as doc. files and checked carefully by the first author before being double-checked by another EFL teacher. Those that were below 200 words were not included. Following other research (e.g., Ryoo, 2018; Laufer & Nation, 1995), the doc. files were then "cleaned" by fixing obvious misspellings and, following Laufer & Nation (1995, p.315) proper nouns were deleted from the samples. As in many studies, incorrect use was not fixed. The cleaned files were then submitted to Lextutor's VocabProfile (Cobb, 2015), a free online program, to obtain word types, tokens, measures of LFP in four-word frequency bands including tokens and percentage of K1, K2, AWL and OWL words of each written text. Lextutor's VocabProfile (Cobb, 2015) have been used in in many studies (e.g., Maamuujav, 2021; Ryoo, 2018). In the present study, following these authors and others (e.g., Crossley et al., 2011; McNamara et al.,

2010), lexical sophistication was operationalised as the proportion of AWL and OWL words. Measures of lexical density (proportion of content words over the total number of tokens) and lexical diversity (TTR) were automatically derived from the software.

The data from Lextutor's VocabProfile were prepared in an Excel spreadsheet and double checked for accuracy by the first author and another EFL lecturer. It was then imported into SPSS (version 20) for analysis. Descriptive statistics were obtained; a series of paired-sample t-tests for the normally distributed data was conducted to compare the effects of topics on lexical measures. Normality of data was checked by Kolmogorov–Smirnov tests showing $p > .05$, indicating normal data distribution (Field, 2005). The significance (alpha) level was conventionally set at .05. The effect sizes of paired sample t-tests were also reported using eta squared values (η^2). According to Cohen (1988), η^2 of .01, .06 and .14 were considered small, medium, and large respectively.

3.2.2. Analysis of interview data

The interviews were transcribed and checked carefully before analysis by the first author and the EFL lecturer who cross-checked the accuracy of the quantitative data mentioned above. Analysis was conducted on transcripts in the original language, Vietnamese, as suggested by Casanave (2010), with the aim of understanding key emerging themes in students' personal accounts and perceptions of the writing topics. Emergent themes were generated, confirmed, and reconfirmed in the iterative open-coding process of analysis (Newman, 2014). According to Yin (2015), studies using such data should publish the original excerpts alongside the translated versions. However, limited space only allowed the latter to be displayed in this paper. The interview transcripts and quotes were double-checked by an EFL teacher who was experienced in doing EFL classroom-based research. Students were de-identified and coded as S1, S2, S3 and so on. Since the students were all aged 20 and in the same year level and writing course, only information about gender was provided alongside the student codes. (F for female and M for male).

Regarding inter-coder reliability, 30% of the interview transcripts were first

randomly selected and then coded independently for themes in an inductive manner by the first author and by an experienced researcher. The percentage agreement was 87% overall, which indicated satisfactory reliability, according to Yin (2015). Any differences were resolved through discussion between the two coders to reach agreement. For the very few themes whose meanings were unclear, and no agreement was reached, they were not included in the analysis. The first author then coded the remaining data.

4. Findings

4.1. Topics and Lexical Density

The descriptive statistics for lexical density across different topics are presented in Table 2.

Table 2
Lexical Density in Essays on Different Topics (N= 64)

	Min	Max	Mean	Std. Deviation	Paired-sample *t*- test		
					t	*df*	p
Facebook	.44	.62	.52	.04	-.065	63	.949
School	.47	.60	.52	.03			
University	.44	.60	.53	.04	-2.685	63	.009
Supermarket	.45	.62	.51	.04			

It can be seen that the mean percentage of content words was just slightly more than 50% in all the topics, suggesting students employed content words and functions words at a similar rate. Regarding the persuasive essay, the level of lexical density in the Facebook task and the School was similar, $t(63)$ = -.065, p = 0.949. However, the results of paired-sample t-test show that performance of the

University task was significantly more dense (M = .53, SD = .04) than that of Supermarket (M = .51, SD = 0.04), $t(63)$ = -2.685, p = 0.009 and this was a medium effect (η^2 = .103). In other words, task topics had selective effects on lexical density.

4.2. Topics and Lexical Diversity

The descriptive statistics for lexical diversity in different topics as measured by TTR are presented in Table 3. The results show an impact of topics on the value of TTR as a measure of lexical diversity. Specifically, students used more diverse words in the Facebook task (M = .53; SD = .06) than in the School task (M = .50, SD = .05) as confirmed by the results of a paired-sample t-test, $t(63)$ = 6.012, p<.001. This was a very large effect (η^2 = .365). Regarding the descriptive essays, the value of TTR was significantly higher in the University task (M = .54, SD = .05) than in Supermarket (M = .53, SD = .06), $t(63)$ = -2.395, p = .020, though the effect size was medium (η^2 = .083). In other words, topic had a clear effect on how varied vocabulary was in students' writing.

Table 3
Lexical Diversity in Essays on Different Topics Measured By TTR (N=64)

	Min	Max	Mean	Std. Deviation	Paired-sample *t*- test		
					t	*df*	p
Facebook	.40	.65	.53	.06	6.012	63	.000
School	.36	.62	.50	.05			
University	.44	.65	.54	.05	-2.395	63	.020
Supermarket	.36	.64	.53	.06			

4.3. Topics And Lexical Sophistication

Lexical sophistication was measured by AWL and OLW words and the LFP results

are summarised in Table 4.

Table 4
Lexical Sophistication in the Persuasive Essays on Different Topics (N= 64)

	FACEBOOK				SCHOOL			
	Min	Max	Mean	Std. Deviation	Min	Max	Mean	Std. Deviation
K1	73.50	86.64	81.85	2.91	81.45	92.70	87.44	2.82
K2	2.05	8.57	4.94	1.41	1.12	8.75	4.58	1.60
AWL	2.61	10.95	5.54	1.64	1.99	9.42	4.98	1.68
OLW	4.18	12.08	7.67	1.96	.66	6.65	2.99	1.27

The results indicated that performance of both topics of the persuasive essays (Facebook and School), contained a large majority of K1 words, 81.85% and 87.44 % respectively on average. Predictably, the frequency of academic words and advanced or uncommon words was far lower than K1 and K2 vocabulary in both tasks. In order to compare the mean percentage of AWL and OLW levels between the two tasks, paired-sample t-tests were conducted for the normally-distributed data, and the results revealed a significant effect of topics on the measure of lexical sophistication. In particular, the Facebook task elicited more academic (AWL) words (M = 5.54, SD = 1.64) than the School task, M = 4.98 (SD = 1.68), t(63) = 2.133, p = .037, with a medium effect size (η^2 = .067). The proportion of uncommon (OLW) words was significantly higher in the former, t(63) = 16.775, p <.001, η^2= .817. Note that the effect size was much larger in the case of uncommon words as a measure of lexical sophistication.

Table 5 presents lexical sophistication in the descriptive essays. As with the persuasive essays, the K1 words constituted the large majority of the total tokens in students' writing in both tasks. A number of paired-sample t-tests were run to compare the frequencies of word levels between the University and Supermarket tasks, and the results revealed that the Supermarket task elicited significantly more rare (OLW) words (M = 7.15, SD = 2.21) than the former task (M = 4.74, SD =

1.69), $t(63) = 7.120$, p <.001, $\eta^2 = .446$. Yet, students produced more academic (AWL) words when they described their university than a supermarket, $t(63)= -4.276$, p< .001, $\eta^2 = .225$. The effect sizes were very large. This shows different effects of task topics on the frequency of sophisticated words that students used.

Table 5
Lexical Sophistication in the Descriptive Essays on Different Topics (N= 64)

	UNIVERSITY				SUPERMARKET			
	Min	Max	Mean	Std. Deviation	Min	Max	Mean	Std. Deviation
K1	76.10	92.49	86.22	3.31	73.13	89.86	82.35	3.69
K2	1.44	8.37	4.89	1.43	3.88	14.35	7.30	2.11
AWL	1.50	9.98	4.14	1.75	.75	8.77	3.20	1.55
OLW	1.21	9.28	4.74	1.69	3.62	13.44	7.15	2.21

4.4. Students' Perspectives

From the students' experience, as revealed in the interview data, writing surfaced as a complex meaning-making process in which topic was a salient factor. The interviewed students commented that topic familiarity facilitated access to ideas for them to write:

Every student has a Facebook account and knows how to use it well. The Facebook topic is more familiar and easier than the School topic; it is easy to address the negative and positive effects of Facebook; I can think of them, give examples more easily. (S3, M)

It is easier to find ideas with the University task than the Supermarket because I know my university well. (S6, F)

Students went further to elaborate on the speed of access to idea generation with familiar topics such as Facebook and University: "*I can think of ideas quickly,*

straight away with the Facebook task and University task (S7, F). This was, in their perceptions, very important in reducing the 'cognitive burden' (Skehan, 1998) since they had to write under time pressure. Students also talked about the novelty associated with the school topic, which made it more challenging to develop an argument about compared to the Facebook topic. One student commented:

The school topic is newer, less familiar; it is difficult to think of arguments right away. (S4, M)

Clearly, topic familiarity, as in the Facebook and University tasks could facilitate ideation and so enhance lexical density (content). For many, a familiar topic gave them a sense of 'security' because they had 'things to say' within time constraint (45 minutes). As one student commented, "*These topics give ideas ready, I feel less stressed*" (S8, F). However, not all the students saw the ease of writing of familiar topics as a positive aspect:

"The Facebook task was familiar, so I needed to invest in time to make my essay new, by thinking of different or new ideas to say." (S2, F)

Because the Facebook task is familiar, it is a bit challenging to produce an original piece of writing (I mean to make my ideas different from others). I needed to provide more updated examples. (S3, M)

These comments show that for at least some of the students, topics that were too familiar made it harder for them to think of interesting and original ideas for their writing. The merits of topic familiarity were, it seems, perceived differently by individual students.

We now turn to students' views as to how topic familiarity might have expedited access to vocabulary to convey the intended meanings. The students had clear views on this:

I have more words to say in the Facebook and University topics than the other ones. (S6, M)

Finding words to write in the Facebook and University topics are easier. (S10, F)

They also saw that familiar topical knowledge also enabled them to shy away from repetitive words:

It is easier to avoid repetition of words in the Facebook task and University task." (S1, F)

To me, it is a bit more challenging to find different words to use in the School task because it is less familiar to me. (S9, F)

Students shared that they had access to "*good*" on-topic vocabulary needed to express their intended meaning in the familiar tasks. As the clock ticked on in a timed performance, many said that they had to drop their intended complex ideas for those that they were linguistically able to encode. Students emphasized that they constantly went back and forth in a problem-solving process between their intended meanings and the lexical resources they had available.

The benefits of topic familiarity recognized by the learners confirm the performance-based data which show greater lexical density and lexical diversity with the Facebook and the University task than the others. In other words, familiar topical knowledge generally aided idea generation and lexical richness.

Some students also mentioned the issue of appeal and interest, noting that the novelty of unfamiliar topics such as School and Supermarket was motivating:

The topic about School and personality education is novel, and as such more appealing to me to write about, but in order to provide good arguments, it is rather difficult. (S5, F)

The Supermarket topic is new to me, so more interesting, though it is more difficult to write." (S2, F)

Students' emotional engagement in the topic was also mentioned. While being more topically demanding, the Supermarket topic appeared to provoke a positive emotional attachment among students because supermarkets had only relatively recently been introduced in this small Vietnamese city:

"The topic of supermarket is more exciting; it is a place to relax, sometimes I go there to shop, or just to window-shop, play games and watch a movie!" (S4, M)

"I love spending time in the supermarket, especially when it is burning hot; it's

cool there." (S10, M)

Further textual analysis reveals that in supermarket-based essays, many topic-related words that fall into the OLW category, as shown by Lextutor's VocabProfile (Cobb, 2015) were used, such as *delicious, fascinating, cosmetics, conditioners, budget, discount, import, candy, noodles, booths, goods, vegetables, magnificent, pizza, hamburger, cookies, souvenirs, snacks, seafood*, etc. These words naturally add to the greater proportion of uncommon words in the Supermarket task performances.

The students commented that they did not explicitly intend to use advanced words ("*If you don't know (rare/uncommon words), you don't know.*") (S5, F). But they did attend to avoiding word repetition *("I often try to think of different words to use such as synonyms or antonyms")* (S9, F)*, which could show an insufficient lexical repertoire ("*Sometimes I don't know an exact synonym, I use similar words.*") (S9, F). Other students recounted they were overtaxed with finding ideas to say and so had to forgo attempts to use more varied lexical items. Generally, students perceived topical knowledge to 'enable' lexical diversity and sophistication, though the latter was not featured as strongly as the former.

Overall, the interview data revealed that the extent to which topics both constrain or facilitate the process of writing depends on students' familiarity with and interest in various topics. According to the students' own accounts, topic familiarity and interest influenced how they allocated attention to idea generation and to the lexical demands of the writing task.

5. Discussion

The current study investigated the lexical richness of the writing performance on different topics within the same genre by 64 Vietnamese EFL university students. It found mixed effects of topics on measures of lexical richness, namely lexical density, lexical diversity and lexical sophistication. The study also probed students' perceptions of their experience of writing on the different topics to provide additional explanatory insights.

The mixed effects of the topics on lexical richness of Vietnamese EFL writing warrants further comment. The non-significant difference in lexical density between the Facebook and School tasks could be because they are both persuasive essays which are academic in nature, and usually contain denser information than other types of writing (Biber & Gray, 2016). On the other hand, a more distinct effect of topic was found in the descriptive essay. Though of the same descriptive genre, the Supermarket task was more on the entertainment side in students' perceptions, with 'cool atmosphere', 'foods' and 'films' to enjoy, which could have entailed a lighter academic content than the University task. At first glance, this seems contradictory since students shared how topic familiarity made generating content easier in the Facebook and University topics than in the others. Given that writing is a meaning-making and problem-solving process (Belcher & Hirvela, 2001; González, 2017), it allowed students space to construct their own content, as some students shared. Their desire to find fresh ways to approach familiar topics and challenge themselves in argumentation would seem to influence the amount of information (lexical density) they want to pack into their essays. This could confound the effect of topic. But it is also the case that lexical density is influenced by the syntactic and cohesive properties of a piece of writing. As Laufer and Nation (1995, p. 309) point out, "Fewer function words may reflect more subordinate clauses, participial phrases and ellipsis, all of which are not lexical but structural characteristics of a composition." In other words, lexical density could be influenced by the syntactic and discoursal choices that the writer makes.

The findings related to lexical density contradict some previous studies which found that familiar topics generated more content words (e.g., Yu, 2010; Zheng, 2016). However, they are in accord with others which show that lexical density is not so strongly affected by topic familiarity (e.g., Lu, 2012; Nasseri & Thompson, 2021). The explanations above are tentative and await more research to gain better subtle understanding about lexical density in different intra-genre topics.

With regards to lexical diversity, the consistent effect of topic on TTR is likely to be a function of prior knowledge and topic familiarity associated with Facebook and University topics which facilitated access to more varied vocabulary, a point that students made in the interview. This finding is in line with previous studies which show how familiar topics lead to writing that has more varied vocabulary

(e.g., Cho, 2019; Ryoo, 2018; Yoon, 2017; Yu, 2010; Zheng, 2016).

The finding that students used more uncommon words in writing on the Supermarket topic than on the University topic is partially in line with other research which has shown greater lexical sophistication with topics that are more relevant to students (e.g., Lo & Hyland, 2007; Yu, 2010; Zheng, 2016). It is also likely that the greater proportion of uncommon words generated by the supermarket task is a simple reflection of the many goods/foods-related words intrinsic to the topic. Thus, both the intrinsic nature of topics and learners' familiarity and engagement with given topics are likely to have shaped the lexical qualities of the writing produced by students in the current study. Students' perspectives in the present study added more explanatory power to understand the impact of topics both cognitively and affectively in relation to their individual orientation and text realisations. Previous studies, though not focusing on lexical richness, shows that students' connection to the task/topic potentially impacts upon task/topic motivation (Cho, 2019; Yoon, 2017). Lambert et al. (2017) study also revealed that personally relevant topics were more engaging, and that topic motivation needs greater attention in research on the effects of task-internal factors on lexical features.

6. Implications and Conclusions

The study offers two implications for how to enhance productive lexical use in writing through task selection and instructional planning. Firstly, it shows that topic familiarity can be a double-edged sword. The interview data showed that while familiar topics can facilitate ideas generation and free attentional resources to attend to using more varied vocabulary, such topics can also be demotivating for some learners. The selection of task topics needs to take into consideration their value to students so as to motivate sustained interest and engagement in writing.

Second, since the Lexical Frequency Profile (LFP) is recommended to use in course design (e.g., Aluthman, 2017; Catalán & Llach, 2017), teachers can use Lextutor's VocabProfile (Cobb, 2015) to obtain information about their students' use of vocabulary in writing to plan appropriate instructional strategies that assist students to attend to lexical aspects of their writing performance. Equally, teachers

could guide students to use it to explore their own LFP. Demonstrating how to use different words is also necessary. Other software tools such as AntConc (Anthony, 2020) can provide focused feedback on use of certain lexical words in students' writing (Nguyen, 2021). Teachers can also raise students' awareness of "lexical aspects that are being overlooked in learner compositions" (González, 2017, p.15). Above all, lexical learning is incremental (Nation, 2013), thus necessitating that teachers provide on-going writing opportunities, using diverse topics to push learners to extend their productive vocabulary knowledge.

7. Limitations and further research

A limitation of the current study that warrants further development in future research concerns the operationalization of topic type in relation to the key learner variables of familiarity with and interest in given topics. To a considerable extent, it is difficult to control either of these variables in a large, varied class of learners, since each learner brings unique life experiences and interests with them to the classroom. Nevertheless, future research could consider manipulating this variable more systematically across genre. Next, the majority of the participants in the present study was female, and though this was a typical feature of English-major classes in Vietnam, future studies could investigate a variety of learner groups to understand how writing topics affect lexical richness in their written language production.

We see a number of other implications for further research. First, understanding the actual AWL and OWL words students employ could provide additional insights into patterns of lexical use by students. Second, future studies could explore lexical accuracy (Read, 2000) alongside the lexical dimensions used in the present study. As Nation and Webb (2011) note, the lexical errors students make in their writing influence our perception of lexical richness. Third, future research could use a more fine-grained approach by using multiple measures of lexical diversity and sophistication, for example examining verb variation, noun variation, adjective variation (see Lu, 2012) and word association (Garner et al., 2019). Finally, since writing proficiency is likely to correlate with the richness of the lexical resources that learners can draw upon in language production, research is needed to investigate how topic familiarity and language proficiency interact to influence

lexical dimensions of written performance.

In conclusion, the current study has contributed to our understanding of the effects of topic on lexical richness in EFL writing, and on how these effects are mediated by learner familiarity and personal engagement with the topics.

Acknowledgements

References

Aluthman, E. S. (2017). Compiling an OPEC word list: A corpus-informed lexical analysis. *International Journal of Applied Linguistics & English Literature*, *6*(2), 78–91. http://dx.doi.org/10.7575/10.7575/aiac.ijalel.v.6n.2p.78

Anthony, L. (2020). Resources for researching vocabulary. In S. Webb (Ed.), The Routledge handbook of vocabulary studies (pp. 561–590). Routledge. https://doi.org/10.4324/9780429291586

Atak, N., Saricaoglu, A. (2021). Syntactic complexity in L2 learners' argumentative writing: Developmental stages and the within-genre topic effect, *Assessing writing*, *47*, (100506). https://doi.org/10.1016/j.asw.2020.100506

Belcher, D., & Hirvela, A. (2001). Introduction. In D. Belcher & A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading–writing connections*. (pp. 1–14). University of Michigan Press.

Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, *69*, 65-78. https://doi.org/10.1016/j.system.2017.08.004

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press. https://doi.org/10.1093/llc/fqx060

Carroll, J.B. (1964). *Language and thought*. Prentice-Hall.

Casanave, C. P. (2010). Distancing: From real life experiences to final research report in qualitative inquiry with multilingual participants. In the 2010 Symposium on Second Language Writing, University of Murcia, Spain.

Catalán, R. M. J., & Llach, M. P. A. (2017). CLIL or time? Lexical profiles of CLIL and non-CLIL EFL learners. *System*, *66*, 87-99. https://doi.org/10.1016/j.system.2017.03.016

Chen, M., & Liu, Y. (2022). A comparative study of lexical richness in English writing by Chinese senior high school students. *Asia Pacific Journal of Education*, 1-16. https://doi.org/10.1080/02188791.2022.2100739

Cho, M. (2019). The effects of prompts on L2 writing performance and engagement. *Foreign Language Annals*, *52*, 576–594. https://doi.org/10.1111/flan.12411

Cobb, T. (2015). VP-compleat. Retrieved from http://www.lextutor.ca/vp/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238. https://www.jstor.org/stable/3587951

Coxhead, A. (2018). Vocabulary and second language writing. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp.1–6). John Wiley & Son. https://doi.org/10.1002/9781118784235.eelt0533

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415–443. https://doi.org/ 10.17239/jowr-2020.11.03.01

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, *29*, 243–263. https://doi.org/0265532211419331

Derakhshan, A., & Karimain Shirejini, R. (2020). An investigation of the Iranian EFL learners' perceptions towards the most common writing problems. *Sage Open*, *10*(2), 1–10. https://journals.sagepub.com/doi/full/10.1177/ 2158244020919523

Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing, 30*(4), 45–57. https://doi.org/10.1016/j.jslw.2015.08.009

Garner, J., Crossley, S., & Kyle, K. (2019). N-gram measures and L2 writing proficiency. *System, 80,* 176–187. https://doi.org/10.1016/j.system.2018.12.001

Gebril, A., & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for Academic Purposes*, *24*, 78–88. https://doi.org/ 10.1016/j.jeap.2016.10.001

González, M. C. (2017) The contribution of lexical diversity to college-level writing. *TESOL Journal*, *8*(4), 899–919. https://doi.org/10.1002/tesj.342

Goodfellow, R., Lamy, M. N., & Jones, G. (2002). Assessing learners' writing using lexical frequency. *ReCALL*, *14*(1), 133–145. https://doi.org/10.1017/S0958344002001118

Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analyzing 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

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgements of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing, 18*(3*)*, 218–238. https://doi.org/10.1016/j.asw.2013.05.002

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. Palmer. https://doi.org/10.4324/9780203209936

He, L., & Shi, L. (2012). Topical knowledge and ESL writing. L*anguage Testing*, *29*(3) 443–464. https://doi.org/10.1177/0265532212436659

Herdan, G. (1964). *Quantitative linguistics.* Butterworths.

Higginbotham, G., & Reid, J. (2019). The lexical sophistication of second language learners' academic essays. *Journal of English for Academic Purposes, 37*, 127–140. https://doi.org/10.1016/j.jeap.2018.12.002

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. https://doi.org/10.1016/j.jslw.2017.06.001

Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychological Monographs*, 56, 1–15. https://doi.org/10.1037/h0093508

Kim, J. Y. (2014). Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study. *English Teaching*, *69*(4), 27–50. https://doi.org/10.15858/engtea.69.4.201412.27

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal*, *102*(1), 120-141. https://doi.org/10.1111/modl.12447

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, *20*(2), 148–161.

Kosmos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, *20*, 148–161. https://doi.org/10.1016/j.jslw.2011.02.001

Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research*, *21*(6), 665–680. https://doi.org/10.1177/1362168816683559

Larsen-Freeman, D. (2014). Another step to be taken: Rethinking the endpoint of the interlanguage continuum. In Z. Han & E. Tarone (Eds.), *Interlanguage: Forty years later* (pp. 203-220). John Benjamins.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics,* *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Lee, C., Ge, H.,& Chung, E. (2021). What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong, *System*, *97*, (102461). https://doi.org/10.1016/j.system.2021.102461

Li, Y. (2000). Linguistic characteristics of ESL writing in task-based e-mail activities. *System,* *28*(2), 229–245. https://doi.org/10.1016/S0346-251X(00)00009-9

Lo, J., & Hyland, F. (2007). Enhancing students' engagement and motivation in writing: The case of primary students in Hong Kong. *Journal of Second Language Writing*, *16*, 219–237. https://doi.org/10.1016/j.jslw.2007.06.002

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

Maamuujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays, *Assessing Writing, 49*, (100540). https://doi.org/10.1016/j.asw.2021.100540

Malvern, D., Richards, B., J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan. https://doi.org/10.1057/9780230511804

McCarthy, P. M., & Jarvis, S. (2007). A theoretical and empirical evaluation of vocd. *Language Testing, 24,* 459–488. https://doi.org/10.1177/ 0265532207080767

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26(1),* 32–47. https://doi.org/10.1093/applin/amh037

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27,* 57–86. https://doi.org/10.1177/0741088309351547

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

Nation, I.S.P. (2013) *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.4236/ce.2016.73042

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle-Cengage.

Newman, L.W. (2014). *Social research methods: Qualitative and quantitative approaches (7th ed)*. Pearson.

Nguyen, B.T.T. (2021). Using AntConc software to facilitate feedback giving on students' use of lexical words in writing. *TESOL Journal, 12(3), 1–5*. https://doi.org/10.1002/tesj.597

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and

writing quality in three genres. *Reading and Writing*, *26*(1), 45–65. https://doi.org/10.1007/s11145-012-9392-5

Park, S. (2013). Lexical analysis of Korean university students' narrative and argumentative essays. *English Teaching*, *68*(3), 131–157. https://doi.org/10.15858/engtea.68.3.201309.131

Pupore, G. (2014). The influence of content on adult L2 learners' task motivation: An interest theory perspective. *The Canadian Journal of Applied Linguistics*, *17*, 69–90. Retrieved from https://journals.lib.unb.ca/index.php/CJAL/article/view/22660

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732942

Ryoo, Y. (2018). Comparing lexical diversity and lexical sophistication in Korean. *Multimedia-Assisted Language Learning*, *21*(3), 63–87. https://doi.org/10.15702/mall.2018.21.3.63

Sadeghi, K & Dilmaghani, S. K. (2013). The relationship between lexical diversity and genre in Iranian EFL learners' writings. *Journal of Language Teaching and Research,4*(2*),* 328–334. https://doi.org/10.4304/jltr.4.2.328-334

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press. https://doi.org/10.5070/L4111005027

Templin, M. (1957). *Certain language skills in children: Their development and interrelationships*: The University of Minnesota Press.

Ure, J. (1971). Lexical density and register differentiation. In G. Perren and J.L.M. Trim (eds), *Applications of Linguistics* (pp.443–452). Cambridge University Press.

Yang, W, Lu, X., & Weigle, S.C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing, 28*, 53–67. https://doi.org/10.1016/j.jslw.2015.02.002

Yin, R. K (2015). *Qualitative research from start to finish* (2nd edition). The

Guilford Press.

Yoon, H. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System, 66*, 130–141. https://doi.org/10.1016/j.system.2017.03.007

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*(2), 236–259. https://doi.org/10.1093/applin/amp024

Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, *56*, 40–53. https://doi.org/10.1016/j.system.2015.11.007

Appendix 1. Writing tasks

Persuasive essays

Task 1

Some people say that Facebook is a positive development that benefits humankind while some others say that it has negative effects on many aspects of life.

What is your opinion?

Provide reasons and examples to support your position.

You have 45 minutes to plan and write your essay. You should write about 250 words.

Task 2

"School plays a more important role than the family in shaping one's personality." What is your opinion?

Provide reasons and examples to support your position.

You have 45 minutes to plan and write your essay. You should write about 250 words.

Descriptive essays

Task 1

You want to introduce your university so that more will know about it. Write an article to a newspaper describing the university you are now attending. You can write about

- its location
- teachers, students (where they come from, their majors, etc.), learning/teaching activities/ extra-curricular activities

- services and facilities (classrooms, library, canteen, playground, clubs), etc.

You have 45 minutes to plan and write your essay. You should write about 250 words.

Task 2

Describe a supermarket that you know. You can write about its

- location
- goods
- services
- facilities
- staff, etc.

You have 45 minutes to plan and write your essay. You should write about 250 words.

About the Authors

Bao Trang Thi Nguyen works as a lecturer at the Faculty of English, University of Foreign Languages, Hue University, Vietnam. Her field of research is task-based language teaching and learning, learner proficiency and SLA. She has published a number of book chapters and articles in different journals such as *TESOL Journal, Language Teaching Research*, *Asia Pacific Journal of Education*, *Language Teaching for Young Learners* and *Language Related Research*.

Jonathan Newton is an Associate Professor and Programme Director for the Masters of Applied Linguistics/TESOL Programmes at the School of Linguistics and Applied Language Studies (LALS), Victoria University of Wellington, New Zealand.

Tran Ngoc Quynh Phuong graduated from University of Foreign Languages, Hue University, Vietnam with an B.A. in Teaching English as a Foreign Language (TEFL). She is doing her Master thesis on English language teaching and learning, focusing on corpus-based language learning research.