



Research Papers in Language Teaching and Learning

Vol. 13, No. 1, January 2023, 121-136

ISSN: 1792-1244

Available online at <http://rpltl.eap.gr>

This article is issued under the [Creative Commons License Deed. Attribution 3.0 Unported \(CC BY 3.0\)](#)

## Syntactic pattern density, connectives, text easability, and text readability indices in students' written essays: A Coh-Metrix analysis

Tlatso Nkhobo & Chaka Chaka

### Abstract

This study sets out to investigate the four main linguistic categories (i.e., syntactic pattern density, connectives, text easability, and readability) together with their related sub-categories and indices in first-year undergraduate university students' written argumentative essay responses as analysed by Coh-Metrix. Employing a mixed-methods approach, the study utilised these written essay responses (n = 14 text files in two equal sets) as its corpora. Four of the findings of this study are worth mentioning. Firstly, concerning syntactic pattern density, all text files exhibited a high incidence of noun phrase density and verb phrase density. Secondly, both sets of text files displayed a lower usage incidence of adversative and contrastive, expanded temporal, and temporal connectives. Thirdly, regarding text easability, most text files in the first set and all text files in the second set had a low narrativity incidence. Fourthly and lastly, both the Flesch-Kincaid Grade Level and Coh-Metrix L2 Readability scores were low for both data sets.

**Keywords:** Coh-Metrix; students' written essay responses; syntactic pattern density; connectives, text easability; text readability.

---

### 1. Introduction

Advances in online technologies has enabled researchers to study massive datasets within a short period of time using online corpus linguistics software tools such as WMatrix (Howlett, 2019; Miura, 2020; Moghadam & Samar, 2020), AntConc (Papangkorn & Phoocharoensil, 2021; Ulfa & Muthalib, 2020), and AntWordProfiler (Halim, 2018; Indarti, 2021) in the process of analysing student writing samples. The main focus of the current study is to analyse students' written essay samples in terms of syntactic pattern density, connectives, text easability, and readability indices as generated by Coh-Metrix (see Azadnia et al., 2019; Graesser et al., 2014; Granados & Lorenzo (2021; McCarthy et al., 2006; Srisunakrua & Chumworatayee, 2019; Westerlund, 2019).

McNamara et al. (2010) define Coh-Metrix as a corpus analysis tool that is able to show cohesion and text difficulty at word and sentence levels in a written text. There has been an increasing number of researchers who have incorporated Coh-Metrix in their studies to analyse student writing samples (Chon & Shin, 2021; Kiselnikov et al., 2020; Latifi & Gierl, 2020; Mahadini et al., 2021; McNamara et al., 2010; Wang et al., 2021; Westerlund, 2019). Corpus linguistics is a strand of linguistics that deals

with corpora (McEnery & Wilson, 1996) by employing some of the corpus tools like the ones mentioned above. Infused into it is student writing analytics (Benetos & Bétrancourt, 2015; Conijn et al., 2018; Xu & Xia, 2019), which involves learning analytics through which student writing samples can also be analysed using some of the online software tools mentioned above.

Fourteen student writing samples (especially written essay responses) were used in the current study. They were retrieved from the Assignments Department for students registered for an undergraduate module, Academic Language and Literacy in English (ENG1503), offered by the Department of English Studies at an open and distance e-learning (ODEL) higher education institution in South Africa, in 2020. The main aim of this study is to analyse students' written texts through an online corpus linguistics software tool, Coh-Metrix, in terms of the syntactic pattern density, connectives, text easability and readability indices. The aforementioned tool has been employed in the South African higher education context to investigate students' written texts in few research studies. Therefore, the current study seeks to answer the following research questions:

- What is the frequency of the syntactic pattern density and connectives in students' written essay responses to Topics 1 and 2 as analysed by Coh-Metrix?
- What is the text easability and readability indices of students' written essay responses to Topics 1 and 2 as rated by Coh-Metrix?

## 2. Related Literature Review

### 2.1. Writing analytics

Consistent with the aim of the current study, which is to analyse students' written essay samples in terms of the usage incidence presented by means of syntactic pattern density, connectives, text easability and readability indices as displayed by the Coh-Metrix analysis software tool, writing analytics forms a significant part of this study. The aforesaid online software tool comprises algorithms that analyse student writing in terms of syntactic pattern density, text easability, and readability indices, which may not generate the same results as the other online software tools such as AntConc, WMatrix and AntWordProfiler due to their different algorithm configurations. In this instance, the studies conducted by Halim (2018) and Indarti (2021) employed AntWordProfiler in order to analyse student writing readability levels. The findings of both studies revealed that the readability indices of students' writing in each case varied slightly. For example, Halim's (2018) study discovered that the lexical richness of undergraduate students' written theses was low because their type token ratio (TTR) was represented below 1 as represented by the AntWordProfiler readability indices. This study attributed its findings to the repetitive vocabulary that students used in their written theses.

Similarly, Indarti's (2021) study found that female and male students' writings in the three semesters did not yield type token ratio (TTR) that was above 1. For instance, female students' TTR ranged from 0.55 to 0.60 across the three semesters. The study also established that male students' writing TTR ranged from 0.55 to 0.58 in the same three semesters. Overall, the findings of these two studies indicated that students did not use different vocabulary items in their written texts; hence the TTR scores that were below 1. This meant that the lexical richness of their respective students' written texts was low. In another scenario, Papangkorn and Phoocharoensil's (2021) and Ulfa and Muthalib's (2020) studies employed AntConc to investigate the similarities and differences of metadiscourse markers and lexical bundles that English (native) and L1 Thai speakers used in English argumentative texts. The two studies did so because AntConc is not configured to analyse and gauge the readability levels of written essays as employed in the current study. Papangkorn and Phoocharoensil's (2021) study found that the use of metadiscourse markers and participant-oriented bundles varied in terms

of their frequencies of use as represented by AntConc in both English (native) and L1 Thai students' written texts.

Furthermore, WMatrix was employed in AlZahrani and Othman's (2019) study to investigate the use of tense-aspect amongst Arabic-speaking EFL learners, whose written essays were compared. The study discovered that the use of progressive and activity verbs was statistically significant in students' written essays. WMatrix has also been used in other studies to investigate other aspects of English language usage in written texts. For example, Miura (2020) used it to examine verb usage, while Howlett (2019) and Moghadam and Samar (2020) employed it to investigate multi-word expressions and metaphor identification, respectively.

Writing analytics is part of learning analytics, and comprises, in this study, various data related to student writing that could be computationally analysed using writing software tools. As a subset of learning analytics, writing analytics is enhanced by corpus analysis software programmes in order to assess written data with the speed and efficiency that the human mind cannot process. Several studies have been conducted using this process, and many deployed keystroking to analyse and assess student writing for different purposes in teaching and learning (see Allen et al., 2016; Barkaoui, 2019; Benetos & Bétrancourt, 2015; Chaka & Nkhobo, 2019; Conijn et al., 2018; Conijn et al., 2019; Fontaine & Aldridge-Waddon, 2015; Gánem-Gutiérrez & Gilmore, 2018; Guo et al., 2019; Leijten et al., 2019; Nkhobo & Chaka, 2021; Sinharay et al., 2019; Van Waes et al., 2019; Xu & Xia, 2019).

In one instance, Guo et al. (2019) investigated the sub-group writing processes that selected participants applied as they produced written essays. A semi-Markov model was used for writing sequence. The students were requested to write an argumentative essay and perform other written tasks. Keystroke logs were used to model the writing processes of the participants on the same written task. The researchers focused on how long students paused and edited, and how long it took them to produce written texts. They found significant writing processes among the chosen groups. Students with a lower socio-economic status and black students showed lower efficiency in producing texts, compared to their counterparts with a higher socio-economic status and white students. Female students, compared to their male counterparts, were more fluent, and showed advanced typing skills. They also used more complex words and spent more time producing texts and editing. This is an indication that students at higher education institutions (HEIs) should be encouraged to structure and produce texts related to a given subject matter.

In another study, Fontaine et al. (2015) investigated the writing processes of 20 second-year students at Cardiff University. The students had to produce written texts: a Facebook message and a short-written essay. The aim of the study was to determine if the chosen participants employed different writing processes. The participants were all first-language English speakers and had been using a keyboard for the past six years. They were given seven minutes to complete each task. The researchers observed that writers with high keyboard efficiency edited their texts more. The students with low keyboard efficiency were more concerned about judgement and made more changes to their writing. The students employed similar writing processes in writing a Facebook message as they did producing a short-written essay. The challenge for many students at HEIs, especially those at the ODeL institution under study, is that a considerable number of them lack computer literacy skills, and this adds to the deficit view held by most academic literacy conformists and traditionalists. This implies that some students may not have the necessary resources for out-of-school literacies that would equip them with the required skills to thrive in teaching and learning situations at an ODeL institution such as the one which is part of this study.

Barkaoui (2019) conducted a study that utilised keystroke logs in investigating second language (L2) students' teaching and learning. He employed keystroke data from his two earlier studies (Barkaoui,

2014, 2015) to examine the forms of cognitive processes that sixty-eight (68) L2 students displayed in their keyboard skills. These skills pertained mainly to both pausing tendencies and L2 mastery when these students generated their own written texts. Two Test of English as a Foreign Language (TOEFL) tests, which consisted of computer-based writing tasks, were administered to these students. Three of the findings that emerged from this study are: students' overall pausing patterns did not vary significantly in relation to keyboard skills and L2 mastery; students who displayed significantly higher keyboard skills and L2 mastery needed lesser time to accomplish their written tasks; and students with a higher-order transition spent shorter pauses within paragraphs, while the converse was true for students who evinced a low-order transition. These findings can better help distance education HEIs understand how to allot the time their students might need for them to carry out pedagogical tasks as students tend to perform their tasks at varying learning rates.

## **2.2. Coh-Metrix**

Many studies have focused on the use of the Coh-Metrix corpus analysis software for investigating students' written samples pertaining to text easability, referential cohesion, lexical diversity, connectives, syntactic complexity, syntactic pattern density, word information, and readability. Among these studies are the following: Chon and Shin (2020); Chon et al. (2021); Crossley and Kyle (2018); Dela Rosa and Genuino (2018); Granados and Lorenzo (2021); Kiselnikov et al. (2020); Kremzer (2021); Latifi and Gierl (2021); Lim (2019); MacArthur et al. (2018); McNamara et al. (2010); McNamara et al. (2013); Nasser and Thompson (2021); Perin and Lauterbach (2016); Wang et al. (2021); Westerlund (2019); and Xu and Tang (2020). The contention of the present study is that the use of corpus analysis software tools such as Coh-Metrix can lead to a paradigm shift in which higher education student writing is not necessarily conceptualised from a deficit perspective (see Nkhobo & Chaka 2021).

McNamara et al.'s study (2013), which took place at the Mississippi State University, examined how linguistic, cohesive, and rhetorical features function in persuasive essay writing. In this study, 313 timed-essays were collected and assessed. One of the key findings of this study is that the quality of essay writing is commensurate with essay length, lexical specificity, lexical diversity, and syntactic complexity. Another study, Chon et al.'s (2021), investigated the importance of machine translation (MT) in L2 writing. Its 70 participants were native speakers of Korean, who were also L2 speakers of English. All of them were college of education students, who majored in English Language Teaching (ELT). They had to write 300-word English composition essays based on anecdotal pictures. The study discovered that MT enabled participants to produce sentences that were syntactically more complex. Another study by Wang et al. (2021) set out to assess the use of Coh-Metrix indices for scoring essay difficulty levels. The essays, which were clustered (cluster 1, cluster 2 and cluster 3 essays), were sourced from south-eastern region schools in the United States. One finding of this study is that cluster 1 and cluster 2 essays had higher readability scores. This meant that these particular essays displayed simpler structures. By contrast, the study discovered that cluster 3 essays exhibited a lower mean percentile, which showed that such essays consisted mainly of complex and unfamiliar structures. The main aim of the current study is to analyse students' written texts by means of the online software tool, Coh-Metrix, by investigating syntactic pattern density, connectives, text easability and readability indices as presented in students' written essays. As mentioned earlier, there are limited research studies conducted in the South African higher education context in which Coh-Metrix has been employed to analyse students' written texts. So, the current study serves to contribute to this paucity of research.

## **2.3. Student writing samples and readability indices – Coh-Metrix**

Several studies have analysed Coh-Metrix readability indices. Some of these studies include Crossley et al. (2011), Dela Rosa and Genuino (2017), Elgort (2017), Mahadini et al. (2021), McCarthy et al. (2022), Zeng and Shen (2016), and Zhao (2014). Other studies that have employed Coh-Metrix to generate readability indices in written texts include those conducted by Chon and Shin (2020), Crossley and McNamara (2014), Kisel'nikov et al. (2020), McNamara et al (2010), and McNamara et al. (2013).

One study, that of Elgort (2017), utilised Coh-Metrix to analyse written texts generated by postgraduate students by means of academic blog posts and written assignments. It was undertaken within 10 weeks under the auspices of a Master's programme in Applied Linguistics and Teachers of English to Speakers of Other Languages (TESOL, and comprised 38 participants based in East and Southeast Asia. This study found that student blog posts were readable and displayed a better sense of agency than written essays. It also discovered that student blog posts manifested less complex textual and lexical structures. In another study by McCarthy et al. (2022), Coh-Metrix and Gramulator were employed to analyse and evaluate linguistic structures in counter-argument paragraphs in 78 student argumentative essays. The paragraphs were analysed and evaluated for the readability and writing quality as generated by Coh-Metrix. One of the findings of this study is that both support paragraphs and counter-arguments did not differ much in their readability and writing quality. Based on this, McCarthy et al. (2022) contend that the paragraphs generated by students displayed some degree of consistency

Likewise, in the study conducted by Zhao (2014), Coh-Metrix was used to investigate the readability indices of written texts by comparing the lexical cohesion of both Chinese EFL learners and English native speakers. The written texts spanned three years at an undergraduate level of study but involved several distinct disciplines at the Huanghuai University in China. The study's prediction was that higher lexical syntactic analysis indices equated with higher readability of the text. Conversely, it predicted that low LSA indices amounted to text complexity and difficulty. In the current study, Coh-Metrix was used to generate the readability indices students' written essay responses.

### 3. Research Methodology

As mentioned earlier, the main aim of this study is to analyse students' written texts through Coh-Metrix, by investigating syntactic pattern density, connectives, text easability and readability indices as presented in students' written essays. Again as mentioned earlier, there is limited research conducted in the South African higher education context in which Coh-Metrix has been employed to analyse students' written texts. So, the current study serves to contribute to this paucity of research and to fill the existing gap. It does so by seeking to answer the following research questions:

- What is the frequency of the syntactic pattern density and connectives in students' written essay responses to Topics 1 and 2 as analysed by Coh-Metrix?
- What is the text easability and readability indices of students' written essay responses to Topics 1 and 2 as rated by Coh-Metrix?

#### 3.1. Study design

The present study was exploratory in nature (cf. Heigham & Croker, 2009; Riazi, 2016). As a result, it investigated syntactic pattern density, connectives, text easability, and readability as its key linguistic categories by analysing them through Coh-Metrix. It also examined the related sub-categories and indices. Riazi (2016) points out that exploratory research is undertaken with a view to investigating new and unexplored phenomena. This view is supported by Leavy (2017), who maintains that exploratory research is useful in investigating less-researched topics.



Exploratory research was appropriate for the current study, as very few studies have investigated linguistic categories displayed in student writing samples, especially at the institution under study. The present study followed a mixed-methods approach (cf. Christensen et al., 2015; Richards et al., 2012), which comprised qualitative and quantitative data (Riazi, 2016; Richards et al., 2012). The data were extracted from students' written essay samples. Qualitative data was in the form of the contents of students' written essay samples. By contrast, quantitative data consisted of:

- linguistic categories, syntactic pattern density, connectives, text easability, and readability;
- sub-categories associated with these main linguistic categories; and
- the way in which these linguistic categories and their subcategories were represented in indices, frequencies, or metrics.

Ivankova and Creswell (2009, p. 136) maintain that: "...mixed methods research is a research approach ... for collecting, analysing, and 'mixing' quantitative and qualitative data at some stage of the research process within a single study in order to understand a research problem more completely" (also see Christensen et al., 2015; Riazi, 2016; Richards et al., 2012).

### **3.2. Sampling**

This study used convenience sampling to select 14 students' written essay samples. As mentioned earlier, the students were first-year students registered for the ENG1503 module in Semester 2 of 2020. All of them spoke English as a second language. All the students who participated in this study were informed through an announcement on *myUnisa* (a UNISA's legacy learning management system) that their written samples would be used for research purposes. The announcement also included informed consent forms, which students had to complete. However, before the study was conducted, an ethical clearance was granted by the UNISA's College of Human Sciences Research Ethics Committee. Its certificate number is: 2017-CHS-026. Christensen et al. (2015) define convenience sampling as requesting people who are readily available and accessible to the researcher to participate in a study. And, Richards et al. (2012) note that convenience samples do not represent the wider population, as the participants are intentionally selected. For his part, Riazi (2016: 60) defines convenience sampling as a non-probability procedure wherein the researcher selects participants based on their availability or easy recruitment, rather than selecting them from a group of potential participants.

### **3.3. Data collection procedure**

The researcher requested 60 previously completed assignments — 30 for each assignment (i.e., Assignment 1 and Assignment 2) — for 2020 Semester 2 from the Assignments Department of the institution under study. From these, a total of 14 Assignment 2 scripts were analysed, which comprised 14 written essay responses that had at least 500 words. Assignment 2 consisted of two topics from which students were required to choose one. The essay responses were divided into Topic 1 and Topic 2. Topic 1 was phrased as follows: Write an essay in which you argue for or against a visible presence of the police in schools as one measure of curbing the scourge of violence. Topic 2 was formulated thus: Write an essay in which you discuss three negative effects of using drugs for mood or behaviour syndromes. Seven responses were selected for each topic.

### **3.4. Data analysis**

Fourteen students' written essay responses were analysed using the corpus analysis software tool, Coh-Metrix. These written essay responses, which were in two sets of responses as described above,

constituted the two sets of corpora for this study. As the corpora were originally in digital files (PDF files), they first had to be extracted and converted into plain texts (henceforth text files or TFs) using Microsoft Word. This was done for each of the text files in the two sets. Then, each text file was anonymised and cleaned for misspellings only. Where necessary, each text file was cleaned for non-English letters or strings. In addition, any missing full stops were added at the end of each sentence for each text file to allow the tool to identify sentence boundaries. Unnecessary gaps between words were also corrected. However, the other punctuation marks and any other grammatical errors were not changed. Thereafter, each text file was converted into Coh-Metrix-readable txt extension format using TextPad (see Azadnia et al., 2019; Khushik & Huhta, 2022; Nkhobo, 2022). After this text conversion process, all the text files were analysed in Coh-Metrix 3.0, through inputting them into this software programme. The quantitative classification and labelling of the measured Coh-Metrix linguistic categories (e.g., syntactic pattern density, connectives, text easability, and readability), together with their related sub-categories and indices (see Tables, 1, 2, 3, and 4), were automatically populated and computed for each text file. The analysed data sets were then exported to and saved in Microsoft Excel files (cf. Westerlund, 2019).

## 4. Findings

The findings of this paper are two-fold: findings for Topic 1 and findings for Topic 2. The discussion of the findings are also presented below.

### 4.1 Coh-Metrix Topic 1 findings

As mentioned above, seven students' written essay responses (in text files, herein shortened as TFs) to Topic 1 were subjected to a Coh-Metrix analysis. The analysis of the four main linguistic categories (i.e., syntactic pattern density, connectives, text easability, and readability) yielded the results, which were in the form of sub-categories and indices, as presented below.

TF1	TF20	TF31	TF42	TF79	TF91	TF95	
393.939	363.790	382.892	362.195	383.333	397.638	442.211	Noun phrase density
256.566	275.804	215.886	265.854	229.630	206.693	198.492	Verb phrase density
12.121	23.689	14.257	25.610	16.667	33.465	27.638	Adverbial phrase density
109.091	103.215	114.053	110.976	133.333	125.984	150.754	Preposition phrase density
22.222	10.152	12.220	12.195	9.259	3.937	10.050	Agentless passive voice density
18.182	8.460	6.110	8.537	14.815	13.780	5.025	Negation density
24.242	13.536	38.697	25.610	22.222	17.717	17.588	Gerund density
38.384	37.225	24.440	26.829	29.630	27.559	22.613	Infinitive density

Table 1. Syntactic pattern density

Table 1 above shows that all the text files had a high incidence of noun phrase density and verb phrase density, with TF95 and TF20 topping in each case. Adverbial phrase density incidence was low in all the text files. All the text files had a high incidence of preposition phrase density, but a lowest incidence of infinitive density, gerund density, agentless passive voice density, and negation density.

TF1	TF20	TF31	TF42	TF79	TF91	TF95	
28.283	21.997	24.440	26.829	35.185	23.622	25.126	Causal connective density
30.303	20.305	30.550	42.683	57.407	41.339	30.151	Logical connective density

8.081	13.536	12.220	17.073	22.222	11.811	5.025	Adversative and contrastive connective density
16.162	10.152	6.110	17.073	22.222	15.748	7.538	Temporal connective density
12.121	11.844	16.293	15.854	16.667	25.591	22.613	Expanded temporal connective density
44.444	40.609	85.540	62.195	55.556	41.339	55.276	Additive connective density
0	0	0	0	0	0	0	Positive connective density
0	0	0	0	0	0	0	Negative connective density

Table 2. Connectives

Table 2 illustrates that all the text files had a low and high usage of additive and logical connectives, with TF31 and TF42 having the highest usage of additive connectives respectively. Causal, adversative and contrastive, temporal, and expanded temporal connectives scored lowest in varying degrees across all the files. Positive and negative connectives had zero occurrences in all the files.

TF01	TF20	TF31	TF42	TF79	TF91	TF95	
22.960	62.550	13.790	39.740	51.600	40.130	55.960	Narrativity, percentile
58.320	1.830	68.080	66.280	46.810	46.410	4.550	Syntactic simplicity, percentile
84.130	89.25	74.220	55.960	37.830	45.220	83.890	Word concreteness, percentile
63.680	55.570	20.330	16.110	29.810	56.360	57.930	Referential cohesion, percentile
67.720	74.220	51.600	82.640	92.510	50.800	47.210	Deep cohesion, percentile

Table 3. Text Easability

As displayed above, Table 3 shows that three text files (TF20, TF79, and TF95) had a significantly high-to-moderate narrativity incidence in terms of percentiles. However, the other text files had low narrativity incidence percentiles. Concerning syntactic simplicity, again, three text files (TF01, TF31, and TF42) had a significantly high-to-moderate syntactic simplicity, while the other text files had a low syntactic simplicity usage. In relation to word concreteness, five text files displayed mostly higher percentiles compared to the other text files. Referential cohesion was mostly higher in four text files, but lower in the other related text files. Lastly, deep cohesion was higher in four text files, moderate in two text files, and low in one text file.

TF1	TF20	TF31	TF42	TF79	TF91	TF95	
51.683	42.781	38.893	65.5	47.856	63.991	46.069	Flesch Reading Ease
10.605	16.469	12.041	7.967	11.744	8.278	14.439	Flesch-Kincaid Grade level
17.705	15.058	11.963	15.672	14.472	17.010	16.742	Coh-Metrix L2 Readability

Table 4. Readability

As indicated in Table 4, the Flesch Reading Ease was high in three text files, but lower in the rest of the text files. Both the Flesch-Kincaid Grade Level and the Coh-Metrix L2 Readability were low in all the text files.

#### 4.2 Coh-Metrix Topic 2 findings

Similarly, for Topic 2, the seven students' written essay responses (in text files, herein shortened as TFs) to Topic 2 were subjected to a Coh-Metrix analysis. The analysis of the four main linguistic categories (i.e., syntactic pattern density, connectives, text easability, and readability) yielded the results, which were in the form of sub-categories and indices, as presented below.



TF48	TF51	TF67	TF69	TF77	TF78	TF92	
339.254	381.862	364.583	379.518	391.544	393.382	350.087	Noun phrase density
227.353	212.411	208.333	244.980	215.074	215.074	261.698	Verb phrase density
17.762	28.640	19.097	28.112	11.029	11.029	39.861	Adverbial phrase density
103.020	97.852	83.333	92.369	99.265	99.265	88.388	Preposition phrase density
0	7.160	5.208	6.024	3.676	3.676	8.666	Agentless passive voice density
8.881	0	6.944	8.032	3.676	3.676	5.199	Negation density
15.986	31.026	19.097	40.161	16.544	16.544	34.662	Gerund density
26.643	16.706	22.569	16.064	7.353	7.353	24.263	Infinitive density

Table 5. Syntactic pattern density

Table 5 above demonstrates that all the text files had a high incidence of noun phrase density and verb phrase density, in which TF78 and TF92 topped in each case. The third syntactic pattern category that had a high density is the preposition phrases. However, the other remaining syntactic pattern categories had a lower density, with both the agentless passive voice density and the negation density having the lowest density. Of these two syntactic pattern categories, TF48 and TF51 had a zero usage occurrence in each case.

TF48	TF51	TF67	TF69	TF77	TF78	TF92	
33.748	35.800	48.611	40.161	40.441	40.441	43.328	Causal connective density
49.734	66.826	45.139	70.281	69.853	69.853	62.392	Logical connective density
19.538	35.800	10.417	18.072	33.088	33.088	24.263	Adversative and contrastive connective density
7.105	21.480	15.625	18.072	14.706	14.706	24.263	Temporal connective density
17.762	16.706	17.361	8.032	20.221	20.221	17.331	Expanded temporal connective density
49.734	57.279	43.403	68.273	68.015	68.015	45.061	Additive connective density
0	0	0	0	0	0	0	Positive connective density
0	0	0	0	0	0	0	Negative connective density

Table 6. Connectives

As illustrated in Table 6, most text files had a high density of logical and additive connectives, with TF69 having the highest density of logical connectives, followed by TF77 and TF78 respectively. Causal, adversative and contrastive, temporal, and expanded temporal connectives had varying degrees of lower and lowest usage density across all the files, with temporal connectives having scored the lowest density. By contrast, positive and negative connectives had zero occurrences in all the files.

TF48	TF51	TF67	TF69	TF77	TF78	TF92	
40.900	28.430	27.430	48.010	24.830	25.140	47.610	Narrativity, percentile
49.200	82.120	37.830	70.540	88.490	88.880	51.600	Syntactic simplicity, percentile
45.620	63.310	8.380	44.040	13.350	12.920	12.510	Word concreteness, percentile
72.570	13.570	44.040	22.960	40.130	40.130	27.430	Referential cohesion, percentile
78.520	97.560	97.5	99.200	97.830	97.830	99.730	Deep cohesion, percentile

Table 7. Text Easability

Table 7 above indicates that all the text files had a low narrativity. In contrast, all the text files had a higher deep cohesion rate. Five text files (TF51, TF69, TF77, TF78, and TF92) had a high usage rate of syntactic simplicity. For word concreteness and referential cohesion, only one file in each case (TF51 and TF48) had a high usage, with the other text files in each category having the low usage.

TF48	TF51	TF67	TF69	TF77	TF78	TF92	
58.622	53.188	47.846	45.350	45.478	45.892	49.700	Flesch Reading Ease
9.906	9.198	11.520	11.692	10.670	10.511	11.636	Flesch-Kincaid Grade level
17.409	11.724	14.016	13.237	14.027	14.463	12.502	Coh-Metrix L2 Readability

Table 8. Readability

As illustrated in Table 8, the Flesch Reading Ease was high in two text files, but lower in all the other text files. Both the Flesch-Kincaid Grade Level and the Coh-Metrix L2 Readability were lower in all the text files.

The aim of the current study is to investigate students' written texts through the syntactic pattern density, connectives, text easability and readability indices. In terms of the syntactic pattern density, similar usage incidence was reported in which students' written essay samples exhibited noun phrases and verb phrases in the text files for both Topics 1 and 2. In relation to connectives, various forms of usage incidence were discovered. Causal connectives had a low usage incidence in certain text files for Topics 1 and 2. Logical connectives had a high usage incidence in some text files, while it had a low usage incidence in others. In contrast, temporal and expanded temporal connectives were low and had a usage incidence in all the text files for both Topics 1 and 2. In terms of text easability, narrativity differed in all the text files in relation to their usage incidence. Similarly, word concreteness and referential cohesion had varying forms of usage incidence in all the text files for both Topics 1 and 2. The readability indices also displayed different levels of readability in all the text files with concomitant varying incidence scores.

## 5. Discussion

The discussion of the findings presented below addresses the research questions of the current study. In addition, it follows the key themes of which the research questions consist.

### 5.1. Syntactic pattern density

As shown in the findings section above, all text files for Topics 1 and 2 displayed a high incidence of noun phrase density and verb phrase density. On a broader level, syntactic density reflects syntactic complexity, which often leads to difficult text comprehension. This is so because syntactic complexity is taken as a Coh-Metric measure in which sentences are embedded and have structurally dense and syntactically vague constituents. By contrast, syntactic simplicity leads to easy text comprehension (Azadnia et al., 2019; Graesser et al., 2004; McCarthy et al., 2006; Srisunakrua & Chumworatayee, 2019). So, at the level of noun and verb phrases, Topics 1 and 2 text files displayed a syntactic complexity, a factor which often leads to the text being syntactically difficult to understand. In relation to this aspect of syntactic pattern density, the findings of the current study tend to corroborate Kim's (2020) findings that reported a prevalence of noun phrases in the discrepant-score essays.

As highlighted earlier, all text files for Topics 1 and 2 text files had a low incidence of adverbial phrase density, agentless passive voice density, negation density, gerund density, and infinitive density. For instance, Westerlund (2019) maintains that a high incidence of adverbial phrase density may indicate that a text is syntactically complex. In the current study, the findings pertaining to this feature of the

syntactic pattern show that the text files in question contained a high degree of syntactic simplicity in their use of adverbial phrases. With regard to the use of the agentless passive voice Westerlund (2019) argues that this feature mainly indicates how the reader processes a text. Concerning the text files in question, the low incidence of the agentless passive voice forms equally reflects the manner in which the writers of these texts viewed the essay topics about which they were writing. They required active agents.

## 5.2. Connectives

In relation to connectives, all the text files for Topic 1 displayed a low and high usage of additive and logical connectives, while most text files for Topic 2 had a high usage incidence of logical and additive connectives. Furthermore, all text files for both topics exhibited a lower incidence of adversative and contrastive, expanded temporal connectives, and temporal connectives. However, there were zero occurrences for positive and negative connectives for both sets of text files in each topic. In this regard, a study by Granados and Lorenzo (2021) reported low causal and adversative/contrastive connectives in secondary school English L2 students' written texts in the first period (28.3). But these connectives gradually increased in periods two (34.3) and three (35.6), even though there was a decrease in extended temporal connectives. The current study was a one-off study that did not involve multiple periods. Often, logical connective indices underscore a semantic linkage between sentences and paragraphs in a text. For example, low scores of logical connectives demonstrate a low textual cohesion, whereas high scores of logical connectives reflect a high textual cohesion (Westerlund, 2019).

## 5.3. Text easability

As mentioned earlier, for Topic 1, more text files had a low narrativity and syntactic simplicity, whereas for Topic 2, all text files had a low narrativity, and most text files had a high syntactic simplicity incidence. In this context, Dela Rosa and Genuino's (2018) study reported a high narrativity incidence. A high narrativity, which also refers to word familiarity (Mahadini et al., 2021; Westerlund, 2019), indicates that a text uses everyday language and is, thus, easy to understand. Conversely, texts with a low narrativity such as those in Topics 1 and 2 show that they are unfamiliar and may be difficult to comprehend (cf. Dela Rosa & Genuino, 2018; Kremzer, 2021). Nonetheless, since these texts were argumentative essay responses and not story-like responses, it is expected that they would employ a lot more nouns, verbs, and in certain instances, modifiers to convey their thoughts, views, and actions (see Mahadini et al., 2021).

Additionally, the syntactic simplicity findings of Topic 2 text files are consistent with Wang et al.'s (2021) findings, which revealed that essays in Clusters 1 and 2 had a higher syntactic simplicity, which meant they were easier to understand as opposed to those in Cluster 3, which were more complex and difficult to understand. On the other hand, those text files that had a lower syntactic simplicity in both topics meant that they were more complex and, thus, not easy to understand. These particular findings are in accord with Dela Rosa and Genuino's (2018) study that had similar findings.

With reference to word concreteness, more text files for Topic 1 had a high occurrence of this category as compared to those for Topic 2 in which only one text file had a high incidence of this category. For referential cohesion, some of the text files for Topic 1 had a high incidence of it, while the other text files had a lower incidence of it. Most text files for Topic 2 had a low incidence of this category. Deep cohesion was, overall, high for Topic 1 text files, and even higher for all Topic 2 text files. Word concreteness is about using concrete words as opposed to using abstract words (Dela Rosa & Genuino, 2018). In this instance, more Topic 1 text files employed more concrete words that enable the reader to make a mental picture of what is written, something that is not the case with topic 2 text files. In

Kremzer's (2021) study, referential cohesion ranged from 16.35% to 77.04%. In the present study, referential cohesion ranged from 16.11 to 63.68 for Topic 1 text files, and from 13.57 to 72.57 for Topic 2. A high referential cohesion indicates that texts employ similar words and conceptual ideas, and a high deep cohesion shows that texts use connectives effectively (Dela Rosa & Genuino, 2018).

#### 5.4. Readability

As mentioned earlier, for Topic 1, the Flesch Reading Ease was high in three text files, but lower in the rest of the text files. Both the Flesch-Kincaid Grade Level scores and the Coh-Metrix L2 Readability scores were low in all the text files. For Topic 2, the Flesch Reading Ease was high in two text files, but lower in all the other text files. Concerning the Flesch-Kincaid Grade Level and the Coh-Metrix L2 Readability, the scores were lower in all the text files. For these three measures, the higher the scores, the easier the text is to comprehend (Greaser et al., 2014). Overall, barring the three text files for Topic 1 and the two text files for Topic 2, the rest of the text files for these three measures had low scores. This means that they were difficult to understand. In Kremzer's (2021) study, the Flesch Reading Ease ranged from 36.86% to 62.65%, which meant that the texts involved were relatively difficult and relatively easy to understand and process (also see Kiselnikov et al., 2020).

### 6. Conclusion, limitations, and recommendations

Overall, in terms of syntactic pattern density, all text files for Topics 1 and 2 exhibited a high incidence of noun phrase density and verb phrase density. This reflected some elements of syntactic complexity or text comprehension difficulty in this aspect of syntactic pattern density. In certain instances, a high syntactic density such as this demonstrates syntactic vagueness. By contrast, all text files for Topics 1 and 2 text files had a low incidence of adverbial phrase density, agentless passive voice density, negation density, gerund density, and infinitive density. All of this evinced a high level of syntactic simplicity for these subcategories of syntactic pattern. These two contrasting results of the syntactic pattern, the latter being one of the main linguistic categories that Coh-Metrix focuses on, indicate how Coh-Metrix differentially interprets these two features of students' written essay texts.

In respect of connectives, Topic 1 text files yielded mixed results, a low and high usage of additive and logical connectives as compared to Topic 2 text files, which had a high usage of these two types of connectives. At the same time, both sets of text files manifested a lower usage incidence of adversative and contrastive, expanded temporal connectives, and temporal connectives. These two sets of results demonstrate how students utilised connectives contrastingly in terms of additive and logical connectives, given the varying usage rates detected in this instance. They also indicate how students employed few adversative and contrastive, and expanded temporal connectives in both sets of text files. A high usage of connectives in a text reflects a high degree of a logical connection of sentences and ideas, while a low usage of connectives signals a low level of a logical connection.

Concerning text easability, most of the text files for Topic 1 and all text files for Topic 2 had a low narrativity incidence. This demonstrates that, overall, text files for these two topics contained unfamiliar words and, as such, they were difficult to process and comprehend. However, it is worth pointing out that all these text files were argumentative and not simply narrative in terms of their genres. Another noteworthy point is that all Topic 2 text files exhibited a high syntactic simplicity, a factor that emphasises their high text comprehension in this instance. For word concreteness, the two sets of text files, again, displayed contrasting results, with Topic 1 text files having a high incidence of this feature as compared to Topic 2 text files. This underscores the manner in which the first set of text files concretely presented their ideas. Similarly, the first set of text files had a higher usage of deep cohesion, which is an indicator for textual coherence, than the second set.

With regard to readability, the scores for the Flesch-Kincaid Grade Level scores and the Coh-Metrix L2 Readability were low in all Topic 1 and Topic 2 text files. Since this measure functions in an inverse way, this demonstrates the text difficulty of these two sets of files in relation to their readability index.

One of the limitations of the current study are the fewer text files it employed. The other limitation is that the study did not employ many different types of data sets, or different sets of text files. On this basis, it is recommended that future studies employ many text files, and focus on different sets of student written corpora. Notwithstanding these limitations, overall, the current study contributes to the current dearth of research that employs Coh-Metrix to analyse student essay writing samples in the South African higher education sector. It is also intended to fill the existing research gap in this ODeL arena in South Africa. Most importantly, the study contributes a South Africa perspective to the existing global research on the use of corpus analysis software applications to investigate higher education student essay writing samples.

## References

- Allen, L. K., Jacovina, M. E., Dascalu, M., Roscoe, R. D., Kent, K. M., Likens, A. D., & McNamara, D. S. (2016). {ENTER}ing the time series {SPACE}: Uncovering the writing process through keystroke analyses. Proceedings of the 9th International Conference on Educational Data Mining, (pp. 22-29). <https://files.eric.ed.gov/fulltext/ED592674.pdf>
- AlZahrani, A., & Othman, N. (2019). Lexical Aspect in the Development of the English Progressive Among Arab EFL Speakers. *Asiatic: IIUM Journal of English Language and Literature*, 13(2), 182-201. <https://journals.iium.edu.my/asiatic/index.php/ajell/article/view/1676>
- Azadnia, M., Lotfi, A. R., & Biria, R. (2019). A study of syntactic complexity via Coh-Metrix: similarities and differences of Ph.D. dissertations written by Iranian university students and English native speakers. *Research in English Language Pedagogy*, 7(2), 232-254. <https://dx.doi.org/10.30486/relp.2018.663453>
- Barkaoui, K. (2014). Examining the impact of L2 proficiency and keyboarding skills on scores on TOEFL iBT writing tasks. *Language Testing*, 31, 241-259.
- Barkaoui, K. (2015). Test-takers' writing activities during TOEFL-iBT writing tasks: A stimulated recall study (TOEFL iBT Research Report No. 25; ETS Research Report No. RR-15-04). Princeton, NJ: Educational Testing Service.
- Barkaoui, K. (2019). What can L2 writers' pausing behavior tell us about their L2 writing processes?. *Studies in Second Language Acquisition*, 41(3), 529-554. <https://doi.org/10.1017/S027226311900010X>
- Benetos, K., & Bétrancourt, M. (2015). Visualization of computer-supported argumentative writing processes using C-SAW. *Revista Română de Interacțiune Om-Calculator*, 8(4), 281-302. <http://archive-ouverte.unige.ch/unige:92518>
- Chaka, C., & Nkhobo, T. (2019). Online module login data as a proxy measure of student engagement: the case of myUnisa, Moya<sup>MA</sup>, Flipgrid, and Gephi at an ODeL institution in South Africa. *International Journal of Educational Technology in Higher Education*, 16(1), 38. <http://doi.org/10.1186/s41239-019-0167-9>
- Chon, T. V., & Shin, D. (2020). Direct writing, translated writing, and machine-translated writing: A text-level analysis with Coh-Metrix. *English Teaching*, 75(1), 25-48. <https://doi.org/10.15858/engtea.75.1.202003.25>.
- Chon, T. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine translated texts: Raters' assessment, linguistic complexity and errors. *System*, 96, 1-12. <https://doi.org/10.1016/j.system.2020.102408>
- Christensen, L. B., Johnson, R. B., & Turner, L. A. (2015). *Research Methods, Design, and Analysis, Global Edition*. New York: Pearson Education.
- Conijn, R., Roeser, J., & Van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing*, 32(9), 2353-2374. <https://doi.org/10.1007/s11145-019-09953-8>
- Conijn, R., Van der Loo, J., & Van Zaanen, M. (2018). What's (not) in a keystroke? Automatic discovery of students' writing processes using keystroke logging. Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge, (pp. 5-9). [http://solaresearch.org/uploads/LAK18\\_Companion\\_Proceedings.pdf](http://solaresearch.org/uploads/LAK18_Companion_Proceedings.pdf)



- Crossley, S. A., & Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing*, 38, 46-50. <https://doi.org/10.1016/J.ASW.2018.06.004>
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66-79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28(3), 282-311. <https://doi.org/10.1177%2F0741088311410188>
- Dela Rosa, J. P. O., & Genuino, C. F. (2018). Correlating writer's block and ESL learners' writing quality. *Indonesian Journal of Applied Linguistics*, 7(3), 604-612. <https://doi.org/10.17509/ijal.v7i3.9810>
- Elgort, I. (2017). Blog posts and traditional assignments by first- and second-language writers. *Language Learning & Technology*, 21(2), 52-72. <https://dx.doi.org/10125/44611>
- Fontaine, L. M., & Aldridge-Waddon, M. (2015). The impact of mode on writing processes: A cognitive functional perspective on student writing. *LyCE Estudios*, 17, 9-34. <http://ffyl.uncuyo.edu.ar/upload/lyce-estudios-17-2014.pdf>
- Gánem-Gutiérrez, G. A., & Gilmore, A. (2018). Tracking the real-time evolution of a writing event: Second language writers at different proficiency levels. *Language Learning*, 68(2), 469-506. <https://doi.org/10.1111/lang.12280>
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse'. *The Elementary School Journal*, 115(2), 210-229.
- Granados, A., & Lorenzo, F. (2021). English L2 connectives in academic bilingual discourse: A longitudinal computerised analysis of a learner corpus. *Revista signos*, 54(106), 626-644. <http://dx.doi.org/10.4067/S0718-09342021000200626>
- Guo, H., Zhang, M., Deane, P., & Bennett, R. E. (2019). Writing process differences in subgroups reflected in keystroke logs. *Journal of Educational and Behavioral Statistics*, 44(5), 571-596. <https://doi.org/10.3102%2F1076998619856590>
- Halim, S. W. (2018). Lexical richness in English language and culture department students' undergraduate theses. *Journal of English Language and Culture*, 8(2), 140-151. <http://dx.doi.org/10.30813/jelc.v8i2.1098>
- Heigham, J., & Croker, R. (2009). *Qualitative Research in Applied Linguistics: A Practical Introduction*. London: Springer.
- Howlett, G. (2019). How Thai students use mobile devices when learning EFL and the effect of urban/rural school location. *Asian EFL Journal*, 23(6.1), 96-114. [https://www.researchgate.net/publication/337908062\\_How\\_Thai\\_students\\_use\\_mobile\\_devices\\_when\\_learning\\_EFL\\_and\\_the\\_effect\\_of\\_urban\\_rural\\_school\\_location](https://www.researchgate.net/publication/337908062_How_Thai_students_use_mobile_devices_when_learning_EFL_and_the_effect_of_urban_rural_school_location)
- Indarti, D. (2021). Lexical richness of students' writings. *Wanastra: Jurnal Bahasa dan Sastra*, 13(1), 47-53. <https://doi.org/10.31294/w.v13i1.9683>
- Ivankova, N. V., & Creswell, J. W. (2009). Mixed methods. In J., Heighman, & R. A. Croker, (Eds.), *Qualitative Research in Applied Linguistics: A Practical Introduction* (pp. 135-161). England: Palgrave Macmillan. <https://doi.org/10.1057/9780230239517>
- Khushik, G. A., & Huhta, A. (2022). Syntactic complexity in Finnish-background EFL learners' writing at CEFR levels A1–B2. *European Journal of Applied Linguistics*, 10(1), 142-184. <https://doi.org/10.1515/eujal-2021-001>
- Kim, H. (2020). Nominal modifiers in argumentative essays as discriminators for writing course placement decisions. *English Teaching*, 75(3), 3-24. <https://doi.org/10.15858/engtea.75.3.202009.3>
- Kiselnikov, A., Vakhitova, D., & Kazymova, T. (2020). Coh-metrix readability formulas for an academic text analysis. *IOP Conference Series: Materials Science and Engineering*, 890(1), 1-7. doi:10.1088/1757-899X/890/1/012207
- Kremzer (2021). Academic reflective writing or anecdotal storytelling: a study on pre-service EFL teaching portfolios. *European Journal of Teacher Education*, 1-25. doi: 10.1080/02619768.2021.1961735
- Latifi, S., & Gierl, M. (2020). Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing. *Language Testing*, 38(1), 62-85. <https://doi.org/10.1177/0265532220929918>
- Leavy, P. (2017). *Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches*. New York, NY: The Guilford Press. <https://doi.org/10.1111/fcsr.12276>

- Leijten, M., Van Waes, L., Schrijver, I., Berolet, S., & Vangehuchten, L. (2019). Mapping Master's students' use of external sources in source-based writing in L1 and L2. *Studies in Second Language Acquisition*, 413, 555-582. <https://doi.org/10.1017/S0272263119000251>
- MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2018). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction?. *Reading and Writing: An Interdisciplinary Journal*, 3, 1553-1574. <https://doi.org/10.1007/s11145-018-9853-6>.
- Mahadini, M. K., Setyaningsih, E., & Sarosa, T. (2021). Using conventional rubric and Coh-Metrix to assess EFL students' essays. *International Journal of Language Education*, 54, 260-270. <https://doi.org/10.26858/ijole.v5i4.19105>
- McCarthy, P. M., Lewis, G. A., Dufty, D. F., McNamara, D. S. (2006). Analyzing writing styles with Coh-Metrix, at <https://www.aai.org/Papers/FLAIRS/2006/Flairs06-151.pdf>, accessed 20 March 2022.
- McCarthy, P. M., Kaddoura, N. W., Al-Harthy, A., Thomas, A. M., Duran, N. D., & Ahmed, K. (2022). Corpus analysis on students' counter and support arguments in argumentative writing. *Pegem Egitim ve Ogretim Dergisi*, 12(1), 256-271. doi: 10.47750/pegegog.12.01.27
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh, UK: Edinburgh University Press.
- McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45, 499-515. doi:10.3758/s13428-012-0258-1
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion'. *Discourse Processes*, 47(4), 292-330. <https://doi.org/10.1080/01638530902959943>
- Miura, K. (2020). Developing semantic-based DDL based on a comparative study of the verb use of British and Japanese students. *Learner Corpus Studies in Asia and the World*, 4, 41-66. doi: <https://ci.nii.ac.jp/lognavi?name=crossref&id=info:doi/10.24546/81011993>
- Moghadam, S. M., & Samar, R. G. (2020). Metaphor in second language academic writing. *Language Awareness*, 29(3-4), 255-271. <http://dx.doi.org/10.1080/09658416.2020.1786577>
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, 100-511. <https://doi.org/10.1016/j.asw.2020.100511>
- Nkhobo, T. I. (2022). Exploring instances of Deleuzian rhizomatic patterns in student writing and online interactions at an open distance eLearning institution in South Africa. Doctoral thesis (University of South Africa, Pretoria). <https://hdl.handle.net/10500/28746>
- Nkhobo, T., & Chaka, C. (2021). Exploring instances of Deleuzian rhizomatic patterns in students' writing and in online student interactions. *International Journal of Learning, Teaching and Educational Research*, 20(10), 1-22. <https://doi.org/10.26803/ijlter.20.10.1>
- Papangkorn, P., & Phoocharoensil, S. (2021). A comparative study of stance and engagement used by English and Thai speakers in English argumentative essays. *International Journal of Instruction*, 141, 867-888. <https://doi.org/10.29333/iji.2021.14152a>
- Perin, D., & Lauterbach, M. (2016). Assessing text-based writing of low-skilled college students. *International Journal of Artificial Intelligence in Education*, 28, 56-78. <https://doi.org/10.1007/s40593-016-0122-z>.
- Riazi, M. (2016). *The Routledge Encyclopedia of Research Methods in Applied Linguistics*. London: Routledge.
- Richards, K., Ross, S., & Seedhouse, P. (2012). *Research Methods for Applied Language Studies: An Advanced Resource Book for Students*. London: Routledge.
- Sinharay, S., Zhang, M., & Deane, P. (2019). Prediction of essay scores from writing process and product features using data mining methods. *Applied Measurement in Education*, 32(2), 116-137. <https://doi.org/10.1080/08957347.2019.1577245>
- Srisunakrua, T., & Chumworatayee, T. (2019). Readability of reading passages in English textbooks and the Thai National Education English Test: A comparative study. *Arab World English Journal*, 10(2), 257-269. <https://dx.doi.org/10.24093/awej/vol10no2.20>
- Ulfa, N., & Muthalib, K. A. (2020). Lexical bundles in students' essay writing' *English Education Journal*, 11(3), 367-379. <http://www.jurnal.unsyiah.ac.id/EEJ/article/view/15934>
- Van Waes, L., Leijten, M., Pauwaert, T., & Van Horenbeeck, E. (2019). A multilingual copy task: Measuring typing and motor skills in writing with inputlog. *Journal of Open Research Software*, 7(1), 1-8. <http://doi.org/10.5334/jors.234>
- Wang, J., Engelhard Jr, G., & Combs, T. (2021). Exploring difficult-to-score essays with a hyperbolic cosine accuracy model and Coh-Metrix indices. *The Journal of Experimental Education*, 1-20. <https://doi.org/10.1080/00220973.2021.1993774>

- Westerlund, M. (2019). Correlations between textual features and grades on the Swedish National Exam in English: A Coh-Metrix analysis, at <https://www.diva-portal.org/smash/record.jsf?dswid=1747&pid=diva2%3A1374786>, accessed 22 March 2022.
- Xu, C., & Xia, J. (2019). Scaffolding process knowledge in L2 writing development: Insights from computer keystroke log and process graph. *Computer Assisted Language Learning*, 34(4), 583-608. <https://doi.org/10.1080/09588221.2019.1632901>
- Zeng, S.C., & Shen, Y. (2016). An exploratory study on the easability of reading tests for English Majors Band 4 (TEM-4) and Band 8 (TEM-8). 2nd International Conference on Social Science and Higher Education, (pp. 148-151). <https://dx.doi.org/10.2991/icsshe-16.2016.38>
- Zhao, C. (2014). Lexical cohesion of Sino-British college students' EAP writing. *Theory and Practice in Language Studies*, 4(10), 2123-2128. doi: 10.4304/tpls.4.10.2123-2128

**Tlatso Nkhobo** ([nkhobti@unisa.ac.za](mailto:nkhobti@unisa.ac.za)) holds a Ph.D. in applied English language studies. He is a lecturer in the Department of English Studies, College of Human Sciences, at the University of South Africa (UNISA), Pretoria, in South Africa. His research interests include: academic writing; electronic learning; learning analytics; writing analytics; and rhizomatic learning.

**Chaka Chaka** ([chakachaka8@gmail.com](mailto:chakachaka8@gmail.com)) holds a Ph.D. in applied English language studies. He is a full professor in the Department of English Studies, College of Human Sciences, at the University of South Africa (UNISA), Pretoria, in South Africa. He previously lectured at four South African universities. His research interests include the following areas: language studies; computer-mediated communication (CMC); electronic learning (e-learning); computer assisted language learning (CALL); mobile learning (m-learning); mobile assisted language learning (MALL); learning and teaching through text and instant messaging (especially online social network messaging); and learning analytics.