ReCALL (2021), 33: 3, 214–228
doi:10.1017/S095834402100001X

RESEARCH ARTICLE

The impact of computer-aided concept mapping on EFL learners' lexical diversity: A process writing experiment

Mohammad Hassanzadeh
Vali-e-Asr University of Rafsanjan, Iran (mhassanzadeh@vru.ac.ir)

Elahe Saffari
Vali-e-Asr University of Rafsanjan, Iran (elahe.saffari89@gmail.com)

Saeed Rezaei
Sharif University of Technology, Iran (srezaei@sharif.edu)

Abstract

Nowadays, many second/foreign language (L2) academic writing instruction programs place a high premium on pre-writing strategies. The current study examined the effect of software-supported concept mapping on lexical diversity (LD) of English learners' argumentative essays within a process writing framework. Additionally, the relationship between the learners' LD and their overall writing quality was investigated. To this end, 53 university English as a foreign language (EFL) undergraduates were assigned to a computer-aided concept mapping (CACM) and a traditional outlining condition over a span of seven weeks. The CACM group was instructed through the graphic organizer software Inspiration®, whereas the comparison group underwent outlining instruction for planning their writing tasks. Measure of textual lexical diversity (MTLD) was used to assess the so-called D values of the assignments. The results revealed that the CACM group outperformed the outlining group in terms of LD scores. Also, no relationship was found between LD and overall quality of the essays. The findings provide L2 researchers and teachers with insights into understanding the use of CACM strategy in process writing. Moreover, exploiting MTLD afforded our experiment the opportunity to counteract potential pitfalls associated with text size. Further implications for the L2 teacher are also discussed.

Keywords: computer-aided concept mapping (CACM); Inspiration®; lexical diversity; MTLD; outlining; writing quality

1. Introduction

Concept mapping (CM) is a cognitive learning strategy and a visual road map that represents the pathways connecting the meanings of concepts (Novak & Gowin, 1984; Novak, Gowin & Johansen, 1983). Two notions form the theoretical framework of CM: "schema theory" (Rumelhart & Norman, 1978; Winn & Snyder, 1996) and Ausubel's (1968) "theory of human cognitive learning," often referred to in the literature as *meaningful learning theory*. According to the schema theory, when learners are confronted with new information, they react in one of three ways: accretion, tuning, or restructuring. In accretion, the learner assimilates new knowledge into an existing schema without changing the overall schema. In tuning, the learner realizes that the existing schema is inadequate for the new knowledge and modifies that.

Cite this article: Hassanzadeh, M., Saffari, E. & Rezaei, S. (2021). The impact of computer-aided concept mapping on EFL learners' lexical diversity: A process writing experiment. *ReCALL* 33(3): 214–228. https://doi.org/10.1017/S095834402100001X

© The Author(s), 2021. Published by Cambridge University Press on behalf of European Association for Computer Assisted Language Learning.

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

In restructuring, the learner creates a new schema that resolves inconsistencies between the new knowledge and the old schema.

Ausubel's (1968) theory of human cognitive learning focuses on the processes of using meaningful learning for acquiring knowledge. At the core of the theory is the distinction between rote and meaningful learning. Ausubel theorized that new meaning develops as a result of previously acquired concepts and propositions reorganizing and merging with new information. While the process of creating a concept map is in the works, a rearrangement of new concepts occurs as a result of obtaining new knowledge and a deeper understanding of the topic. As a continuation of Ausubel's legacy, Novak's (2002) "theory of meaningful learning" connects cognitive gains with emotional sensitivity. According to Novak, meaningful learning is dynamic and is modified over time as new information is acquired.

One means of developing students' academic competence and performance includes writing through meaningful practice. As an essential language skill, writing compels second/foreign language (L2) learners to think, concentrate, and organize their ideas. It also cultivates their analytical abilities and demands highly meaningful cognitive practice. Meanwhile, students learning to write in an L2 come across multiple challenges, namely deciding on the content knowledge related to a writing topic, making proper lexical and grammatical choices to form correct sentences, and organizing sentences into a paragraph or an entire essay. Hence, it is vital to identify efficient learning strategies and constructive planning tools, as composing a well-developed multiple-paragraph essay can be a daunting task for learners (Graham, Collins & Rigby-Wills, 2017). According to Novak (2002), learning by rote contributes very little to the learners' knowledge structure and hampers creative thinking and innovative problem-solving. As an attempt to counter this inefficiency, CM was put forth for augmenting conceptual understanding such that new concepts can be meaningfully assimilated into the learners' pre-existing knowledge base. Focusing on knowledge structure in such a way is assumed to promote meaningful learning.

Although modern practices are oriented towards pre-writing as a key component of the writing process, it is often not given substantive thought by instructors in many educational settings (Chang, Sung & Chen, 2001; Gardner, 2015). Teachers often express concern that enhancing learners' writing ability is easier said than done, which, ideally, requires ample time, energy, and expertise to fulfill (Zumbrunn & Krause, 2012). In this process, learners' lexical competence can be an asset to their written productions. An essential ingredient of lexical competence is lexical diversity (LD), which refers to the range or, in fact, breadth of vocabulary usage in a text or piece of writing (Malvern, Richards, Chipere & Durán, 2004). For a text to be highly lexically diverse, the writer has to produce a wide range of different words, with little repetition of the words already used (McCarthy & Jarvis, 2007). In short, the current study attempted to investigate the effectiveness of computer-aided concept mapping (CACM) as a pre-writing strategy versus the traditional outlining approach on English as a foreign language (EFL) learners' LD as a key component of linguistic competence in written productions. In addition, the study sought to explore the relationship between learners' LD and their writing quality. Hence, the following research questions were proposed:

1. Is CACM as a pre-writing strategy more effective than the traditional outlining approach in developing EFL learners' LD?
2. Is there a relationship between LD (represented by D values) and the writing quality scores of the two conditions (CACM vs. outlining)?

2. Literature review

To date, the effects of various pre-writing strategies (e.g. outlining, free writing, brainstorming, mind mapping) on various aspects of English as a second language (ESL) and EFL learners'

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

writing ability have been investigated (e.g. Ellis & Yuan, 2004; Lan, Sung, Cheng & Chang, 2015; Ojima, 2006). In principle, it is held that skills such as meaningful organization of the text, critical thinking, and idea generation are positively affected by the aforementioned pre-writing strategies. Furthermore, several studies (e.g. Blunt & Karpicke, 2014; Novak & Gowin, 1984; Redford, Thiede, Wiley & Griffin, 2012) have proposed that CM is an effective strategy that can improve both the quality and quantity of written compositions. Nevertheless, the majority of studies to this day have adopted a paper-and-pencil approach to CM. In retrospect, the traditional approach has several drawbacks (Chang *et al.*, 2001; Reader & Hammond, 1994; Schau, Mattern, Zeilik, Teague & Weber, 2001). On the one hand, the process of constructing and modifying concept maps produced in this way is often tedious and frustrating for students, and occasionally viewed by teachers as inconvenient for evaluation due to their often cluttered image. On the other hand, the development process of the concept maps cannot be recorded; hence, the quality of learning may only be judged on the basis of learners' final production.

2.1 CACM and aspects of writing

In light of the advances in computer-aided educational applications, a number of software packages have been developed to cope with the limitations of the paper-and-pencil approach. Mapping software programs enable learners to construct, revise, and record their individual maps and receive instant feedback from an instructor; meanwhile, the instructor is also able to trace and further examine the students' idea-generation trajectories. Thus, CACM is seen to have the potential to overcome most of the shortcomings of the paper-and-pencil approach. A plethora of studies (e.g. Gardner, 2015; Kwon & Cifuentes, 2009; Liu, 2011; Liu, Chen & Chang, 2010) corroborate the effectiveness of incorporating CACM on learners' writing performance in terms of various dimensions including accuracy, fluency, coherence, and syntactic complexity. Overall, the bulk of research in this area suggests that LD plays a crucial role in distinguishing between high and low writing proficiency levels (Crossley, Salsbury & McNamara, 2012; Gebril & Plakans, 2016; González, 2017; Olinghouse & Wilson, 2013). However, there are studies surrounding LD and writing proficiency levels that found little or no association between the two variables (e.g. Jarvis, 2002; Wang, 2014; Yu, 2010). In fact, evidence implies that moderator variables such as first language (L1) background and planning strategies may affect this relationship (Jarvis, 2002). Dujsik (2008), for example, examined the effects of computer-aided pre-writing strategy instruction on intermediate ESL students' strategy use, writing quantity, and writing quality. The results of the sequential mixed-methods study demonstrated the significant impact of training on the learners' strategy use, but failed to detect tangible effects on the quantity and quality of their writing.

A number of studies have examined the role of CACM as an instructional strategy. Sung (2008) compared the effects of computerized CM versus no-mapping on the writing performance of 125 ESL undergraduate students. The results indicated that participants of the CM group – using a piece of software known as Inspiration – outperformed the comparison group in terms of content and organization of their writing due to ample opportunity for devising their assignments through visualization. Elsewhere, Liu (2011) examined the effect of using different CM treatments (no-mapping, individual-mapping, and cooperative-mapping) during the pre-writing phase on ESL learners' writing performance across various proficiencies. The participants were 94 university freshmen divided into three groups of high-, middle-, and low-level proficiency on the basis of their baseline writing scores. Participants received all three rounds of treatment for accomplishing three writing assignments. The results indicated that the two CACM conditions were more productive than the no-mapping condition. Furthermore, the study revealed the sophistication of the learners' concept maps, which were positively correlated with their writing performance. In a multiple-baseline study, Evmenova *et al.* (2016) investigated the impact of a computer-based graphic organizer (CBGO) on the quantity and quality of persuasive essays.

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

The participants were 6th- and 7th-grade ESL students with high-incidence disabilities who were assigned to the three conditions of baseline (writing on the computer but not using the graphic organizer), CBGO use (writing on the computer with the graphic organizer), and maintenance (writing on the computer without the graphic organizer). Data were collected on the number of words, sentences, transition words, essay parts, and holistic writing quality score, revealing that the writing quality and quantity of the CBGO planning condition improved more significantly than the baseline and maintenance groups. The CBGO used for the second group allowed students to approach planning and writing either in a vertical (i.e. brainstorm all parts first and then write complete sentences) or horizontal manner (i.e. brainstorm one idea and turn it into a complete sentence before moving to the next one). Furthermore, Gardner (2015) attempted to explore the effects of CACM versus the so-called four-square graphic organizers on the quality of persuasive writing compositions administered through essays to ESL students. The four-square organizer is a type of pre-writing strategy in which a complete topic sentence is written in the center of a large square, which is divided into four smaller squares of equal size. Learners are then supposed to write three sentences that develop the thesis statement. The results indicated that CACM more efficiently affected the persuasive content and engagement with content.

2.2 Lexical diversity and writing

Having discussed the capability of CACM for the pre-writing stage, we now proceed to a discussion of the role of lexical competence in the actual writing process. Research suggests that lexical knowledge is a crucial factor in writing proficiency. Several studies (Crossley & McNamara, 2012; Crossley *et al*., 2012; Crossley, Salsbury, McNamara & Jarvis, 2011; Engber, 1995; Gebril & Plakans, 2016; González, 2017; Goodfellow, Lamy & Jones, 2002) have observed the central role of LD in academic writing proficiency. Gebril and Plakans (2016), for example, investigated issues related to the influence of textual borrowing on LD and the difference in LD across test scores on integrated tasks. To this end, 130 ESL students from a Middle Eastern university completed a reading-based integrated task that required students to read two passages on "global warming" and write an argumentative essay on the topic. The writing samples were analyzed via CLAN software as a measure of their LD. The scoring procedure was based on the Test of English as a Foreign Language (TOEFL) iBT's integrated writing task rubrics. The results demonstrated a positive correlation between LD and the integrated writing scores with a fairly large effect size.

In another study, Crossley *et al*. (2012) explored how L2 texts can be classified using computational indices that characterize lexical competence. In so doing, writing samples derived from 100 L2 learners at various proficiency levels were analyzed using lexical indices available in Coh-Metrix 3.0. The writing samples were categorized into beginning, intermediate, and advanced groupings based on the learners' scores on the Institutional TOEFL or the TOEFL iBT, and the ACT ESL Compass reading and grammar tests. A discriminant function analysis was used to predict the level categorization of the texts using lexical indices related to breadth of lexical knowledge (i.e. word frequency and LD), depth of lexical knowledge (hyponymy, polysemy, and word meaningfulness), and access to core lexical items (word concreteness and familiarity). The findings indicated that word frequency, LD, and word familiarity were the strongest predictors of the writing proficiency level. Further evidence came from González (2017), who investigated the extent to which lexical frequency and LD contribute to writing proficiency scores on monolingual English-speaking writers' and advanced multilingual writers' academic compositions. The data comprised 172 essays written by the multilingual and monolingual learners. The essays were evaluated by three independent raters based on the TOEFL iBT's independent writing task rubrics. Results from a binary logistic regression indicated that LD had a significantly greater contribution to the writing scores than lexical frequency, thereby situating LD as an essential component of academic writing proficiency.

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

2.3 Measure of textual lexical diversity and other measures of lexical diversity

The concept of LD has been of interest to linguists since the 1930s, leading to the introduction of several indices for its measurement. As a traditional measure of range of a writer's vocabulary, LD has often been calculated through type-token ratio (TTR) (Malvern & Richards, 2002), which represents the ratio between the number of different words (types) and the total number of words (tokens). TTR is claimed to be inherently flawed, as it varies as a function of sample length. To counter this problem, several researchers developed indices with diverse algebraic transformations of the TTR (e.g. Guiraud, 1954; Maas, 1972). Nevertheless, these TTR variations could not resolve the problem of text length influence. Later, Yule's K measure was tested on texts containing several thousands of tokens, and as a result, text length influence remained a significant factor in evaluating LD. Yule's K measure (Yule, 1944) is a probability model of the changes that take place in the lexical frequency spectrum of a text as the text becomes longer.

Computational linguistics has introduced a new generation of tools for measuring LD that produce length-invariant estimates without removing any data. One such tool is known as vocd-D, which was developed by Malvern and Richards (1997) and computationally implemented by McKee, Malvern and Richards (2000). This measure, derived from the vocd algorithm and the process utilized in its generation, counteracts the problem of sample size through random sampling and reference to ideal curves. The procedure is run three times to reach the optimum diversity (D) value. Vocd-D is also sensitive to short samples that contain less than 150 tokens (Koizumi, 2012; McCarthy & Jarvis, 2007; Owen & Leonard, 2002). Furthermore, McCarthy and Jarvis (2007, 2010) proposed the HD-D index. HD stands for hypergeometric distribution and HD-D is an alternative to the vocd-D index. The estimation of LD using the HD-D is based on probabilities of word occurrence in a language sample (McCarthy & Jarvis, 2010). HD-D is mathematically less demanding than vocd-D, although both indices are highly correlated and also sensitive to short samples of less than 150 tokens (deBoer, 2014).

The measure of textual lexical diversity (MTLD) is another contemporary measure developed by McCarthy (2005). MTLD measures TTR after every word of a sample until it reaches a value of 0.72. Two MTLD values are regularly calculated: one for forward processing and one for reverse processing. The average of the forward and backward MTLD scores produces the final D value through a sequential analysis of the sample (Fergadiotis, Wright & Green, 2015; Koizumi, 2012). According to several research studies, the MTLD measure presents several advantages: it is more robust with regard to text length variations than vocd-D or HD-D (e.g. Fergadiotis, Wright & West, 2013; Treffers-Daller, 2013), and it demonstrates no text length bias for text samples that encompass between 100 and 2,000 tokens (Crossley, Salsbury & McNamara, 2009; McCarthy, 2005). Further evidence was found from a study by González (2017), who observed that a reliable instrument was necessary for calculating the D values of the written essays that varied greatly in text length. Consequently, González used the MTLD index within the Coh-Metrix (Version 3.0), which produced reliable estimations of LD.

To validate the MTLD approach, McCarthy and Jarvis (2010) compared it against the performances of the basic (TTR and Yule's K) and sophisticated competing indices (vocd-D, HD-D, and Maas) in the field. Two corpora were used in the study: the MJ corpus and the M&C corpus, with each text originally composed of approximately 2,000 words. Following common practice in LD assessment (e.g. McCarthy & Jarvis, 2007; McKee *et al.*, 2000), each text of each register was divided into smaller sections (i.e. one section of 2,000, two sections of 1,000, etc.) in order to examine the sensitivity of the LD indices when texts of varying lengths were assessed. Thus, a total of 1,584 textual units were included. The comparisons involved assessments of convergent validity, divergent validity, internal validity, and incremental validity. The results suggested that MTLD correlates highly with all the established LD indices and satisfies convergent validity to at least the same degree as other sophisticated and established LD indices. In the case of divergent validity, the results indicated that MTLD, similar to Maas, K, vocd-D and HD-D, did not highly

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

correlate with TTR, which was considered as a flawed index. In sum, MTLD is one of the most versatile indices of measuring LD. As noted, it is less sensitive to text length variations than TTR and D as traditional algorithms (e.g. Fergadiotis *et al.*, 2013; Treffers-Daller, 2013) and even other sophisticated measures such as vocd-D and HD-D.

Although there has been a relatively large body of research into the effect of CACM on a number of variables (e.g. accuracy, text cohesion, vocabulary learning, writing motivation), one issue that has not received due attention by researchers is the extent to which CACM can affect LD in essay writing. Thus, in the first place, this study integrated the affordances of computer-assisted language learning into the exploration of LD. Second, even though the majority of studies testify to the role of LD in distinguishing between high and low levels of writing proficiency (e.g. Crossley *et al.*, 2012; Gebril & Plakans, 2016; González, 2017), further evidence from a process writing experiment may shed more light on their partly inconclusive association. In addition, as discussed earlier, the basic indices of LD (e.g. TTR, Guiraud, and Yule's K) have been inherently problematic to a greater or lesser extent. Even sophisticated measures of LD (e.g. vocd-D and HD-D) have been criticized on the grounds that they are sensitive to text length variations. Thus, MTLD was ultimately opted for this project.

3. Method

3.1 Participants

Initially, the sample included 55 mixed-gender sophomores, aged between 19 and 24 (M age $= 20.3$, $SD = 1.59$) from two intact classes at Vali-e-Asr University of Rafsanjan in Iran. They were native speakers of Farsi majoring in English language literature who had managed to gain entrance to the university on the basis of their ranking obtained in a competitive national examination. Out of the two groups, one was randomly chosen as the experimental ($N = 27$) and the other as the comparison group ($N = 26$). The groups were from two independent essay-writing classes, and the course was a core subject in their undergraduate program. However, a mock Oxford Placement Test (OPT) (Version 1.1, 2001) was administered to further establish whether they were at a homogeneous level of language proficiency. Based on a rating descriptor accompanying the OPT, 53 students who scored between 30 and 47 – the acceptable range for the intermediate level – were considered qualified to partake in this study.

3.2 Instrumentation

To achieve the objectives of the study, the following instruments were applied: (a) a writing task (as pre-test and post-test), (b) the Inspiration software (Version 9), (c) a writing analytic scale (Jacobs, Zinkgraf, Wormuth, Hartfiel & Hughery, 1981), and (d) MTLD.

3.2.1 Inspiration

Inspiration, a product of Inspiration Software, Inc., is a concept-mapping program designed for adolescent and adult students (Liu, 2011; Sung, 2008). This visual learning tool helps users to organize their ideas by creating hierarchical concept maps and to make immediate changes by adding concepts, creating links, labeling propositions, and moving items around. Inspiration also contains cross-curricular templates designed to help students and educators jump-start their writing process. A built-in help system provides support when users have inquiries about how to apply it. Moreover, this software consists of a presentation manager that can transform diagrams and concept maps into polished presentations that communicate ideas and notions.

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

3.2.2 MTLD

MTLD (McCarthy, 2005), a product of textinspector.com, was used in this study to calculate the learners' LD scores. In addition to not varying as a function of text length (McCarthy & Jarvis, 2010), MTLD allows for comparisons between text segments of considerably different lengths (at least 100 to 2,000 words) and produces reliable results while strongly correlating with other LD indices (McCarthy, 2005). The MTLD scores have an indefinite range; however, validation studies identified the typical score as ranging from 70 to 120, with higher scores indicating greater diversity (McCarthy & Jarvis, 2010). To enter the data for MTLD analysis, the spelling mistakes were manually corrected according to the guidelines provided by Text Inspector, as the presence of errors in texts can have a confounding effect on LD ratings. First, the written texts were checked for any spelling, and typographical errors and certain misspellings were corrected. These principles for correcting spelling mistakes are based on the premise that the focus of LD is to explore how diverse the range of words are. Therefore, to fulfill this primary focus, the words themselves must be in correct and recognizable forms. Second, contracted forms were separated (e.g. *I'm* was changed to *I am*, *don't* to *do not*, *I'd* to *I would*, etc.). Third, all digits and numbers were excluded by Text Inspector (see supplementary material for the Text Inspector tool and the LD profile score, respectively).

3.3 Procedure

First and foremost, the students' and university administrators' written consent to work the intervention into the program was sought in advance. Prior to the treatment, a five-paragraph argumentative essay of approximately 300 words was assigned within a 45-minute time limit as a pre-test to both conditions. Next, the experimental group underwent CACM treatment via Inspiration in a computer-equipped classroom, spanning seven sessions of their regular class time. During the first session, students were given instructions on how to sketch concept maps via Inspiration. The first segment comprised an introduction to CM in which the notion was delineated by the instructor[1] and some terminology associated with it, including concepts, linking, propositions, and cross-links, was presented. Moreover, the instructor provided the class with some samples of concept maps as well as a video clip for additional support. During the second part of the session, the learners created their own concept maps using Inspiration under the instructor's guidance. In each session, approximately 20–25 minutes were devoted to implementing the procedure. At the outset of each session, the participants received a writing prompt, and each individual was required to construct their hierarchical map with the main concepts and sub-concepts incorporated and labeled with propositions within the diagram view of the software. Then the instructor examined the concept maps and had students go through the drafting process via Microsoft Word on their computers. The instructor also provided prompt-like advice with regard to the lexical aspects of the learners' drafts (e.g. elaboration, LD, relevance to the topic, etc.). This procedure was repeated over the subsequent sessions until the treatment was completed (see supplementary material for a hierarchical concept map constructed by one student).

Alongside the experimental group, students in the comparison (outlining) group underwent the same task of developing a five-paragraph essay. Similarly, the pre-writing phase as an essential component of process writing was followed discreetly by the instructor in the comparison group as well. The students were given similar topics for their writing assignments but used an alternative outlining strategy over a seven-session period that took place during seven weeks. They received a topic each session and were instructed to generate an outline before composing a draft. To create a well-developed outline, the students were taught to categorize the main points and ideas relating to the introduction, body, and conclusion paragraphs. Then the outline structures (i.e. clustering and ordering ideas) were checked by the instructor. The students were told to

[1]One of the authors of this manuscript.

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

follow the outline content and transform their ideas into paragraphs by providing relevant details. The instructor also gave the students the necessary feedback to revise their written drafts during each stage of the writing process before completing their texts (final products). It should be noted that the outlining group's writing prompts, essay length, and administration timing were identical to those of the experimental group. In fact, they were parallel classes that were held once a week.

3.4 Scoring

Immediately after the seven-week span of the treatments, a post-test identical in topic and genre to the pre-test was assigned to both groups. Two independent, trained raters[2] anonymously rated all 106 essays in the original format (i.e. handwritten) using Jacobs *et al.*'s (1981) analytic scoring scale. This multi-trait scoring rubric has been extensively used in L2 writing studies to measure text quality (Polio & Friedman, 2017). In spite of being labor intensive, we used this scale in particular to evaluate the writing quality of the essays due to the added value it provides beyond the total score. In this scale, scripts are rated on five aspects of writing and are weighed differentially on the basis of content (30 points), language use (25 points), organization (20 points), vocabulary (20 points), and mechanics (5 points). The inter-rater reliability was also found to be at an acceptable range ($r = .72$). In the end, the compositions were entered into Text Inspector for calculating the D value. It is worth mentioning that for evaluating LD, only the lexical aspect of the essays is taken into account. The respective score typically ranged from 70 to 120.

4. Results

To address the first research question, the data were submitted to a one-way analysis of covariance (ANCOVA). The aim was to explore the impact of CACM on LD of the written essays. The independent variable was the type of intervention (CACM vs. outlining); the dependent variable was the LD scores on the post-test, and the covariate was the writing pre-test. Preliminary checks were made to ensure there was no violation of the assumptions of normality, linearity, homogeneity of variances, and homogeneity of regression slopes. Table 1 presents the descriptive data output for LD scores provided by Text Inspector.

The ANCOVA test (Table 2) revealed that after adjusting for the pre-test scores, there was a statistically significant difference between the LD indices of the two conditions, $F(1, 50) = 5.08$, $p = .029$, $\eta_p^2 = .09$. The magnitude of difference in the mean scores can be inferred from the effect size index, which suggests a medium impact. Overall, the results imply that CACM has affected LD scores more strongly than the outlining treatment. Figure 1 illustrates the performance of both groups across the tests.

With regard to the second research question, Pearson product-moment correlation was used to explore the association between learners' LD and writing quality scores (see Table 3 for the descriptive writing quality scores before and after the treatments). The results, based on Jacobs *et al.*'s (1981) rubric (with a maximum score of 100), indicated no significant relationship between the post-test scores of the two variables in either of the conditions: CACM, $r = .06$, $p = .77$; outlining, $r = -.01$, $p = .94$. Figure 2 illustrates the status of the CACM group. As can be seen, the scores are spread out across the diagram, indicating no meaningful orientation. Similarly, no association was found between learners' LD and their writing quality scores on the pre-test: CACM, $r = -.16$, $p = .42$. A similar lack of association was found for the outlining condition, $r = .26$, $p = .19$.

[2]Both raters belonged to the academic community; one was a North American native speaker of English and the other an Iranian PhD holder in applied linguistics with adequate experience in assessing writing.

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

Table 1. Descriptive statistics for lexical diversity scores

		Pre-test		Post-test	
Group	*N*	*M*	*SD*	*M*	*SD*
CACM	27	71.77	16.89	83.61	21.47
Outlining	26	68.00	15.15	70.45	17.35

Table 2. ANCOVA output for lexical diversity (LD) scores

Source	*df*	*F*	Sig.	Partial eta squared
LD pre-test	1	7.44	.009	.129
Groups	1	5.08	.029	.092

Table 3. Descriptive statistics for writing quality

		Pre-test		Post-test	
Group	*N*	*M*	*SD*	*M*	*SD*
CACM	27	59.33	4.13	73.81	6.45
Outlining	26	57.77	6.51	60.27	5.61



Figure 1. Lexical diversity scores from pre-test to post-test

5. Discussion

The first research question inquired whether applying Inspiration as a graphic organizer would affect the D values in the learners' written essays. The results revealed that learners in the CACM group scored significantly higher than those in the outlining group, implying that using Inspiration for the pre-writing strategy enabled learners to build more complex cognitive structures to organize their conceptual system. As a matter of fact, learners' ability to give variety to their lexical choices and produce linguistically diverse texts received a boost. The second research question sought to explore the relationship between LD and overall writing quality. The findings

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

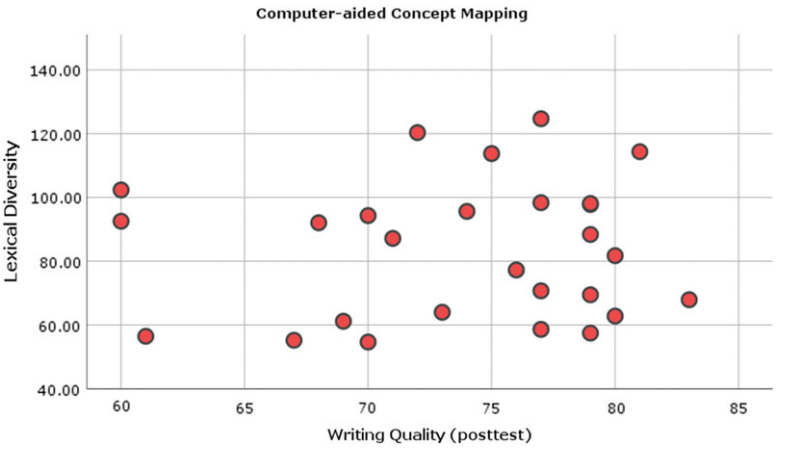


Figure 2. Association between lexical diversity and writing quality in the CACM group

revealed that there was no correlation between the two variables, neither on the post-test nor the pre-test.

To create concept maps, the individual learner needs to take charge of retrieving and restructuring the relevant concepts. The process helps them to envision the key themes and constituents of the text and their relationships. In this way, the participants of the current study were able to assimilate the new concepts into their pre-existing knowledge base. Therefore, the findings corroborate Ausubel's (1968) theory of human cognitive learning, which underscored the processes of using meaningful learning for acquiring knowledge. Although not many studies to date have specifically focused on the impact of CACM on LD in writing, the findings of the present study are, in one way or another, consistent with a number of previous research studies (e.g. Gardner, 2015; Kwon & Cifuentes, 2009; Liu, 2011; Liu *et al*., 2010) whose findings suggested that CACM can serve as a cognitive tool to enhance learners' thinking, problem-solving, and reasoning skills. With regard to the substantial impact of graphic organizers at the pre-writing stage, Gardner's (2015) findings are comparable to those of the current study. Although LD was not evaluated as a dependent variable in Gardner's study, the greater contribution of CACM, compared to the four-square or the no-organizer conditions, became evident. A possible reason for Gardner's convergent findings to those of ours might be attributed to CACM's propositional structure, which helps learners to organize their ideas and make connections between their argument and the content they are using to support their argument. One thing that can be deduced from our findings is that CACM instruction through process writing possibly made learners more attentive to drafting and revision stages of the writing process, thereby leading to improvement in their compositions. Furthermore, Liu's (2011) findings revealed that the application of different computerized concept-mapping treatments (individual-mapping and cooperative-mapping) as a pre-writing strategy had distinctly more impact on the writing performance of the learners than did the no-mapping treatment.

In retrospect, it has been argued that integrating computers in writing instruction may not always yield a favorable outcome due to various reasons such as classroom time constraints and dearth of the teacher's expertise. As mentioned earlier, Dujsik (2008) had found no difference between his experimental and control groups in terms of writing quality and quantity. In fact, there is a plethora of factors that can result in such varying outcomes in different studies. For instance, one possible cause for the conflicting findings with those of the current study might come from Dujsik's shorter intervention span. It therefore could be the case that the length of treatment, among other elements, can be a crucial factor.

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

Concerning the second research question, it was seen that EFL students' ability to produce diverse vocabularies may not always correlate with their mastery over composing texts. This finding fails to accord with previous conclusions that suggested that LD is positively associated with writing proficiency (Crossley & McNamara, 2012; Crossley *et al.*, 2012; Gebril & Plakans, 2016; González, 2017; Olinghouse & Wilson, 2013). Regarding the relationship between linguistic complexity features and L2 writing proficiency, the results here were not aligned with those of Crossley and McNamara's (2012), who observed that LD and lexical frequency can predict differences between low- and high-scoring learners. González (2017) also came up with dissimilar results to those of the current study. He used MTLD to investigate the extent of the contribution that LD and lexical frequency impart to English writing quality in multilingual and monolingual English academic compositions. The study concluded that LD had a greater impact on writing scores than lexical frequency. Moreover, our findings are in line with those who found that LD was not a defining characteristic of text quality (e.g. Jarvis, 2002; Wang, 2014; Yu, 2010). Wang (2014), for example, found a lack of association between LD and EFL writing proficiency. The study was carried out through an analysis of email texts written by 45 Chinese high school students. The TTR and vocd-D indices were used therein to assess LD. His findings revealed that the LD scores of email texts graded at higher proficiency levels were not significantly different from those of lower proficiency levels. Addressing this dissociation, one can argue that learners who produced a more diverse range of vocabulary could have committed more usage errors, thus impinging on their overall writing score. Yet, considering the dissimilar genres and LD measures across the two studies, our claim remains merely a conjecture and conclusive evidence should be garnered to verify that by future researchers. In a similar vein, Yu (2010) investigated the relationship between LD and quality of writing tasks, where 200 compositions on five different topics by English learners from different L1 backgrounds were analyzed in the form of written discourse using the vocd-D. The results revealed that for samples from the two largest L1 groups (i.e. Filipino and Chinese), LD was not a significant predictor of the quality of learners' written compositions. A possible explanation for this discrepancy between that study and ours could be that in Yu's experiment, the subjects from several language backgrounds were tested. On balance, our study provided further empirical evidence that the overall quality of a written essay may not always correlate with the use of a wider range of words or decrease in lexical monotony.

6. Conclusion

Theoretically, our findings demonstrate that applying the CACM strategy ahead of the drafting stage helped to facilitate and provoke effective thinking. This strategy affected the learners' diction through developing the generation and organization of ideas. In other words, CACM can possibly provide more chances for writers to view their own writing through visualization. This possibility can provide a justification for the improvements in the D values of student compositions. The present study was a pioneering attempt in the Iranian context to determine whether the retrieval of a wider range of lexical items by EFL learners could be realized through employing the Inspiration software. As can be perceived from the preceding studies, concept maps portray relationships in a logical and hierarchical fashion and meaningfully connect branches of content ideas. We, therefore, found that the application of CACM as a pre-writing strategy can help promote learners' LD in writing. Correspondingly, L2 writing instructors and practitioners can work graphic organizer training into their lesson plans in order to encourage their students to rehearse before homing in on the main draft. Researchers should take stock of the significant role of CACM in boosting linguistic complexity in general and LD in particular across different genres of writing in their future endeavors. Evaluating LD through a quantitative measure, such as MTLD, may provide teachers with a more accurate picture of their students' lexical development

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

and offer a respectable benchmark for distinguishing writing proficiency levels (also see Vögelin, Jansen, Keller, Machts & Möller, 2019). We would also wish to emphasize that accentuating CACM should take on a more substantive role in academic writing instructional courses. The finding that this pre-writing strategy can affect LD more than a traditional technique such as outlining holds worthwhile implications for methods of teaching and assessing EFL compositions. Teachers could set up varied practice activities on how to diversify learner diction during essay drafting and revision processes and how to identify alternatives for repetitive lexical items. We recommend that more attempts be made to introduce apps such as Inspiration to the L2 classroom. As noted, one merit of the study resides in the utilization of MTLD, which enabled us to counteract potential pitfalls that are associated with other indices, as it is maximally independent of text size. MTLD can also be used as a diagnostic tool to target students who are struggling with their unsatisfactory lexical knowledge.

The study also shed light on the extent of association between overall writing quality and LD. It is to be acknowledged that because the main thrust of the current study was to operationalize CACM as well as LD measures and software, we did not further scrutinize the analytic writing subscores. The main purpose of the second research question was to compare LD and writing quality as an overall construct. Hence, we sufficed to report a representative mean score. Nevertheless, the analytic rubric provides an overall score as well as the subscores on the learners' detailed performance. Analytic scales such as that of Jacobs *et al.*'s (1981), therefore, can be useful for formative assessment purposes where the provision of diagnostic information on the learners' writing performance is taken into account. We would also like to acknowledge that the respective data are available for sharing and further research, and we recommend that researchers analyze writing quality in terms of its subscale dimensions in connection with the other variables of their study. Thus, researchers of future studies may prefer to focus on a collection of textual features, including content, organization, accuracy, error production, style, and other contextual factors, and examine how such variables are affected by CACM or other computer-assisted pre-writing strategies.

6.1 Limitations

Furthermore, as LD mainly pertains to the breadth of lexical knowledge, it can only account for the range of words used during the writing process. It therefore cannot judge the quality of the lexicon such as how appropriately the words have been used. It is suggested that future studies examine depth of knowledge over and above the size of vocabulary. Schmitt's (2014) synthesis of studies involving vocabulary depth is also a thought-provoking resource for avid researchers and practitioners alike. Although the present study strived to open new directions for further investigation, it is recommended that researchers carry out a more comprehensive evaluation of the effect of various software applications on other dimensions of semantic complexity, namely lexical accuracy, lexical frequency, lexical sophistication, lexical density, and lexical repetition (see, e.g., Ballance, 2021). In light of the fact that pre-writing strategies instruction is an under-researched area, studies in the future can delve into different aspects of this enterprise. Care must be taken to consider the spillover effect of learner variables such as motivation, learning styles, and language proficiency. Hence, aptitude-treatment interaction[3] research can be illuminating in this regard. We also suggest that future researchers draw on other computerized learning tools such as those involving idea generation/organization software, which were not tapped in the present study. One last point of note is that our study was primarily designed for an experimental

[3]Aptitude-treatment interaction (ATI) refers to "studies that explore how learners' individual differences (e.g., aptitude, cognitive creativity, motivation, learning styles, strategies, working memory) are related to the effectiveness of varied kinds of instruction and pedagogical decisions. ATI studies empirically investigate how second language instruction can be optimized to fit the individual needs of a given learner" (Mackey, 2017: 550).

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

set-up, and the sample size was by no means adequately large to allow correlational extrapolation (to the population). As a result, the findings pertaining to the second research question should be treated with more caution and deemed perhaps as the equivalent of an action research project.

Supplementary material. To view supplementary material referred to in this article, please visit https://doi.org/10.1017/S095834402100001X

Ethical statement. The research involved no conflict of interest. All participants consented to contribute to the experiments and were personally informed about the outcome of their writing tasks. All individuals' performance data remained anonymous throughout the project.

References

Ausubel, D. P. (1968) *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.

Ballance, O. J. (2021) Narrow reading, vocabulary load and collocations in context: Exploring lexical repetition in concordances from a pedagogical perspective. *ReCALL*, 33(1): 4–17. https://doi.org/10.1017/S0958344020000117

Blunt, J. R. & Karpicke, J. D. (2014) Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, 106(3): 849–858. https://doi.org/10.1037/a0035934

Chang, K. E., Sung, Y. T. & Chen, S. F. (2001) Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning*, 17(1): 21–33. https://doi.org/10.1111/j.1365-2729.2001.00156.x

Crossley, S. A. & McNamara, D. S. (2012) Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2): 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Crossley, S., Salsbury, T. & McNamara, D. (2009) Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, 59(2): 307–334. https://doi.org/10.1111/j.1467-9922.2009.00508.x

Crossley, S. A., Salsbury, T. & McNamara, D. S. (2012) Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2): 243–263. https://doi.org/10.1177/0265532211419331

Crossley, S. A., Salsbury, T., McNamara, D. S. & Jarvis, S. (2011) Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4): 561–580. https://doi.org/10.1177/0265532210378031

deBoer, F. (2014) Evaluating the comparability of two measures of lexical diversity. *System*, 47: 139–145. https://doi.org/10.1016/j.system.2014.10.008

Dujsik, D. (2008) *The effects of pre-writing strategy training guided by computer-based procedural facilitation on ESL students' strategy use, writing quantity, and writing quality*. University of South Florida, unpublished PhD.

Ellis, R. & Yuan, F. (2004) The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26(1): 59–84. https://doi.org/10.1017/S0272263104026130

Engber, C. A. (1995) The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2): 139–155. https://doi.org/10.1016/1060-3743(95)90004-7

Evmenova, A. S., Regan, K., Boykin, A., Good, K., Hughes, M., MacVittie, N., Sacco, D., Ahn, S. Y. & Chirinos, D. (2016) Emphasizing planning for essay writing with a computer-based graphic organizer. *Exceptional Children*, 82(2): 170–191. https://doi.org/10.1177/0014402915591697

Fergadiotis, G., Wright, H. H. & Green, S. B. (2015) Psychometric evaluation of lexical diversity indices: Assessing length effects. *Journal of Speech, Language, and Hearing Research*, 58(3): 840–852. https://doi.org/10.1044/2015_JSLHR-L-14-0280

Fergadiotis, G., Wright, H. H. & West, T. M. (2013) Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech-Language Pathology*, 22(2): S397–S408. https://doi.org/10.1044/1058-0360(2013/12-0083)

Gardner, M. E. (2015) *Concept mapping: Effects on content knowledge and engagement with content in elementary students' persuasive writing*. University of Kentucky, doctoral dissertation.

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

Graham, S., Collins, A. A. & Rigby-Wills, H. (2017) Writing characteristics of students with learning disabilities and typically achieving peers: A meta-analysis. *Exceptional Children*, 83(2): 199–218. https://doi.org/10.1177/0014402916664070

Guiraud, P. (1954) *Les caractères statistiques du vocabulaire: essai de mèthodologie*. Paris: Presses Universitaires de France.

Jacobs, H. L., Zinkgraf, S., Wormuth, D., Hartfiel, V. & Hughery, J. (1981) *Testing ESL composition: A practical approach*. Rowley: Newbury House.

Jarvis, S. (2002) Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1): 57–84. https://doi.org/10.1191/0265532202lt220oa

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

Koizumi, R. (2012) Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction*, 1(1): 60–69. https://doi.org/10.7820/vli.v01.1.koizumi

Kwon, S. Y. & Cifuentes, L. (2009) The comparative effect of individually-constructed vs. collaboratively-constructed computer-based concept maps. *Computers & Education*, 52(2): 365–375. https://doi.org/10.1016/j.compedu.2008.09.012

Lan, Y.-J., Sung, Y.-T., Cheng, C.-C. & Chang, K.-E. (2015) Computer-supported cooperative prewriting for enhancing young EFL learners' writing performance. *Language Learning & Technology*, 19(2): 134–155. https://doi.org/10125/44421

Liu, P.-L. (2011) A study on the use of computerized concept mapping to assist ESL learners' writing. *Computers & Education*, 57(4): 2548–2558. https://doi.org/10.1016/j.compedu.2011.03.015

Liu, P.-L., Chen, C.-J. & Chang, Y.-J. (2010) Effects of a computer-assisted concept mapping learning strategy on EFL college students' English reading comprehension. *Computers & Education*, 54(2): 436–445. https://doi.org/10.1016/j.compedu.2009.08.027

Maas, H.-D. (1972) Über den Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik*, 2(8): 73–79.

Mackey, A. J. (2017) Classroom-based research. In Loewen, S. & Sato, M. (eds.), *The Routledge handbook of instructed second language acquisition*. New York: Routledge, 541–561. https://doi.org/10.4324/9781315676968-30

Malvern, D. D. & Richards, B. J. (1997) A new measure of lexical diversity. In Ryan, A. & Wray, A. (eds.), *Evolving models of language: Papers from the annual Meeting of the British Association for Applied Linguistics held at the University of Wales, Swansea, September 1996*. Clevedon: British Association for Applied Linguistics in association with Multilingual Matters, 58–71.

Malvern, D. & Richards, B. (2002) Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1): 85–104. https://doi.org/10.1191/0265532202lt221oa

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

McCarthy, P. M. (2005) *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)*. University of Memphis, unpublished PhD. Retrieved from '¼https://umdrive.memphis.edu/pmmccrth/public/Phil's%20papers.htm?uniq¼xq6bv

McCarthy, P. M. & Jarvis, S. (2007) *vocd*: A theoretical and empirical evaluation. *Language Testing*, 24(4): 459–488. https://doi.org/10.1177/0265532207080767

McCarthy, P. M. & Jarvis, S. (2010) MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2): 381–392. https://doi.org/10.3758/BRM.42.2.381

McKee, G., Malvern, D. & Richards, B. (2000) Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15(3): 323–338. https://doi.org/10.1093/llc/15.3.323

Novak, J. D. (2002) Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, 86(4): 548–571. https://doi.org/10.1002/sce.10032

Novak, J. D. & Gowin, D. B. (1984) *Learning how to learn*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139173469

Novak, J. D., Gowin, D. B. & Johansen, G. T. (1983) The use of concept mapping and knowledge vee mapping with junior high school science students. *Science Education*, 67(5): 625–645. https://doi.org/10.1002/sce.3730670511

Ojima, M. (2006) Concept mapping as pre-task planning: A case study of three Japanese ESL writers. *System*, 34(4): 566–585. https://doi.org/10.1016/j.system.2006.08.003

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

Owen, A. J. & Leonard, L. B. (2002) Lexical diversity in the spontaneous speech of children with specific language impairment: Application of D. *Journal of Speech, Language, and Hearing Research*, 45(5): 927–937. https://doi.org/10.1044/1092-4388(2002/075)

Polio, C. & Friedman, D. A. (2017) *Understanding, evaluating, and conducting second language writing research*. Abingdon: Routledge. https://doi.org/10.4324/9781315747293

Reader, W. & Hammond, N. (1994) Computer-based tools to support learning from hypertext: Concept mapping tools and beyond. *Computers & Education*, 22(1–2): 99–106. https://doi.org/10.1016/0360-1315(94)90078-7

Redford, J. S., Thiede, K. W., Wiley, J. & Griffin, T. D. (2012) Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, 22(4): 262–270. https://doi.org/10.1016/j.learninstruc.2011.10.007

Rumelhart, D. E. & Norman, D. A. (1978) Accretion, tuning, and restructuring: Three modes of learning. In Cotton, J. W. & Klatzky, R. (eds.). *Semantic factors in cognition*. Hillsdale: Lawrence Erlbaum Associates, 37–60.

Schau, C., Mattern, N., Zeilik, M., Teague, K. W. & Weber, R. J. (2001) Select-and-fill-in concept map scores as a measure of students' connected understanding of science. *Educational and Psychological Measurement*, 61(1): 136–158. https://doi.org/10.1177/00131640121971112

Schmitt, N. (2014) Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4): 913–951. https://doi.org/10.1111/lang.12077

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press

Sung, Y. L. (2008) *A case study on the use of inspiration to assisted college EFL students' writing*. Fu-Jen University, unpublished master's thesis.

Treffers-Daller, J. (2013) Measuring lexical diversity among L2 learners of French: An exploration of the validity of D, MTLD and HD-D as measures of language ability. In Jarvis, S. & Daller, M. (eds.), *Vocabulary knowledge: Human ratings and automated measures*. Amsterdam: Benjamins, 79–104. https://doi.org/10.1075/sibil.47.05ch3

Vögelin, C., Jansen, T., Keller, S. D., Machts, N. & Möller, J. (2019) The influence of lexical features on teacher judgements of ESL argumentative essays. *Assessing Writing*, 39: 50–63. https://doi.org/10.1016/j.asw.2018.12.003

Wang, X. (2014) The relationship between lexical diversity and EFL writing proficiency. *University of Sydney Papers in TESOL*, 9: 65–88. http://faculty.edfac.usyd.edu.au/projects/usp_in_tesol/pdf/volume09/Article03.pdf

Winn, W. & Snyder, D. (1996) Cognitive perspectives in psychology. In Jonassen, D. H. (ed.), *Handbook of research for educational communications and technology: A project of the Association for Educational Communications and Technology*. New York: Macmillan, 112–142.

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

Yule, G. U. (1944) *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press.

Zumbrunn, S. & Krause, K. (2012) Conversations with leaders: Principles of effective writing instruction. *The Reading Teacher*, 65(5): 346–353. https://doi.org/10.1002/TRTR.01053

About the authors

Mohammad Hassanzadeh is an assistant professor of applied linguistics with the Department of English Language and Literature at Vali-e-Asr University of Rafsanjan, Iran. He is currently serving as a guest professor at Sharif University of Technology. His research interests involve instructed second language acquisition with a focus on technology-enhanced pedagogy.

Elahe Saffari is an MA graduate in applied linguistics from Vali-e-Asr University of Rafsanjan, Iran. Having taught English across several language institutes and schools in the city of Kerman, she has developed a vested interest in L2 vocabulary and writing instruction research.

Saeed Rezaei is currently an associate professor of applied linguistics at Sharif University of Technology. His main areas of interest are in social issues in TESOL and linguistics. His most recent publications have appeared in *TESOL Quarterly*, *Teaching and Teacher Education*, *New Writing*, *Multilingua*, *Open Learning*, *Asian Englishes*, and *Higher Education*.

Author ORCiD. Mohammad Hassanzadeh, https://orcid.org/0000-0003-1510-1149
Author ORCiD. Elahe Saffari, https://orcid.org/0000-0001-8601-1437
Author ORCiD. Saeed Rezaei, https://orcid.org/0000-0003-0296-0382

https://doi.org/10.1017/S095834402100001X Published online by Cambridge University Press