



## OPEN ACCESS

## EDITED BY

Aysel Saricaoglu,  
Social Sciences University of Ankara, Turkey

## REVIEWED BY

Dogan Yuksel,  
Kocaeli University, Turkey  
Erdem Akbaş,  
Erciyes University, Turkey

## \*CORRESPONDENCE

Zhi Li  
✉ z.li@usask.ca

## SPECIALTY SECTION

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

RECEIVED 02 September 2022

ACCEPTED 10 February 2023

PUBLISHED 02 March 2023

## CITATION

Li Z, Makarova V and Wang Z (2023) Developing literature review writing and citation practices through an online writing tutorial series: Corpus-based evidence. *Front. Commun.* 8:1035394. doi: 10.3389/fcomm.2023.1035394

## COPYRIGHT

© 2023 Li, Makarova and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Developing literature review writing and citation practices through an online writing tutorial series: Corpus-based evidence

Zhi Li<sup>1\*</sup>, Veronika Makarova<sup>1</sup> and Zhengxiang Wang<sup>2</sup>

<sup>1</sup>Department of Linguistics, University of Saskatchewan, Saskatoon, SK, Canada, <sup>2</sup>Department of Linguistics, Stony Brook University, Stony Brook, NY, United States

Writing a literature review (LR) in English can be a daunting task for non-native English-speaking graduate students due to the complexities of this academic genre. To help graduate students raise genre awareness and develop LR writing skills, a five-unit online tutorial series was designed and implemented at a large university in Canada. The tutorial focuses on the following features of the LR genre: logical structure, academic vocabulary, syntax, as well as citation practices. Each tutorial unit includes an interactive e-book with explanations, examples, quizzes, and an individual or collaborative LR writing assignment. Twenty-nine non-native English-speaking graduate students from various institutions participated in the tutorials and completed five writing tasks. This study reports on their developmental trajectories in writing performance in terms of cohesion, lexical features, syntactic features, and citation practices as shown in three individual writing tasks. Corpus-based analyses indicate that noticeable, often non-linear, changes are observed in several features (e.g., use of connectives, range and frequency of academic vocabulary) across the participants' writing samples. Meanwhile, citation analysis shows a steady increase in the use of integral citations in the participants' writing samples, as measured with occurrence by the number of sentences, along with a more diverse use of reporting verbs and hedges in their final writing samples. Pedagogical implications are discussed.

## KEYWORDS

literature review writing, ESL graduate students, online tutorials, corpus-based analysis, linguistic complexity, citation practices, reporting verbs

## 1. Introduction

Well-developed academic writing skills in English are an indispensable prerequisite for academic success of graduate students since these skills are commonly expected in thesis writing and academic publication. Even after passing the entry English language requirements, many non-native English-speaking international students still experience multiple difficulties with academic writing (Cheng and Fox, 2008; Zhang, 2011). Most studies of writing address either the contexts of standard English proficiency tests, such as TOEFL (Riazi, 2016), or undergraduate academic writing courses (e.g., Dafouz, 2020). By contrast, very little is known about the specifics of writing skills development by graduate students (e.g., Shi and Dong, 2015). The available studies are limited in their scope, disciplines, number of participants, ethnic backgrounds (Cheng, 2007), or methods involving opinions rather than actual writing features (Zhang, 2011). To the best of our knowledge, there are no studies that consider the potential of extracurricular tutorials to impact the writing proficiency of graduate students.

Our focus on the genre of “literature review” is motivated by its high frequency among the types of graduate writing assignments across multiple disciplines in Canadian graduate programs (Shi and Dong, 2015). “Literature review” is defined as a survey of “the relevant literature to discuss the state of knowledge or identify gaps in research” (Shi and Dong, 2015: 131). Literature review has been one of the challenges faced by graduate students, native and non-native-English speaking ones alike (Chen et al., 2016; Badenhorst, 2018). Graduate-level international students often come to Canadian universities unaware of the specific requirements of this genre and experience multiple difficulties with it, such as avoiding plagiarism, building coherent narratives, citation formats, the use of vocabulary, and grammar forms, which impede their academic progress and cause other negative consequences (Shi and Dong, 2015). While various efforts have been made in English for academic purposes (EAP) programs to help graduate students develop academic writing skills and/or genre awareness (e.g., Storch and Tapper, 2009; Li et al., 2020; Crosthwaite et al., 2021), programs or courses dedicated to literature review writing are still sparse and under-investigated in research.

From the theoretical perspective, this study is motivated by exploring the specific linguistic structure and textual parameters associated with successful academic writing in graduate context. In general, the major domains of writing success (or failure) have been reported to include lexis (including academic vocabulary), syntax (at phrasal and sentence levels), cohesion, and citation/referencing conventions (Biber et al., 2011; Mansourizadeh and Ahmad, 2011; Mazgutova and Kormos, 2015). On the practical side, our project is aimed at identifying a cost-effective way to develop graduate students’ writing skills with minimal pressure on the faculty and university resources. In Canada, not all universities can yet provide adequate support for the academic writing needs of international students at the graduate level (Okuda and Anderson, 2018). As a potential solution, we developed an extracurricular writing tutorial series with free access to an online learning platform. In this paper, we are exploring whether the developed tutorial delivers practical outcomes in terms of improving academic writing proficiency in the genre of Literature review.

Our study, therefore, answers a strong demand for additional means of assisting graduate students with their academic writing skills development in the context of Literature review genre. We consider the feasibility and potential benefits of online academic writing tutorials for graduate students’ writing performance in terms of linguistic development and citation practices. The materials of the study come from writing samples collected from our online writing tutorial series on literature review writing.

## 2. Literature review

### 2.1. Learners’ development of academic writing

While we acknowledge the role of “history, ideology, and socio-cultural structures” in writing skills development (Zhang, 2011, p. 41), in this particular paper, we treat writing from an “autonomous” perspective of the academic literacies model

which sees writing as a development of specific skills and genre repertoires (Lea and Street, 1998). More specifically, we adopt the combination of “the linguistic and genre” approach focusing mostly on linguistic structures and writing within a specific genre (Xu, 2019). For the genre of literature review, it is not surprising to see studies applying genre analysis to graduate students’ writing samples (Flowerdew and Forest, 2011) or focusing on writer stances (Shahsavari and Kourepaz, 2020). However, specific linguistic features in this genre, especially by graduate students, have not been extensively researched. When it comes to gauging learners’ development of academic writing in general, linguistic measures are often operationalized at various levels such as lexical and syntactic complexities (Bulté and Housen, 2014). At the lexical level, measures of complexity and sophistication have been employed in several empirical studies. For example, in Mazgutova and Kormos (2015) pretest-posttest study, changes in five lexical measures, i.e., two diversity measures, two-word frequency measures, and one semantic similarity measure, were investigated in the argumentative writing samples by two groups of students in a 4-week intensive EAP program. While the two groups enrolled in the same EAP program as two cohorts, it was found that at the end of the program, the L2 students in one group were able to gain significant improvement in all five lexical measures, while the students in the other group showed progress in only one diversity measure (squared verb variation) and a frequency measure of AWL (Academic Word List) words. With their corpus tools named the Tool for the Automatic Analysis of Lexical Sophistication (TAALES), Kyle and Crossley (2015) added a series of new measures (e.g., word or n-gram frequency and range indices with various corpora references, academic vocabulary indices, and word information indices) to represent lexical sophistication, which is also related to lexical diversity and difficulty of lexical items. Through examining the relationships between the lexical sophistication indices yielded by TAALES and the holistic scores of TOEFL writing samples from two writing task types, Kyle and Crossley confirmed that several features were significantly correlated with the writing holistic scores. Nevertheless, task effect in Kyle and Crossley (2015) seemed noticeable as a set of four lexical sophistication features (i.e., all word range and bigram frequency logarithm referenced to BNC written corpus, number of hypernyms, and word information) were able to explain over 30% of the variance in the holistic scores of TOEFL independent writing task, while only two features (e.g., number of noun hypernyms and a genre-based range index) were retained in the final regression model to explain <10% of score variances for the TOEFL integrated writing task. Since TOEFL integrated writing only bears some resemblance to source-based academic writing, it is worth investigating what kind of lexical features may exhibit progressive trends in the genre of literature review as sampled in our online tutorial series.

Lexical features are also closely related to cohesion quality, an important aspect of writing performance. For example, the use of connectives and lexical overlaps across sentences or paragraphs can contribute to local cohesion or global cohesion (Crossley et al., 2019). However, previous studies on L2 student essays have established that cohesion features may have different effects on writing performance, often with local cohesion features, such as

sentence-level lexical overlaps and use of conjunctions, associated negatively with writing scores (Crossley et al., 2016). It is less clear whether similar patterns can be observed in literature review samples written by L2 graduate students.

For syntactic features in academic writing, phrasal complexity has attracted much attention because features like various phrasal embedding are found to be more distinctive than clause-level features in academic writing (Biber et al., 2011). This trend was confirmed by Kyle (2016) in a series of analyses with his corpus tool named Tool for Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC). One goal of Kyle (2016) was to identify a group of syntactic features that are associated with writing proficiency as measured with TOEFL independent writing tasks. With 13 syntactic features, Kyle's final regression model was able to explain 34.2% of variances of TOEFL holistic scores. Six of the features were phrasal complexity indices (i.e., counts and standard deviation of dependents per nominal, prepositions per non-pronoun nominal, dependents per non-pronoun object of a preposition, counts, and standard deviation of dependents per direct object), contributing to an explanation of 17.1% of variances. In other words, the essays written by high-proficiency test takers are more likely to show higher phrasal complexity strengthened by more dependents for noun phrases that function as objects. The other features included five indices of syntactic sophistication that are related to the association strengths or frequencies of construction, and two clausal complexity indices that conceptualize the counts and types of clausal dependents.

Differences in phrasal complexity have also been noted in comparative studies of the writings of novice writers and expert writers. For example, following Biber et al. (2011) noun phrase development model, Ansarifar et al. (2018) compared the phrasal complexity features in the abstracts of research articles written by Persian L1 graduate students (MA and Ph.D. students) and experts in applied linguistics. Their analysis results indicate that four complex noun phrase features, i.e., nouns as pre-modifiers, -ed participle as post-modifiers, attributive adjectives/nouns as pre-modifiers, and multiple preposition phrases as post-modifiers, were strongly associated with the level of expertise as more experienced writers tended to use highly modified noun phrases.

Somewhat different from the general findings in cross-sectional studies on linguistic complexity, Kyle (2016) longitudinal analyses of two small corpora of English language learners' writing samples show that while some common trends were observed as with those in the cross-proficiency comparison, the learners' progress was mainly manifested in clausal complexity such as mean length of t-units and complex nominals per clause, as well as syntactic sophistication indices such as verb-VAC (verb-argument construction) frequency. These differences may be related to the nature of writing tasks (untimed vs. timed tasks), genres, as well as educational levels of the writers.

## 2.2. Citation practices in academic writing

Academic writing, especially the section of literature review, by nature, requires citing other studies to make an argument and

to engage in academic conversations. Citation practices have been studied from the following perspectives: Citation forms or types, citation functions, and reporting verbs (Hyland, 1999; Swales and Feak, 2012).

Citation forms or the surface form of citations are often distinguished as integral and non-integral in-text citations based on whether a cited work functions syntactically as a part of a sentence (Swales and Feak, 2012). Both expertise level and nativeness seem to affect writers' citation practices. For example, Mansourizadeh and Ahmad (2011) noted that at a Malaysian university, the research papers in chemical engineering produced by non-native English-speaking expert writers showed a higher density (normalized frequency per 1,000 words) of citations in each section of the research papers than the papers written by non-native novice writers (students) in the same field (e.g., 23.19 vs. 20.1 citations per 1,000 words in the section of Introduction). In addition, in terms of citation types, non-integral citations constitute the majority of the citations (86.47% in experts' papers and 73.23% in novice writers' papers, respectively). However, expert writers demonstrated a balanced use of the two forms of integral citations, namely, verb-controlling where a cited work is introduced by a verb and naming forms which include a cited work as a part of a noun phrase, whereas novice writers heavily relied on the former. With regard to citation functions, while both writer groups have utilized citations for different purposes, expert writers seemed to have more evenly used four main functions while almost half of the citations in novice writers' papers fulfilled the function of attribution. In a study on 100 source-based papers by first-year university ESL students, Lee et al. (2018) revealed a lower citation density (about 10 citations per 1,000 words) and a more balanced use of integral and non-integral citations (53 vs. 47%). This pattern may be attributed to the design of the first-year writing class at the university, where students may be taught to use both citation forms.

Likewise, the use of reporting verbs may vary across expertise levels as well as disciplines both in terms of frequency and types (Hyland, 1999). In a comparison of a learner corpus of academic essays written by first-year undergraduate students and a reference corpus of published research articles in the same disciplines, Liardét and Black (2019) revealed that novice writers on average used fewer reporting verbs than expert writers. This finding is supported by Marti et al. (2019) comparative study on reporting verbs. Furthermore, as shown in Liardét and Black (2019), compared to the reporting verbs in learners' essays, a larger proportion of the reporting verbs used by experts are used to convey writers' evaluative stance. Related to stance-making in citations is the use of boosters and hedges. Aull and Lancaster (2014) corpus-based comparative analysis indicates that the frequencies of hedges and boosters were affected by writers' expertise levels. More specifically, first-year university students' argumentative essays tend to use fewer hedges and more boosters than the ones written by upper-level students and experts. It is not clear whether similar patterns can be observed in L2 graduate students' literature reviews.

As discussed above, novice writers, including L2 graduate students, need to pay attention to citation practices for better academic socialization. Nevertheless, only limited studies have focused on L2 learners' development of citation practices in academic writing.

## 2.3. Research questions

Considering the importance of linguistic performance and citation practices in academic writing, as well as the lack of empirical studies on L2 learners' development of literature review writing, this study aims to answer two research questions in the context of an online tutorial series of literature review writing:

1. What linguistic features may change in L2 graduate students' literature review texts over the course of an online academic writing tutorial series?
2. How do L2 graduate students' citation practices change in terms of citation forms and stance features (reporting verbs, hedges, and boosters) over the course of an online academic writing tutorial series?

## 3. Research methodology

### 3.1. The writing tutorial series

To help address some challenges faced by ESL graduate students in writing literature reviews, we developed a 5-unit online tutorial series with a focus on the following aspects of literature review writing: genre requirements, logic and structure in literature review, sentence structures, academic vocabulary, and grammar of reported speech (see [Table 1](#)). The selection of the themes was made based on the genre features of literature review ([Flowerdew and Forest, 2011](#)) as well as the typical challenges in academic writing ([Chen et al., 2016](#)). Each unit consists of a core tutorial explaining key concepts with both positive and negative examples, a writing task with peer review or group writing activity, and supplementary activities such as awareness-raising activities for comprehension check, discussion forum for participants to share writing experiences and ideas, and a wrap-up quiz.

The tutorial series is delivered via MoodleCloud, a cloud-based learning management system, as self-paced sessions but with fixed due dates for writing tasks. Each core tutorial is presented in the format of H5P (HTML5 Package) interactive e-book (see [Figure 1](#)), which allows participants to browse the content and complete awareness-raising activities embedded in the e-book. The total duration of the tutorial series is 3.5 months. The writing tasks require the participants to write a literature review on a given topic. The expected length is 700–800 words, excluding the references section. Three out of five writing tasks are to be completed individually with a round of online peer review whereas the other two tasks are collaborative tasks contributed by two to three participants. The individual writing tasks were evenly arranged in this tutorial series, with one at the beginning, one in the middle (Unit 3), and one at the end. Such a distribution allows us to study participants' writing development without direct input from their peers as in collaborative writing tasks. A bibliography on each topic is provided to the participants as a starting point for literature search and synthesis. Since this tutorial series was fully online and primarily managed by a non-teaching administrator, there was no restriction regarding the resources that a participant could access to help prepare the writing tasks. This study focuses

on individual writing tasks only to track participants' writing development throughout the online tutorial series.

### 3.2. Participants

To reach out to the targeted participants, we disseminated recruitment flyers digitally via university newsletters, online announcement, and emails to university faculty members. We also employed snowball sampling strategy to encourage the registered participants to share the project information with their peers. Despite strong interest from prospective participants, only 29 participants have completed the tutorial series and the corresponding writing tasks, partly due to the impact of the COVID-19 pandemic. There were 20 female participants, seven male participants, and two did not provide gender information. Based on their responses to a background survey, 19 participants were current master's students, four were doctoral students, and six were prospective graduate students without clear education level information. The main study fields of the participants include Linguistics (7), Education (6), Political Science (5), Computer Science (2), Veterinary Science (2), Health Science (2), Humanities (1), Agriculture (1), and Business (1). The first languages of the participants included Chinese (4), Farsi (4), Russian (2), Bengali (2), Hindi (2), Punjabi, Turkish, Vietnamese, Urdu, Spanish, Igbo, Luo, Portuguese, Czech, Jamaican Patois, Lusaka, Arabic, and Ukrainian. The average score of their most recent standardized English language tests is about IELTS 7 or equivalent after a score conversion from other tests, with a range from IELTS 6.5 to 8.5 and a standard deviation of 0.55.

### 3.3. Data collection and analysis

After a brief online orientation *via* email, the participants started self-paced study of tutorial content and prepared for the writing task accordingly. The duration of a tutorial unit was 3 weeks when its writing task is designed to be completed individually by the participants. In the case of collaborative writing tasks, participants had 2 weeks to complete the unit (Units 2 and 4). The writing samples were rated by three raters based on a 10-point analytical rating scale at the end of each unit. The rating scale covers ten aspects of the writing quality, such as quality of selected source materials, integration of source materials, overall structure, clarity of ideas, grammatical accuracy, coherence and cohesion, and vocabulary quality. Written feedback, along with the ratings, from the raters was shared with the participants. Considering that the rating practices in this project focused more on written feedback and this paper primarily employed a more corpus-based approach, we chose not to include the ratings in data analyses. The literature reviews written to three individual writing tasks (Tasks 1, 3, and 5) were collected and then processed to retain the body of the literature review only for text analysis (see [Table 2](#)). Overall, the average length of the writing samples is 733 words. Nevertheless, a possible topic effect is reflected in average essay lengths, with Task 1 (*Online learning: Pros and cons*) elicited relatively longer responses (782 words and 42 sentences on average)

TABLE 1 Overall structure of the writing tutorial series.

Unit	Theme	Writing task type	Learning activities
1.	The genre of literature review and its requirements	Individual writing	Awareness-raising activities with examples, wrap-up quizzes, discussion forums, peer review (for individual writing tasks)
2.	Logic and structure of presentation	Collaborative writing on “Canadian linguistic landscape”	
3.	Sentence structures	Individual writing	
4.	Academic vocabulary	Collaborative writing on “Lessons from COVID-19”	
5.	Grammar of reported speech	Individual writing	

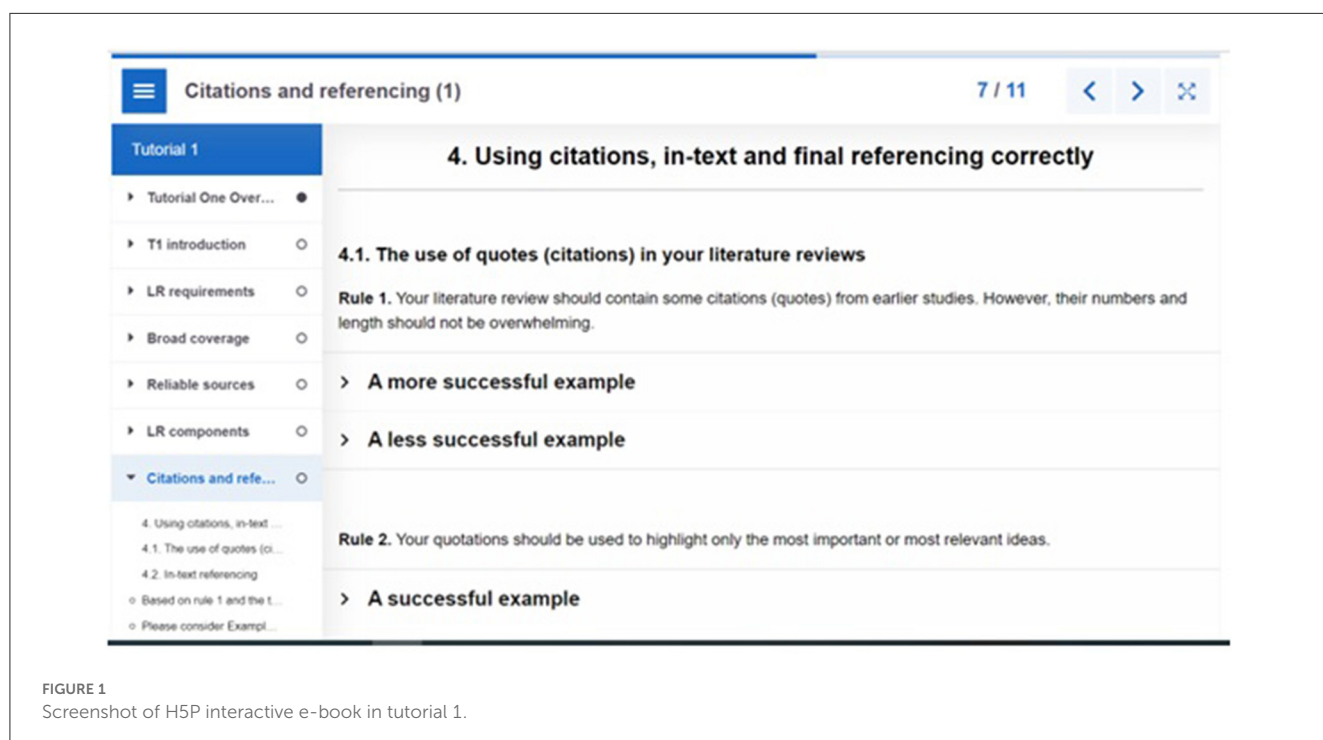


FIGURE 1 Screenshot of H5P interactive e-book in tutorial 1.

and Task 5 (Pacifism, peace-making, just/justifiable war) shorter responses (702 words and 33 sentences on average), while average sentence lengths are mostly comparable across the topics.

The corpus data were analyzed with three open-source tools developed by Kyle et al. (available at [www.linguisticanalysisistools.org/](http://www.linguisticanalysisistools.org/)) for the linguistic features related to text cohesion (Tool for the Automatic Analysis of Cohesion or TAACO 2.0, Crossley et al., 2019), syntactic features (Tool for Automatic Analysis of Syntactic Sophistication and Complexity or TAASSC 1.3.8, Kyle, 2016), and lexical features (Tool for Automatic Analysis of Lexical Sophistication or TAALES 2.2, Kyle and Crossley, 2015). While these three analytical tools produce a large number of features or indices, we selected a small portion of the normalized features based on their relevance to the quality of academic writing in general (Crossley et al., 2016; Kyle and Crossley, 2016, 2017). For example, While TAACO yields 194 features in seven categories, such as Type token ratio (TTR) and density, sentence-level lexical overlap, paragraph-level overlap, semantic overlap, connectives, givenness, and source text similarity, four categories are either overlapped with the output

from another tool (e.g., TTR from TAALES) or less relevant to the writing genre in this study (e.g., givenness and source text similarity). As a result, only 41 out of 194 features from TAACO in three categories were selected partially based on the findings mentioned in Crossley et al. (2019). Section 4.1 provides the details of the selected features.

A Python script was developed to automatically extract in-text citation instances for further analyses of citation types (i.e., integral vs. non-integral citations) and to identify reporting verbs, boosters, and hedges used in those instances based on corresponding word lists A list of 122 reporting verbs was prepared based on Liardét and Black (2019); a list of 56 hedges and a list of 38 boosters were adapted from Aull and Lancaster (2014).

Considering the common patterns of non-normal distribution of frequency or ratio-based linguistic features (McEneary and Hardie, 2012), we employed the Friedman test, a non-parametric statistical test, which is similar to repeated measures ANOVA, to compare the selected features across the three individual writing tasks to gauge participants’ development in academic writing. Nemenyi *post-hoc* test was used for pairwise comparisons when

TABLE 2 Participant writing sample corpus (excluding references sections).

Task	Topic	Word count: Mean (SD)	Sentence count: Mean (SD)	Words per sentence: Mean (SD)
1.	Social consequences of legalized marijuana (cannabis)	716 (145)	36 (15)	21 (5)
3.	Online learning: Pros and cons	782 (308)	42 (28)	20 (4)
5.	Pacifism, peace-making, just/justifiable war	702 (202)	33 (10)	22 (4)

a Friedman test detected differences of statistical significance. We understand that using Friedman tests multiple times on the data tends to introduce Type I errors and we decided to choose a more stringent p-value ( $p \leq 0.001$ ) to minimize such influence. Python packages SciPy1.8.1 (Virtanen et al., 2020) and scikit-posthocs (Tepilowski, 2019) were used for statistical analyses.

## 4. Results

### 4.1. Research question 1: Linguistic features

Regarding the lexical features, we narrowed down to nine features that are related to academic writing from the output of the software TAALES: six range or frequency features in content words, bigrams, and trigrams with reference to the academic register in COCA (Corpus of Contemporary American English), two percentage features related to Academic Word List (AWL) and Academic Formulas List (AFL), and academic lemma TTR (Type-Token Ratio, ratios of unique lemma count to token count). The results of Friedman tests indicate that two features exhibited at least one statistically significant difference among the writing samples from the three tasks (see Table 3): frequency of trigrams in COCA Academic (chi-squared = 13.52,  $p = 0.001$ , Kendall's  $W = 0.233$ ) and proportion of AWL (chi-squared = 20.07,  $p < 0.001$ , Kendall's  $W = 0.346$ ). Meanwhile, the differences in other three features were close to statistical significance level: frequency of content words in COCA Academic (chi-squared = 10.21,  $p = 0.006$ ), bigram range in COCA Academic (chi-squared = 6.69,  $p = 0.035$ ), proportion of core AFL (chi-squared = 9.93,  $p = 0.007$ ), and academic lemma TTR (chi-squared = 9.17,  $p = 0.010$ ).

The first feature in Table 3 "COCA academic frequency content words" reflects the level of frequency scores of the content words that appear in the academic register or sub-corpus in COCA, with a higher value suggesting more frequent content words. Follow-up pairwise comparisons confirm that the content word frequency in Task 1 was significantly lower than in Task 3 ( $p = 0.023$ ) and Task 5 ( $p = 0.011$ ), while no statistically significant difference exists between content word frequency in Tasks 3 and 5. The feature of the bigram range in COCA Academic shows the distribution range of academic bigrams in COCA Academic. A significant difference in this feature is noted between the texts in Tasks 3 and 5 ( $p = 0.033$ ), with the latter having a larger value in bigram range. The third COCA Academic-reference feature is the frequency of trigrams. The participants' writing exhibited a decreasing trend in this feature from Task 1 to Task 3 ( $p = 0.001$ ), meaning that less frequent trigrams were used in the texts for Task 3. There is no statistically

significant difference in this feature between the texts for Tasks 3 and 5.

The texts also differed in the proportions of words from AWL in Task 1 and the other two tasks (Task 1,  $p = 0.001$  and Task 5,  $p = 0.005$ ), with the texts in Task 1 having a higher proportion of AWL words. For the percentage feature related to AFL, texts in Task 1 had a higher proportion of AFL multi-word units than those in Task 3 ( $p = 0.005$ ) while Tasks 3 and 5 had similarly lower proportions. On the other hand, the texts in Task 1 had a lower TTR of academic lemma than those in Task 3 ( $p = 0.010$ ), while Tasks 3 and 5 had a similar level of academic lemma TTRs.

For these lexical features, different developmental patterns are observed. Two features (i.e., content words frequency and academic lemma TTR) showed increasing trajectories and two (i.e., trigram frequency and Core AFL percentage) had somewhat downward trends. The other two features remained similar values in Tasks 1 and 5, but differed more remarkably in Task 3.

To explore participants' development of academic writing in terms of text cohesion, we focused on 41 cohesion features in three categories from TAACO, namely, 12 features related to lexical overlaps across sentences, 4 features related to semantic similarity, and 25 connective-related features. These features were compared across the three tasks using the Friedman test. While 14 features were found to be statistically significant at alpha = 0.05 level, only three were retained and reported in Table 4 because some of the features were either highly correlated and/or were extremely low in value. For example, the correlation between the feature word2vec-based semantic similarity across a 2-sentence span and word2vec-based semantic similarity across a 1-sentence span is 0.77, which shows a strong relationship with probably some overlaps in their assessed constructs.

Table 4 shows the mean and standard deviation of the three cohesion features that exhibited noticeable differences across the three writing tasks. The feature "adjacent overlaps of content words" captures local cohesion achieved through repeating content words or lemmas (e.g., lexical verb, noun, adjective) in a target sentence and the following two sentences (chi-squared = 7.103,  $p = 0.029$ ). A higher value suggests more occurrences of repeated content words (lemmas) divided by type count in the same text. Nemenyi *post-hoc* tests for pairwise comparison indicate a significant difference in this feature between Task 1 and Task 5 ( $p = 0.023$ ), with texts for Task 5 showing a lower level of adjacent overlaps in content words but a larger standard deviation.

Different from the lemma overlap-based measure of cohesion, most semantic similarity measures in TAACO take advantage of statistical representation of word meanings as used in latent semantic analysis or LSA, latent Dirichlet allocation or LDA,

TABLE 3 Lexical features across three writing tasks.

Task	COCA academic frequency (content words)	COCA academic bigram range	COCA academic trigram frequency	AWL %	Core AFL %	Academic lemma TTR
1	691.483 (95.589)	0.140 (0.016)	15.790 (5.092)	0.092 (0.018)	0.046 (0.020)	0.60 (0.10)
3	771.736 (142.709)	0.131 (0.012)	11.326 (2.400)	0.110 (0.022)	0.032 (0.016)	0.65 (0.09)
5	774.636 (177.876)	0.145 (0.015)	12.401 (3.752)	0.092 (0.028)	0.038 (0.021)	0.64 (0.08)

TABLE 4 Cohesion features across three writing tasks.

Task	Adjacent overlaps (content words, 2 sentences)	Semantic similarity (word2vec, 2 sentences)	Basic connectives
1	0.211 (0.067)	0.811 (0.040)	0.037 (0.011)
3	0.200 (0.055)	0.812 (0.045)	0.042 (0.010)
5	0.194 (0.092)	0.924 (0.018)	0.044 (0.012)

and word2vec, respectively to numerically evaluate the semantic distance between text blobs. In this study, the difference in the word2vec-based semantic similarity across a 2-sentence span shows statistical significance across the three tasks (chi-squared = 43.655,  $p < 0.001$ , Kendall's  $W = 0.753$ ). The follow-up pairwise comparisons confirm that the texts for Task 5 exhibited stronger semantic similarity than the ones written for Task 1 ( $p = 0.001$ ) and Task 3 ( $p = 0.001$ ). The use of basic connectives (e.g., *for*, *and*, *nor*) also distinguishes the writing samples (chi-squared = 7.103,  $p = 0.029$ ) to some extent with more basic connectives used in the Task 5 texts than in the Task 1 texts ( $p = 0.023$ ) and no significant difference between Task 1 and Task 3.

Overall, the patterns of cohesion features observed in this study suggest that the participants developed better local cohesion through the use of connectives and also by higher sentence-level semantic similarity, but with fewer incidences of repetition of content words across sentences.

From the output of TAALES for syntactic feature analysis, we started with 25 distinguishing features that are relevant to academic writing, including 20 noun phrase complexity features and five clause-complexity features. Four features showed statistically significant differences across the three writing tasks (see Table 5).

The feature "average number of dependents per nominal phrase" saw a decreasing trend from Task 1 to Task 5 (chi-squared = 13.24,  $p = 0.001$ , Kendall's  $W = 0.228$ ), with the texts for Task 1 being significantly higher in this feature than those for the other two tasks ( $p = 0.005$ ,  $p = 0.005$ ). For this noun-phrase complexity feature, the dependents could be modifiers fulfilled by an adjective, noun, clause, and preposition phrase. The other three distinguishing features belong to clause-complexity: number of clausal complements per clause (chi-squared = 6.90,  $p = 0.032$ ), coordinate phrases per clause (chi-squared = 6.28,  $p = 0.043$ ), and complexity nominal per clause (chi-squared = 7.66,  $p = 0.022$ ). For the last two clause complexity features, the main difference in the ratio of coordinate phrases per clause exists between Tasks 1 and 3 only ( $p = 0.066$ ). The difference in complex nominal per clause between Tasks 1 and 5 is statistically significant ( $p = 0.016$ ).

Overall, two of the significant features showed some decrease in both phrasal complexity and clausal complexity in the final writing

task. The other two features exhibit a drip and a peak in Task 3, respectively. In other words, the development of syntactic features is not linear as shown in the writing samples.

## 4.2. Research question 2: Citation practices

The second question concerns participants' citation practices in terms of citation types and stance features in citation instances. Table 6 presents the descriptive statistics of both raw counts and normalized counts of integral and non-integral citations used by the participants in three writing tasks. The results of the Friedman tests indicate that there are no statistically significant differences in citation types across the three tasks. Overall, non-integral citations were dominant in participants' writing in this study. For example, the average raw count of non-integral citation is about 11 whereas that of integral citations is close to 2. The normalized counts of citation types also show similar trends. Similar to the distribution pattern of some linguistic features discussed above, the texts for Task 3 had higher raw counts of citation instances, compared to the other two tasks. This again points to the possible topic effect on writing performance.

As for the stances in citation instances, there were no statistically significant differences in the frequencies of occurrence of reporting verbs, boosters, and hedges across the writing tasks while increasing trends are evident for these features, especially the normalized frequencies (see Table 7). For example, the texts written for Task 5 employed slightly more reporting verbs, hedges, and boosters than the first two individual writing tasks.

A closer look at the specific instances of reporting verbs, hedges, and boosters reveals that all three features have noticeable differences in their type-token ratios or TTR, especially from Task 1 to the other tasks. For example, in Task 1, a total of 93 unique reporting verbs appeared in the texts. With many of the reporting verbs repeated in the texts, the count of reporting verbs is 225. Therefore, the TTR of reporting verbs in the Task 1 texts is 0.41 (93 types/225 tokens). By contrast, the TTRs for Task 3 and Task 5 are 0.48 (119/246) and 0.46 (118/255), respectively, suggesting higher diversity in participants' use of reporting verbs along with

TABLE 5 Syntactic features of statistical significance across three writing tasks.

Task	Dependents per nominal	Clausal complements per clause	Coordinate phrases per clause	Complex nominal per clause
1	1.427 (0.168)	0.107 (0.047)	0.552 (0.180)	2.380 (0.490)
3	1.339 (0.161)	0.077 (0.044)	0.658 (0.261)	2.348 (0.611)
5	1.324 (0.164)	0.106 (0.047)	0.531 (0.244)	2.056 (0.466)

TABLE 6 Citation practices across three writing tasks.

Task	Integral citations: Mean (SD)	Non-integral citations: Mean (SD)	Integral citations per 1,000 words: Mean (SD)	Non-integral citations per 1,000 words: Mean (SD)
1	1.34 (2.18)	11.03 (3.99)	1.65 (2.29)	15.70 (5.68)
3	2.00 (2.49)	11.41 (6.30)	2.98 (3.97)	14.88 (7.29)
5	1.79 (2.04)	10.45 (5.22)	2.69 (3.44)	15.53 (7.43)

TABLE 7 Reporting verbs, hedges, and boosters in citation instances across three writing tasks.

Task	Reporting verbs: Mean (SD)	Reporting verbs per 1,000 words: Mean (SD)	Hedges: mean (SD)	Hedges per 1,000 words: Mean (SD)	Boosters: mean (SD)	Boosters per 1,000 words: Mean (SD)
1	7.41 (3.57)	10.30 (4.70)	8.86 (4.79)	12.58 (6.95)	2.41 (1.88)	3.34 (2.71)
3	8.48 (4.31)	11.29 (5.92)	10.97 (6.49)	12.31 (7.41)	3.83 (3.62)	4.84 (4.34)
5	8.79 (4.93)	13.05 (7.4)	7.69 (4.33)	13.89 (9.3)	3.79 (2.26)	5.47 (3.22)

more frequent use of reporting verbs. Likewise, the TTR of hedges increases from Task 1 (24/257, 0.093) to Task 3 (30/313, 0.096) and Task 5 (29/223, 0.130). For the specific cases of boosters, the texts in Task 1 have a TTR of 0.24 (17/79), whereas Tasks 3 and 5 saw higher frequencies of occurrence of boosters (111 and 110, respectively) but with the same count of types (20) and therefore lower TTR values (0.18). This is mainly because of the relatively small list of boosters (38) used in this study for booster extraction. Since we only calculated the TTRs of these features at task level to explore possible differences, we did not run inferential statistics.

Table 8 lists the top 10 words in each category across three writing tasks. Each category shares several common items across three writing tasks with some minor differences emerging. For example, “suggest”, “conduct”, “demonstrate”, and “identify” are unique among the top 10 reporting verbs in Task 1. While it is difficult to pinpoint possible causes of these differences, we speculate that they may be linked to the topic effects as the sources cited for different topics may differ in terms of perceived certainty and acceptance of findings, which in turn may affect writers’ stance-making decisions.

## 5. Discussions and conclusions

As shown above, the development of literature review writing was not clear-cut over the course of the online tutorial series. On the one hand, participants’ writing displays progress in local cohesion quality and some academic lexical features from the first individual task to the last one (Task 5). However, their writing samples also see some decreases in syntactic features and some academic lexical features, which can be an indication of the

participants’ adjustment of sentence variety in response to one of the most frequent comments from raters on long sentences and the grammatical issues found in them. The participants’ citation practices remained at similar levels in terms of citation types and stances manifested through reporting verbs, hedges, and boosters. However, they seem to have more stance-making resources at their disposal as evidenced by more types of reporting verbs, hedges, and boosters. It should be noted that the increase or decrease in values does not necessarily suggest improvement in academic writing yet. These observations are to be discussed in light of the relevant literature.

### 5.1. Linguistic features

If we consider the complex nature of language learning in general, the diverse patterns in the changes of linguistic features from Task 1 to Task 5 are less surprising as linear progressions in academic writing are rarely achieved. For example, in a longitudinal case study of an L2 graduate student at an Australian university, Rosmawati (2014) tracked the development of complexity and accuracy in the student’s argumentative essays over a semester. It was found that a high level of variability existed for both complexity and accuracy development, supporting the non-linear and dynamic nature of L2 learning. Furthermore, a clear difference in the interactions between complexity and accuracy measures was revealed at different measurement points or stages in the semester.

Non-linear patterns are obvious with the lexical features reported in this study as three out of six features (i.e., COCA



TABLE 8 Top 10 reporting verbs, hedges, and boosters in citation instances across three writing tasks.

Task	Reporting verbs (count)	Hedges (count)	Boosters (count)
1	Find (17), show (17), report (13), suggest (8), relate (7), consider (6), conduct (5), demonstrate (5), identify (5), and indicate (4)	Can (66), could (59), may (37), likely (19), might (16), possible (13), around (8), often (7), rather (6), and almost (3)	Should (28), very (11), certain (7), clear (5), never (3), must (3), indeed (2), evident (2), decided (1), and undeniable (1)
3	Show (10), find (10), report (9), provide (7), consider (6), offer (6), argue (5), focus (5), mention (5), and indicate (5)	Can (121), may (79), could (27), might (19), around (15), likely (6), possible (6), usually (5), often (4), and almost (3)	Should (43), must (17), very (15), always (8), certain (8), evident (3), of course (2), sure (2), undeniable (2), and conclusively (1)
5	Argue (8), justify (8), see (7), believe (7), consider (6), understand (6), report (6), claim (5), present (5), and explain (4)	Can (60), may (32), might (20), could (19), often (12), possible (10), around (10), rather (8), generally (6), and about (6)	Should (36), certain (12), very (11), must (10), never (8), always (6), really (5), clear (4), clearly (3), and actually (3)

Academic-based bigram range, COCA Academic-based trigram frequency, and percentage of core AFL units) showed a U-shaped development. In other words, the literature reviews written on the third topic (“Online learning: Pros and cons”) had lower values in these three features. Compared with the other two topics (i.e., legalization of marijuana and pacifism), online learning is a relatively familiar topic to the L2 graduate students, especially for those who have experienced online learning during the COVID pandemic. This may be the reason that the participants did not use the bigrams of higher range value (that is, the bigrams used more widely in COCA’s academic register) or trigrams of higher frequency value (that is, the trigrams appearing more frequently). Likewise, the L2 graduate students may have incorporated their personal experiences in this literature review, which may reduce the need to use the units from AFL (academic Formulas List). Interestingly, the literature reviews on Topic 3 had a higher percentage of words from AWL. This may be accounted for by the fact that the topic itself is academic and common academic vocabulary is expected.

Overall, three lexical features (i.e., content word frequency, bigram range, academic lemma TTR) showed higher values in the last writing task, suggesting some improvement in the use of academic vocabulary at the end of the tutorial series. While these features are different from the ones as strong predictors of TOEFL writing performance in Kyle and Crossley (2015), the features share some basic characteristics, such as range and frequency indices with different reference corpora (COCA Academic in this study vs. BNC written in Kyle and Crossley (2015)). The increase in academic lemma TTR is roughly in line with the findings in Mazgutova and Kormos (2015) in which TTR-based measures improved at the end of a 4-week writing program.

As for the cohesion features, the retained distinguishing ones (i.e., content word overlaps in adjacent two sentences, word2vec-based similarity across two sentences, and use of basic connectives) in this study highlight the notable changes in local cohesion across the three writing tasks. These features are not necessarily specific to the genre of literature reviews, though, or reflect improvement in academic writing made by the L2 graduate students. Compared to the findings in the studies that used the same or similar measures, these three cohesion features have not been reported as major predictors of writing performance yet. In Crossley et al. (2016) study of descriptive essays written by L2 university students in EAP courses, they identified 32 cohesion features that were significantly correlated to students’ essay scores. Nevertheless, some of the

relevant cohesion features were found to be significantly correlated with essay scores in Crossley et al. (2016), for example, adjacent overlaps of words in general across two sentences ( $r = 0.32$ ), LSA-based similarity from initial to middle paragraphs ( $r = 0.23$ ), and use of conjunctions ( $r = 0.24$ ). Therefore, we may speculate that the increase in these cohesive features may be related to the content of the online tutorials, specifically the ones about academic vocabulary and sentence structures, thus contributing positively to writing performance.

The syntactic features also showed some unexpected patterns as all four features ended up with smaller values in the last writing task. This suggests that L2 graduate students’ writing became less complex syntactically at both phrasal and clausal levels, likely as a result of addressing raters’ comments on long and complex sentences found in previous writing tasks. For example, the last literature reviews appeared to have fewer dependents per nominal structure, fewer clausal complements, coordinate phrases, and complex nominals per clause. Since previous studies have established positive correlations between syntactic complexity features and writing performance (Kyle, 2016), we may suppose that with a relatively abstract topic (i.e., pacifism), the last literature review task may have presented some challenges to L2 graduate students who may be less able to manage syntactic structures at the expense of other writing aspects such as topic familiarity. While it is difficult to be certain about the causes of the decreases, it is equally important to keep in mind that the majority of the syntactic features (21 out of 25) examined did not have statistically significant changes across the three writing tasks. It appears that the L2 graduate students have reached a relatively stable level of syntactic development and only a handful of features fluctuated in their writing samples.

## 5.2. Citation practices

In terms of the frequency of citations, the literature reviews written by the L2 graduate students have fewer normalized counts of citation instances or lower citation density, compared to the citation counts in the Introduction sections of chemical engineering papers reported in Mansourizadeh and Ahmad (2011). Meanwhile, similar to the dominant use of non-integral citations found in both expert papers and novice papers in their study, the writing samples in this study also showed a strong preference for non-integral citations (85.1 to 89.1% of raw citation counts).

However, it is worth mentioning that disciplinary differences in preferred citation forms exist (Hu and Wang, 2014). For example, Hu and Wang noted that published research articles in applied linguistics tend to have a more balanced distribution of integral and non-integral citations, compared to the articles in general medicine. Lillis et al. (2010) reported higher proportions of non-integral citations (over 70%) in the research articles in psychology. Considering that the topics in our study have an orientation toward social science or humanities, we may expect to see a somewhat balanced use of both citation forms as well. It requires further inquiry to understand why the L2 graduate students in this study chose to use more non-integral citations as several factors may affect a writer's preference for citation forms. In a quantitative analysis of the relationships among individual factors and citation competence, Ma and Qin (2017) reported that cognitive proficiency in source use (operationalized as knowledge of source use and ability to detect plagiarism) can significantly influence English major students' citation competence as measured with intertextual strategies, writers' stance, citation typology, and citation function in a read-to-write task. Meanwhile, students' academic reading proficiency exerted a direct impact on citation competence.

Regarding the use of reporting verbs, hedges, and boosters in the citation instances, our study shows a steady, but not statistically significant, increase in these stance-making features. More importantly, the TTRs of these lexical items increased as well, showing that the L2 graduate students may have become more aware of stance-making with these items at the end of the tutorial series. As discussed above, this progress was also related to the student's development of academic vocabulary, likely from the benefit of the tutorials on genre requirements and sentence structures. Since the changes in these features were not statistically significant, these phenomena should be interpreted with caution.

## 6. Implications and future studies

Before discussing the pedagogical implications of the findings, we need to point out the limitations of this study. Firstly, this study focuses on the written output of the online tutorial series. More qualitative input from the L2 graduate students as well as the instructors (e.g., feedback and scores) would supplement the quantitative analysis and facilitate the understanding of students' development of academic writing. In addition, the tutorial series has a fixed number of topics for the participants. Even with the provided bibliographies, topic familiarity may still influence the writer's performance and writing motivation. To avoid this potential issue, future studies may consider collecting writing samples from graduate students' work either from course assignments or their research projects. Lastly, for this purely online tutorial series open to voluntary participation during the pandemic, it was challenging to recruit more participants and several participants were not able to complete the whole series. Consequently, with a small sample size, the diversity in participants' background and educational level may impact the generalizability of the findings. Future projects with similar materials design may consider a hybrid mode with in-person consultation opportunities to better attract and support participants.

The format and content of the online tutorial may be useful for EAP practitioners and researchers. The materials in the tutorial series were prepared with specific aspects of literature review writing in mind, covering both genre features (e.g., logic and structure) and linguistic features (e.g., academic vocabulary, sentence structure, and reported speech). These materials can be used to supplement EAP teaching. In addition, the interactive e-book and accompanying activities used in this series are made to be open-access resources. Other researchers and EAP instructors may use some of the content and activities directly from this tutorial series, with or without modifications. At the same time, the quantified linguistic features generated by the software provided rich information regarding graduate students' writing development as well as tutorial performance. These analytical tools are free and user-friendly. With basic training, practitioners may be able to gain more insight into graduate students' writing. Furthermore, the decrease in some syntactic features, along with the lack of changes in many other linguistic features, deserves some attention. These trends may be a result of missing or under-represented components in the tutorial series. For example, discipline-specific activities on the rhetorical functions of citation and roles of citation types can be added or expanded so that L2 graduate students can better understand the expectations of citation practices in their fields. In addition, more example structures of appropriate phrasal and clausal complexity levels can be incorporated into the tutorial on "sentence structures." For the features that did vary across the writing tasks, it would be beneficial to identify the ones that are associated positively with academic writing performance so that corresponding activities can be developed to raise students' awareness of those important features.

While the developmental patterns vary across the targeted linguistic features and citation practices, the largely positive findings of this corpus-based study (i.e., progress in cohesion and academic vocabulary) are very promising for a standalone online tutorial series that requires limited interventions from instructors.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Research Office, University of Saskatchewan. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## Funding

This project is funded by the Insight Development Grants (430-2020-00179) from Social Sciences and Humanities Research Council of Canada (SSHRC).

## Acknowledgments

We would like to thank SSHRC for supporting this project.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Ansarif, A., Shahriari, H., and Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *J. English Acad. Purp.* 31, 58–71. doi: 10.1016/j.jeap.2017.12.008
- Aull, L. L., and Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing. *Written Commun.* 31, 151–183. doi: 10.1177/0741088314527055
- Badenhorst, C. (2018). Citation practices of postgraduate students writing literature reviews. *London Rev. Educ.* 16, 121–135. doi: 10.18546/LRE.16.1.11
- Biber, D., Gray, B., and Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quart.* 45, 5–35. doi: 10.5054/tq.2011.244483
- Bulté, B., and Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *J. Second Lang. Writing* 26, 42–65. doi: 10.1016/j.jslw.2014.09.005
- Chen, D. T. V., Wang, Y. M., and Lee, W. C. (2016). Challenges confronting beginning researchers in conducting literature reviews. *Stud. Contin. Educ.* 38, 47–60. doi: 10.1080/0158037X.2015.1030335
- Cheng, A. (2007). Transferring generic features and recontextualizing genre awareness: Understanding writing performance in the ESP genre-based literacy framework. *English Specific Purp.* 26, 287–307. doi: 10.1016/j.esp.2006.12.002
- Cheng, L., and Fox, J. (2008). Towards a better understanding of academic acculturation: Second language students in Canadian universities. *Canadian Mod. Lang. Rev.* 65, 307–333. doi: 10.3138/cmlr.65.2.307
- Crossley, S. A., Kyle, K., and Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behav. Res. Methods* 51, 14–27. doi: 10.3758/s13428-018-1142-4
- Crossley, S. A., Kyle, K., and McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *J. Second Lang. Writ.* 32, 1–16. doi: 10.1016/j.jslw.2016.01.003
- Crosthwaite, P., Sanhueza, A. G., and Schweinberger, M. (2021). Training disciplinary genre awareness through blended learning: An exploration into EAP students' perceptions of online annotation of genres across disciplines. *J. English Acad. Purp.* 53, 101021. doi: 10.1016/j.jeap.2021.101021
- Dafouz, E. (2020). Undergraduate student academic writing in English-medium higher education: Explorations through the ROAD-MAPPING lens. *J. English Acad. Purp.* 46, 100888. doi: 10.1016/j.jeap.2020.100888
- Flowerdew, J., and Forest, R. W. (2011). "Schematic structure and lexicogrammatical realization in corpus-based genre analysis: The case of research in the PhD literature review," in *Academic Writing: At the Interface of Corpus and Discourse*, eds. M. Charles, D. Pecorari, and S. Hunston (London, UK: Bloomsbury Publishing) 15–26.
- Hu, G., and Wang, G. (2014). Disciplinary and ethnolinguistic influences on citation in research articles. *J. English Acad. Purp.* 14, 14–28. doi: 10.1016/j.jeap.2013.11.001
- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Appl. Linguist.* 20, 341–367. doi: 10.1093/applin/20.3.341
- Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. Unpublished doctoral dissertation. Georgia State University.
- Kyle, K., and Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *J. Second Lang. Writ.* 34, 12–24. doi: 10.1016/j.jslw.2016.10.003
- Kyle, K., and Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Lang. Test.* 34, 513–535. doi: 10.1177/0265532217712554
- Kyle, K., and Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quart.* 49, 757–786. doi: 10.1002/tesq.194
- Lea, M. R., and Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Stud. Higher Educ.* 23, 157–172. doi: 10.1080/03075079812331380364
- Lee, J. J., Hitchcock, C., and Elliott Casal, J. (2018). Citation practices of L2 university students in first-year writing: Form, function, and stance. *J. English Acad. Purp.* 33, 1–11. doi: 10.1016/j.jeap.2018.01.001
- Li, Y., Ma, X., Zhao, J., and Hu, J. (2020). Graduate-level research writing instruction: Two Chinese EAP teachers? localized ESP genre-based pedagogy. *Jou of Eng for Acad Purp.* 43. doi: 10.1016/j.jeap.2019.100813
- Liardet, C. L., and Black, S. (2019). "So and so" says, states and argues: A corpus-assisted engagement analysis of reporting verbs. *J. Second Lang. Writ.* 44, 37–50. doi: 10.1016/j.jslw.2019.02.001
- Lillis, T., Hewings, A., Vladimirov, D., and Curry, M. J. (2010). The geolinguistics of English as an academic lingua franca: citation practices across English-medium national and English-medium international journals. *Int. J. Appl. Linguist.* 20, 111–135. doi: 10.1111/j.1473-4192.2009.00233.x
- Ma, R., and Qin, X. (2017). Individual factors influencing citation competence in L2 academic writing. *J. Quantit. Linguist.* 24, 213–240. doi: 10.1080/09296174.2016.1265793
- Mansourizadeh, K., and Ahmad, U. K. (2011). Citation practices among non-native expert and novice scientific writers. *J. English Acad. Purp.* 10, 152–161. doi: 10.1016/j.jeap.2011.03.004
- Marti, L., Yilmaz, S., and Bayyurt, Y. (2019). Reporting research in applied linguistics: The role of nativeness and expertise. *J. English Acad. Purp.* 40, 98–114. doi: 10.1016/j.jeap.2019.05.005
- Mazgutova, D., and Kormos, J. (2015). Syntactic and lexical development in an intensive English academy purposes programme. *J. Second Lang. Writ.* 29, 3–15. doi: 10.1016/j.jslw.2015.06.004
- McEnery, T., and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. London: Cambridge University Press. doi: 10.1017/CBO9780511981395
- Okuda, T., and Anderson, T. (2018). Second language graduate students' experiences at the Writing Center: A language socialization perspective. *TESOL Quart.* 52, 391–413. doi: 10.1002/tesq.406

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2023.1035394/full#supplementary-material>

- Riazi, A. M. (2016). Comparing writing performance in TOEFL-iBT and academic assignments: An exploration of textual features. *Assess. Writ.* 28, 15–27. doi: 10.1016/j.asw.2016.02.001
- Rosmawati, R. (2014). Dynamic development of complexity and accuracy: A case study in second language academic writing. *Australian Rev. Appl. Linguist.* 37, 75–100. doi: 10.1075/aryl.37.2.01ros
- Shahsavari, Z., and Kourepaz, H. (2020). Postgraduate students' difficulties in writing their theses literature review. *Cogent Educ.* 7, 1784620. doi: 10.1080/2331186X.2020.1784620
- Shi, L., and Dong, Y. (2015). Graduate writing assignments across faculties in a Canadian university. *Canadian J. Higher Educ.* 45, 123–142. doi: 10.47678/cjhe.v45i4.184723
- Storch, N., and Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *J. English Acad. Purp.* 8, 207–223. doi: 10.1016/j.jeap.2009.03.001
- Swales, J., and Feak, C. (2012). *Academic Writing for Graduate Students (3rd ed.)*. Ann Arbor, MI: University of Michigan Press. doi: 10.3998/mpub.2173936
- Terpilowski, M. (2019). scikit-posthocs: Pairwise multiple comparison tests in Python. *J. Open Source Softw.* 4, 1169. doi: 10.21105/joss.01169
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2
- Xu, F. (2019). “Learning the language to write for publication: The nexus between the linguistic approach and the genre approach,” in *Novice Writers and Scholarly Publication* (Springer International Publishing) 117–134. doi: 10.1007/978-3-319-95333-5\_7
- Zhang, Z. (2011). A nested model of academic writing approaches: Chinese international graduate students' views of English academic writing. *Lang. Liter.* 13, 39–59. doi: 10.20360/G27G6R