Text Differences between Individual and Collaborative Writing: A Lexico-Grammatical Perspective

WILLIAM J. CRAWFORD
Northern Arizona University, USA
KIM McDONOUGH
Concordia University, Canada

Although collaborative writing studies have found that collaboratively-written texts are more accurate than individually-written texts, previous studies in this framework have not identified differences in the grammatical features of texts written individually or collaboratively (Fernández-Dobao, 2012 Wigglesworth & Storch 2009; Storch & Wigglesworth, 2007; Storch, 2005). The current study further investigates variation in individual and collaborative texts by using a lexico-grammatical, corpus-driven approach (Chen & Baker, 2010; Hyland, 2008 a, b) to identify keyword differences and recurrent word sequences (lexical bundles). English L2 university students in Thailand (n = 99) wrote problem and solution paragraphs on the same topic in two different conditions: individually and collaboratively. The texts were analyzed to identify keyword differences and lexical bundles, and the lexical bundles were classified further as to their form and function. The results indicated differences between the two groups in their use of key words as well as the frequency of lexical bundles. Implications for these findings are discussed in light of the method used in previous collaborative writing studies as well as in the benefit of collaborative writing in general.

Keywords: *collaborative writing, corpus linguistics, key word analysis, lexical bundles, second language writing*

INTRODUCTION

The contribution of collaboration to second language (L2) learning has been the focus of theoretical research from a variety of perspectives, including interactionist (e.g., Mackey, 2012; Gass, 2003) and sociocultural (e.g., Lantolf, 2011; Swain, 2010) approaches to L2 acquisition. The theoretical benefits of interaction have provided support for the use of pair and group work in several pedagogical approaches to L2 teaching, such as Communicative Language Teaching and Task-based Language Teaching. Until recently, however, the majority of this research has focused on oral communication (Mackey, 2007), even though researchers have argued that written communication may provide learners with greater opportunities to focus on language form and accuracy (Williams, 2012), due to its greater processing time as compared to speaking

(Kormos, 2012). Although peer interaction may occur at different phases of the writing process, such as brainstorming or peer review, collaborative writing involves a sustained pattern of interaction between writers and shared responsibility in the co-construction of texts through cooperation and interaction during the entire writing process (Storch, 2013), which differentiates it from cooperative writing in which students may divide up a writing task and subsequently combine individually-written sections into a single text. Although some collaborative writing studies have implemented using technology, such as wikis or blogs, our focus here is on collaborative writing in face-to-face classroom settings. Collaborative writing may benefit L2 learners by creating opportunities for them to engage in conversations about how language works, which Swain has described as "languaging" (2010), and by helping them create grammatically accurate texts.

Studies that have described the interaction that occurs between learners when they carry out collaborative writing tasks have focused largely on the occurrence of language-related episodes (LREs). Interactions with peers may generate LREs at a more developmentally appropriate level than instructor feedback (Brooks & Swain, 2009; Ohta, 2000), with studies showing that that L2 learners incorporate the forms they discuss into their written texts (e.g., Brooks & Swain 2009; Watanabe & Swain 2007). Besides engaging in LREs, L2 learners also discuss the content and organization of their written texts during collaborative writing tasks (McDonough, Crawford & De Vleeschauwer, 2016; Wigglesworth & Storch, 2009; Storch & Wigglesworth, 2007; Storch, 2005).

Collaborative writing research has also compared the linguistic features of collaborative and individual texts by using measures of complexity, accuracy, and fluency to identify potential variation in language use. These studies have found that whereas collaborative texts do not differ in their complexity or fluency, they are more accurate (Fernández-Dobao, 2012; Wigglesworth & Storch, 2009; Storch & Wigglesworth, 2007; Storch, 2005). One potential explanation for the lack of complexity findings concerns difficulty in defining and operationalizing the construct of complexity (Housen, Kuiken, & Vedder, 2012. Despite criticisms of the t-unit as being a broad measure that fails to distinguish among different grammatical forms (Biber et al., 2011) or capture variation in lower-proficiency levels (Norris & Ortega, 2009), it has been widely used for complexity measures in collaborative writing research.

Although the concept of complexity can be a valuable perspective on understanding language development, one can also consider the concept of variation as relevant to understanding L2 writing performance. Rather than relying on constructs such as t-units or dependent/independent clauses to reveal potential language differences between individual and collaborative texts, a descriptive lexico-grammatical approach may provide insight into potential variation in L2 performance. Drawing upon corpus linguistics' well-

established tradition of describing linguistic variation (Reppen, Fitzmaurice, & Biber, 2002; Biber, 1988), this type of analysis aims to identify variation in texts that can be attributed to situational variables such as participants, mode, production circumstances, setting, purpose, and topic (Biber & Conrad, 2009). For example, register analysis (Biber, 1988) has shown that the situational differences between speech and writing result in the use of different lexico-grammatical features, with speech characterized by contractions, first and second person pronouns, and indefinite pronouns. The occurrence of these features in spoken language can be attributed to the fact that spoken language occurs in a shared physical context, which requires reference to the participants in the speech event. Lexico-grammatical descriptive analysis has been used to characterize English grammar across multiple registers (Biber et al., 1999), including the academic register (Biber, 2009; Biber et al., 2004). These studies have used numerous methods for identifying differences, including corpus-informed approaches (i.e., using pre-existing lexico-grammatical features as the basis for analysis) and corpus-driven approaches (i.e., using computer programs to identify potential lexico-grammatical features that emerge from the text themselves). The present study adopts a corpus-driven approach involving two different analyses: keywords and lexical bundles.

Keyword analysis has a long history in corpus research (Bondi & Scott, 2010; Scott & Tribble, 2006) where it has been used to compare the lexical features associated with two different sets of texts. The procedure is fairly straightforward as it involves compiling a word list from each set of texts and then using a statistical analysis to identify words that are more frequent (positive keywords) or less (negative keywords) frequent across the two lists. Keyword analysis has been used to describe various genres (Tribble, 2000), variation within genres (Baker, 2004), discourse differences in the field of critical discourse analysis (Muldering, 2008) and educational texts (Leone, 2010). As noted by Baker (2004), keyword analysis provides researchers with words that distinguish two different sets of texts, but the researcher must decide how to analyze and interpret any potential difference in relation to a specific research issue. Given the relative ease with which potentially relevant data can be retrieved, this approach is promising for collaborative writing researchers.

The second corpus-driven analysis used in this study was lexical bundles, which involves the identification of recurrent word strings. Compared to other types of multi-word units that are identified using semantic criteria, such as idiomatic expressions like *kick the bucket*, lexical bundles are identified purely in terms of recurrent word strings (Granger & Paquot, 2008). Unlike n-grams (i.e., “n” refers to the number of words and “gram” denotes word), lexical bundles have frequency and dispersion criteria and have identifiable functions associated with them (Cortes, 2015). Corpus research carried out by Biber and colleagues has characterized different registers in terms of the structure and function of lexical bundles (Biber & Barbieri, 2007; Biber, 2006; Biber et al., 2004; Biber et al., 1999).

Whereas academic writing contains more three and four-word bundles based on noun and prepositional phrases that serve referential functions, academic spoken language contains more bundles based on verb phrases that serve discourse organizing functions. This analytic framework also has been used to explore variation in the texts written by novice and experienced writers, with novice writers using fewer bundles (Cortes, 2004) that serve different functions (Cortes, 2002) than experienced writers. Furthermore, novice writers tend to use more verb phrase and discourse organizing bundles than authors of published academic journals, with rates for L2 novice writers higher than L1 novice writers (Chen & Baker, 2010). Taken as a whole, this body of research provides evidence that the lexico-grammatical approach captures language variation in texts written in different registers and by different authors.

Situated within this approach, the current study compares the lexico-grammatical features of individual and collaborative texts composed by English L2 writers. The research question was: *Do collaborative and individual texts differ in terms of their lexico-grammatical features?* Although previous comparative studies have not identified systematic variation in the language forms or functions found in collaborative and individual texts, their reliance on traditional complexity measures may account for the absence of findings; therefore, due to the lack of findings in previous comparative studies, no predictions were made about whether individual and collaborative texts would show lexico-grammatical variation.

METHODOLOGY

Participants and instructional context

The participants were 99 second-year undergraduate students (53 women and 46 men) enrolled in the Faculty of Medicine at a public university in northern Thailand. They ranged in age from 17 to 20 years, with a mean age of 18.7 years (*SD* = .6). They had studied English previously from nine to 18 years, with a mean of 13.5 years (SD = 1.9) of prior English instruction in elementary and secondary schools. The majority of the students (65/99) had never traveled to a country that required English for communication. Of those that did, only 11 students reported visits of 30 days or longer. Only 24 of the 99 students reported any knowledge of additional languages besides Thai and English, which were Chinese (13), French (3), Spanish (3), along with Danish, German, Italian, Japanese, and Korean (1 each). The participants had never taken a standardized tests or in-house proficiency exam. However, based on their written texts, the majority of the participants represented basic users (A2) with a few proficient users (B1) according to the *Common European Framework of Reference for Languages*.

The participants were enrolled in two sections of a required EFL critical reading and writing class taught by the same instructor. Each class met for two, 75-minute class periods per week for 18 weeks, with breaks for holidays and examinations. The class was designed to develop students' critical thinking and reading abilities and foster paragraph-level writing skills, and used a textbook created by faculty members in the English Department of the university. The textbook was designed for A2 level students, and contained reading passages in the form of jokes, headlines, advertisements, and news reports. Most of the reading passages were adapted from *Reader's Digest,* and had Flesch-Kincaid grade levels ranging from 1.1 to 8.4. The writing component targeted sentence-level writing in the first two chapters, and then emphasized short (80-120 word) paragraph writing in the remaining chapters. Four paragraph types were taught, which were summary, compare/contrast, cause/effect, and problem/solution.

Materials and procedure

As part of a larger dataset (McDonough, Crawford & De Vleeschauwer, 2016), this study compared the lexico-grammatical features of collaborative and individual problem/solution paragraphs written by students in their EFL class. The assignment was administered as part of the normal instructional routine of the participants' EFL class. The prompt for writing tasks stated a problem (games addiction among CMU students) and instructed the students to write a short paragraph (80-100 words) in which they described possible solutions to that problem. At the time the paragraphs were written, the students had already been taught about problem/solution paragraphs using the information and practice tasks in their textbooks but the focus of instruction was not on the descriptive/narrative /personal writing in their previous courses, but was intended to introduce students to writing tasks that were more academic in nature. Whereas approximately half of the texts were written by individual students (35 texts), the other texts were co-constructed by two students working together (32 texts written by 64 individuals working in pairs).

The collaborative and individual writing tasks were administered by the classroom instructor following the procedures used for in-class writing assignments. For the class that wrote collaboratively, the students self-selected a partner, were given the task instructions, and wrote their collaborative paragraphs by hand. The students were allowed to talk to their partner, but discussion across pairs was not allowed. Students were free to choose which language(s) to speak when interacting with each other. In this context, students typically choose to work with a friend with whom they feel comfortable interacting. The second researcher was present during data collection and did not observe any learners within a pair who failed to engage in the collaborative task, such as by allowing a partner to do all the work. In contrast, for the class that wrote individually, students were not allowed to discuss anything with their peers. They received the prompt and wrote their paragraphs by hand. Students

in both conditions were permitted to consult a paper-based dictionary while writing their paragraphs, but no electronic dictionaries or other electronic devices (phones, tablets, or laptops) were allowed. Their performance was monitored to ensure that no additional resources were being consulted. Across both writing conditions, the students wrote for a mean of 55 minutes (*SD* = 12.5), with writing time ranging from 25 to 78 minutes.

Data coding & analysis

To facilitate data exchange among the researchers, the students' hand-written problem/solution paragraphs were converted into Microsoft Word documents by three research assistants. No changes were made to the formatting, spelling, punctuation, or language use in the students' texts during the conversion process, with the exception of ignoring words and sentences that the students had crossed out in their texts. As shown in Table 1, two sub-corpora were created for the collaborative and individual problem/solution texts. Although there were fewer collaborative texts (*n* = 32), they contained almost the same number of words as the individual texts (*n* = 35). The collaborative texts were approximately 18 words longer that the individual texts.

Table 1
Sub-corpus characteristics

Text type	Total texts	Total words	Mean words per text (*SD*)
Collaborative	32	5,412	169.1 (47.8)
Individual	35	5,437	151.2 (31.7)

Two software programs were used for the analysis. For the keyword analysis, word lists were generated and compared using WordSmith version 6.0 (Scott, 2012). To ensure a representative distribution, only words that occurred in at least 10% of the texts were analyzed. Log likelihood was used to determine keywords differences in the texts and a *p* value of .000001 was set for significance. To identify four-word lexical bundles, the software program AntConc 3.4.3 (Anthony, 2015) was used, with a lexical bundle defined as a continuous sequence of four words that are frequent in texts and have identifiable functions associated with them. Only four-word lexical bundles were included in the initial analysis because they are long enough to capture potential lexico-grammatical relations, but not so long that they occur infrequently (Cortes, 2015; Chen & Baker, 2010). However, as reported in the Results section, longer bundles were included after it was determined that some four-word sequences were in fact components within the same, longer sequence. To ensure a representative distribution, only bundles that occurred at least once in 10% of the texts in each sub-corpus were included in the analysis. Because the two groups of texts were nearly the same size, counts were not normalized. The bundles were classified in terms of their structural composition and discourse functions following Biber's structural and functional framework (Biber &

Barbieri, 2007; Biber, 2006; Biber et al., 2004). From a structural perspective, although four word lexical bundles do not form syntactic constituents, they contain fragments of grammatical units such as verb phrases (*is going to be*), noun phrases (*a little bit about*), and prepositional phrases (*of the things that*). In terms of their discourse functions, lexical bundles can fulfill referential, discourse-organizing, and stance functions. However, no stance bundles that met the frequency and dispersion criteria were found in the current dataset. Whereas referential bundles serve to stipulate some characteristic/object or provide an orientation point in the text, discourse-organizing bundles contribute to the overall discourse coherence of a text. Examples of the referential and discourse-organizing lexical bundles that met the criteria are provided in Table 2.

Table 2
Lexical bundles by discourse function and form

Type	Definition	Examples
Referential	Used to specify an attribute/entity or provide time, place, amount, or text reference	*problem of games addiction* [NP] *addiction among CMU students* [NP]
Discourse organizing	Used to structure a text such as introduce, elaborate or focus on a topic/argument	*to solve this problem* [VP] *there are a lot* [VP]

RESULTS

The research question asked whether collaborative and individual texts differed in terms of their lexico-grammatical features, which were operationalized in terms of key words and lexical bundles.

Key words distinguishing individual and collaborative texts

First, for the keyword analysis using WordSmith, five words with positive key values were identified, *you, your, it, we* and *firstly,* which means that these words were associated with the individual texts. In addition, four words with negative values were identified, *among, the, university,* and *students*, which means that these words were more frequent in the collaborative texts. Table 3 provides the list of all keywords that distinguished the individual and collaborative texts, along with their frequency, the number of texts, the keyness values, and the p values.

Table 3
Key words distinguishing individual and collaborative texts

Key word	Individual texts		Collaborative texts		Keyness value	p value
	Frequency	# Texts	Frequency	# Texts		
you	120	22	11	5	106.56	0.01
your	45	16	3	1	44.29	0.01
it	65	21	14	12	35.94	0.01
we	46	19	10	7	25.25	0.01
firstly	17	15	0	0	23.65	0.01
among	5	5	35	26	-25.17	0.01
the	145	32	251	31	-28.24	0.01
university	7	5	45	18	-30.83	0.01
students	58	22	171	32	-57.76	0.01

Initial lexical bundles

For the second lexico-grammatical feature, the initial search for 4-word bundles identified 13 lexical bundles that met the distribution requirements. Seven of the bundles were found in the collaborative texts, while six occurred in the individual texts. The complete list of bundles by text type along with the range of (i.e., number of different) texts, form and function is provided in Table 4.

Table 4
Initial lexical bundles by text type and frequency

Collaborative (*n* = 32)				**Individual (*n* = 35)**			
Bundle	**Frequency (# texts)**	**Form**	**Function**	**Bundle**	**Frequency (# texts)**	**Form**	**Function**
addiction among CMU students	25 (20)	NP	R	problem of games addiction	14 (8)	NP	R
games addiction among CMU	19 (19)	NP	R	the problem of games	10 (6)	NP	R
to solve this problem	17 (14)	VP	DO	to solve this problem	9 (7)	VP	DO
there are many solution	9 (8)	VP	DO	are a lot of	6 (5)	VP	DO
addiction among CMU students	6 (5)	NP	R	there are a lot	6 (5)	VP	DO
among	6 (5)	NP	R	many	5 (5)	Clausal	DO

CMU students problem			CMU students face
game addiction among CMU	6 (5)	NP	R

Note: DO = Discourse organizing, R = Referential

As seen in the list in Table 4, there was some overlap in the four-word lexical bundles, which needed to be addressed to avoid inflating frequency and function counts (Chen & Baker, 2010). In cases where two bundles were identical except for variation in the singular or plural form of a noun (e.g., *game* versus *games*), they were considered instances of the same bundle. Similarly, bundles that were identical except for variation in a function word (*there are many solutions* and *there are some solutions*) were considered instances of a single bundle with a variable slot (Biber, 2009; Hunston & Francis, 2000). To address complete overlap, which occurs when a four word bundle is part of a larger unit, longer bundles consisting of five or six words were identified. For example, the four-word lexical bundle *games addiction among CMU* occurred with equal frequency with the five word string g*ames addiction among CMU students*, which reflected the fact that the four-word bundle occurred exclusively within the five-word sequence. In such cases of subsumption (i.e., when one bundle is included within another bundle), only the longer bundle was included. In other words, if any initial four-word bundle was subsequently identified as being part of longer unit, only the longer five- or six-word bundle was included.

Final lexical bundles

After addressing overlap and subsumption, the total number of lexical bundles in the data set was reduced to eight, with four in the collaborative texts and four in the individual texts. The final list of lexical bundles were then categorized for both form and function In one case, the function of an initial bundle was changed once it was identified as being a part of a longer bundle. Specifically, the initial bundle *many CMU students face* was classified as a VP-based, but it occurred within a longer bundle that started with the word *problem* and included an "optional" relativizer (*that*). As a result the final lexical bundle was classified as clausal. Table 5 shows the frequency, form, and function of the eight bundles that were identified. In terms of overall frequency, lexical bundles were more frequent in the collaborative texts (60 bundles) than the individual texts (34 bundles). In addition, the lexical bundles were dispersed across a larger number of collaborative texts with three bundles occurring in at least ten texts. In contrast, no lexical bundles occurred in more than eight individual texts.

Table 5
Final lexical bundles by text type: Frequency, Form, and Function

Collaborative (*n* = 32)				**Individual (*n* = 35)**			
Bundle	**Frequency (# texts)**	**Form**	**Function**	**Bundle**	**Frequency (# texts)**	**Form**	**Function**
game(s) addiction among CMU students	25 (20)	NP	R	the/a problem of game(s) addiction	14 (8)	NP	R
to solve this problem	17 (14)	VP	DO	to solve this problem	9 (7)	VP	R
there are many/some solution(s)	11 (10)	VP	DO	there are a lot of	6 (5)	VP	DO
game addiction among CMU students problem	7 (6)	NP	R	problem (that) many CMU students face	5 (5)	Clausal	DO
Total	**60**			**Total**	**34**		

Note: DO = Discourse organizing, R = Referential

Finally, in terms of the form and function of the bundles, also shown in Table 5, the collaborative texts contained 32 NP-based referential bundles and 28 VP-based discourse organizing bundles. In contrast, the individual texts contained 14 NP-based referential bundles and 20 VP-/Clausal bundles fulfilling discourse organizing functions. A closer examination of the structural configurations of the bundles showed some similarity in the collaborative and individual texts (*to solve this problem* or infinitive + det + noun) as well as some the shared noun sequences that are related to the writing prompt (*game(s) addiction* and *CMU students*), but the collaborative texts had more NP-based sequences. The collaborative texts contained noun sequences on both the head NP position (*game(s) addiction*) as well as in the position of the prepositional object (*CMU students* and *CMU students problem*).

DISCUSSION

To summarize the findings, both the keyword and lexical bundle analyses revealed differences in the lexico-grammatical features of collaborative and individual texts. The keyword analysis showed that individual texts contained more pronouns (*you, your, it, we*) and use of *firstly*. In contrast, collaborative texts contained more nouns (*university, students*) as well as the determiner *the* and the preposition *among*. The lexical bundles analysis revealed that

collaborative texts contained nearly twice the number of bundles than individual texts. Although there were no clear functional differences in the lexical bundles in individual and collaborative texts, there was a notable difference in form; namely, the collaborative texts had more noun sequences with variable slots. In this section, we highlight how the findings may shed light on the benefits of collaborative writing.

The keyword analysis showed that the majority of words distinguishing the individual texts from the collaborative texts were words associated with interactional discourse as opposed to informational discourse (see Biber, et. al., 1999 and Biber 1988). The interactional features found in the individual texts may be the result of an absence of collaboration during the individual writing process, which led the individual writers to "interact" with the text and use more interactional features. Because this was the first EFL class where the students were required to address academic writing tasks as opposed to personal, descriptive, and narrative writing, the presence of a partner may have helped them to employ features more representative of academic writing. As seen in the individual text provided in Example 1, second person pronouns are used to address the reader or general audience throughout the text (*when you get addicted; your bod[y]; your time; you can find new friends; you just manage your time; destroy you*). The third person pronoun *it* is also prevalent to refer to a variety of previously established referents (e.g., technology, games addiction) and some potentially confusing referents (*it has it's own solution*). Additional interactive features associated with the individual texts included questions, as in the example below, which points toward a tendency for the individual writers to align directly with the reader/audience in ways that were more reflective of spoken, interactional discourse than with (academic) writing.

Example 1: Interactivity in individual texts

> *Nowadays, the game technology is developing so quick.* ***It*** *make a lot of students love and interesting to play games but some of them addict* ***it.*** *When* ***you*** *get addicted with the game* ***it*** *cause many problem for* ***your*** *life such as* ***your*** *bod and socially will be weak,* ***it*** *wastes* ***your*** *time and etc. So* ***we*** *need to solve this problem may be* ***we*** *should think about* ***it*** *seriously. How to get /---/ form games addiction? I think they should find some activities to do instead playing game.* ***It's*** *such a good idea because some activity can make* ***your*** *body stronger and your feeling more better. Sometime* ***you*** *can find new friends form activities to. How about wasting of time?* ***It's*** *very serious problem for student who's playing games. However* ***it*** *has* ***it****s own solution.* ***You*** *just manage your time and make* ***it*** *well. So playing game is good but don't let them destroy* ***you****.*

Unlike the individual writers, students who wrote collaboratively had opportunities to interact with their peers throughout the writing process. This

interaction may have helped these L2 writers use forms that are more indicative of informational discourse, as opposed to the features of interactional discourse found in the individual texts. The keywords that distinguished the collaborative texts involved full noun phrases (*the university, the students*) as opposed to the pronouns found in the individual texts. As illustrated in Example 2, the collaborative texts contained determiner + noun sequences using keywords identified in the keyword analysis (*students* and *university*) as well as other determiner + noun sequences (e.g., *the (first) solution, the ways, the government*).

Example 2: Collaborative text with noun phrases

> *There are many solutions for this problem. The first solution is about to manage* ***the students'*** *free time. In* ***the university****,* ***the students*** *have to study hard. This is the cause of stress, so they have to relax themselves in their free time. Playing games is one of the ways to relax, but some* ***students*** *spend all time to play games. This is the cause of game addiction. Therefore to stop this problem,* ***the students*** *should do other activities for relaxing themselves in their free time such as do the exercise, listen to music or etc. Another solution is about the government. There are many game centers around* ***the university****. Thus, it is easy to find the places for playing games. The government should not allow the game centers to locate near* ***the university****. Therefore,* ***the students*** *should control themselves and the government should help them to stop this problem together.*

The collaborative writers' tendency to use full noun phrases was also evident by the functions associated with pronoun use. Whereas individual texts contained 120 occurrences of the pronoun *you*, most of which occurred in subject position (109), the collaborative texts contained only 11 instances of *you* in subject position. A similar pattern was found with the pronoun *it*, occurring as a subject 59 times in the individual texts but only 12 times in the collaborative texts.

In addition to keyword variation, the comparison of collaborative and individual texts also revealed variation in lexical bundle use. Collaborative texts contained nearly twice as many lexical bundles (60 vs. 34) as the individual texts, and those bundles occurred in a greater number of texts. For example, the most frequent bundle in the collaborative texts (*game(s) addiction among CMU students*) was found in 20 of the 32 texts (63%). However, the most frequent bundle in the individual texts (*problem of game(s) addiction*) occurred in only 8 of the 35 texts (23%). Interestingly, the prompt for the task specifically mentioned "games addiction among CMU students" which suggests that the collaborative writers may have employed a strategy of using information from the prompts during a timed-writing task. A similar pattern of bundle frequency was also found in the collaborative texts with *to solve this problem* and *there are many/some solution(s)*), occurring in 14 (44%) and 10 (31%) of the collaborative texts,

respectively. These bundles tended to occur toward the beginning or the end of the texts. The bundle *game(s) addiction among CMU students* occurred in the collaborative texts at the beginning of the text as a way to state the problem before providing a solution in almost all cases (19/20 times). However, in the 20 collaborative texts in which this bundle appears, five of them used this bundle at both the beginning and the end of the text as a way of framing the paragraph. This use of lexical bundles in the collaborative texts is illustrated in Example 3.

Example 3: Collaborative text with lexical bundles as framing devices

> ***There are many solution*** *to* ***games addiction among CMU students problem****. First, prevent student from accessing the games. The university should be responsible this solution by blocking online games that use internet. This methot block only games but can access normal website. Blocking online games causes more difficulty to playing game. Students who are very addict to games may find the game shop outside the university. lead to second solutions. Second, do something to change student's interest from games. Parents should take care student when student is at home. For example, doing activity with their child, spend time with their child and don't leave their child with game for a long time otherwise children who stay with games will be interested and addict to the games. Students should be responsible theirselves. Studying is main duty of students, not playing game. Student should study, do their work or anything else to improve quality of life. Students should play games only for relax. In addition, playing sports, reading books should also be done for relax. In conclusion, the ways to solve* ***Game addiction among CMU students*** *are online blocking system by university and good-take care by their parents. These are effective way* ***to solve this problem****. However, everything relies on students who will choose their own way.*

In sum, the lexical bundle analysis suggests that having students write collaboratively provided them with opportunities to interact with each other and the topic content during the writing process, which, in turn, allowed them to use lexical bundles more often and in different ways than the individual writers. Such lexical differences are noted in the use of noun-noun sequences in the bundles such as *game(s) addiction, CMU students*, and *CMU students problem*, and greater variety in word forms (*solve* and *solution*). It should be noted that we are not suggesting here that more lexical bundles leads to better writing; instead, we are suggesting that the collaborative condition permitted the students to use different lexico-grammatical features indicative of academic writing, including more noun-noun sequences.

IMPLICATIONS

Although previous comparative studies failed to identify differences in the grammatical features of individual and collaborative texts, this could be due to their focus on measures of complexity, such as t-unit and clausal-based measures (Fernàndez-Dobao, 2012; Wigglesworth & Storch, 2009; Storch & Wigglesworth, 2007; Storch, 2005). The current study employed bottom-up corpus–driven methods to compare the lexico-grammatical features of collaborative and individual texts, which identified differences in the occurrence of keywords and lexical bundles across the text types. The keyword analysis showed that individual writers tended to use lexical features associated with in interactive spoken discourse, as opposed to informational discourse or academic writing. In addition, Biber and colleagues (Biber et al., 2011) have argued that noun phrase length, which is affected by the use of prenominal and post-nominal modification, is an important characteristic of academic writing. Subsequent research has shown that texts written by more experienced L2 writers (Parkinson & Musgrave, 2013) and texts that received higher ratings (Taguchi et al., 2013) also contain longer noun phrases. Similarly, the collaborative texts analyzed here contain a higher frequency of lexical bundles based on noun and prepositional phrases. From a methodological perspective, it may be useful to incorporate lexico-grammatical features into future comparative studies of collaborative and individual texts in order to clarify their potential differences.

The findings raise potential pedagogical implications for the use of collaborative writing tasks with novice L2 writers. The collaborative texts analyzed here contained fewer features of interactive discourse and more noun-based lexical bundles. These findings suggest that collaborative writing tasks may 'push' learners to deploy their linguistic resources in ways that help them approximate features of academic writing. Consequently, it may be beneficial to include collaborative writing tasks in EFL classes as a way to help learners acquire academic writing skills. An important issue, however, is whether carrying out collaborative writing tasks facilitates individual writing development. Ideally, students will improve their individual writing abilities, and studies to date have not clarified under what conditions collaborative writing tasks contribute to individual writing development. In order for collaborative writing to positively impact individual writing development, activities that raise learners' awareness of what transpired during the collaborative writing activity may be useful. Having students write short reflective essays on the collaborative writing experience or compare individual and collaborative texts may lead to greater awareness of how writing collaboratively positively impacted their language use and perhaps translate more readily into individual writing ability.

Integrating collaborative writing tasks into instructional sequences that include explicit instruction about the rhetorical structure of academic English may also

help facilitate individual writing development. For example, previous studies of individual writers have found that explicit instruction on the rhetorical structure of source-based written texts helped EFL writers acquire discourse features associated with academic writing (McDonough, Crawford & De Vleeschauwer, 2014; Zhang, 2013). By targeting lexical bundles that fulfill referential functions, instructors may help writers develop the ability to use those structures when writing individually, as opposed to using them only during collaborative writing tasks. For example, collaborative writers used referential bundles to specify the problem being addressed and maintain reference to that problem throughout their texts, which individual writers did not. By focusing on these structures and drawing learners' attention to their form and function, instructors may help facilitate transfer to individual writing tasks.

LIMITATIONS

Although the findings revealed lexico-grammatical variation in the collaborative and individual texts written by these Thai EFL learners, there are several limitations that impact the study's generalizability. First, the sub-corpora were fairly small, with each one only approximately 5,400 words. Previous studies considering key words and lexical bundles in corpus linguistics research have used much larger corpora (for lexical bundle studies, generally over one million words, although Chen & Baker, 2010 use a corpora of 120,000 words), and considerable debate remains about whether a minimum size is necessary to identify lexico-grammatical variation (Cortes, 2015). Next, as the current study examined only problem/solution paragraphs, it is not clear whether the findings would apply to other text types commonly used in EFL settings, such as cause and effect, summary, narrative, and argumentative texts. It is possible that lexico-grammatical variation between individual and collaborative texts is mediated by text type, and future research involving more diverse texts is needed to clarify its potential role in variation.

Finally, although both the individual and collaborative tasks were administered under similar conditions, experimental research under more tightly controlled conditions, rather than classroom based research, can help clarify the extent to which time on task may influence students' use of lexico-grammatical features while writing both collaboratively and individually. Future research might also consider the relationship between the learners' collaborative dialogue and their texts. Although it was beyond the scope of the current study, analyzing the learners' interaction while carrying out the collaborative writing task may help shed light on how collaboration allows students to access and deploy linguistic forms that they may not use or use less frequently during individual writing tasks.

CONCLUSION

The findings revealed lexico-grammatical differences between individual and collaborative texts involving keywords and lexical bundles. By adopting the lexico-grammatical approach associated with corpus linguistics, the study identified variation in individual and collaborative texts that has gone undetected in previous comparative research that adopted clausal measures of linguistic complexity. Broadening the methodological scope of comparative studies by adopting a textual variation approach has opened the door for further investigations of collaborative writing tasks and raised questions about how the process of collaboration can facilitate individual writing development. Our future research aims to address these issues in order to provide greater insight into the positive impact of collaboration on EFL writers' development.

ACKNOWLEDGEMENTS

Funding for this study was provided by the Canada Research Chairs program (950-221304). We would like to thank the instructor, Jindarat De Vleeschauwer, for her assistance with data collection. We are also grateful to the research assistants who assisted with data coding: Zach Alderton, César García Fuentes, and Maxime Lavallée.

THE AUTHORS

William J. Crawford is an Associate Professor in Applied linguistics/TESL in the Department of English at Northern Arizona University. His research interests include corpus linguistics, second language writing and English grammar.
bill.crawford@nau.edu

Kim McDonough is a Professor of Applied Linguistics and Canada Research Chair in the Department of Education at Concordia University. Her research interests include interaction and usage-based approaches to second language acquisition, structural priming, joint attention, peer interaction, and task-based language learning.
kim.mcdonough@concordia.ca

REFERENCES

Anthony, L. (2015). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Baker, M. (2004). Querying keywords: questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, *32*(4), 346-359.

Biber, D. (1988). *Variation across speech and writing.* New York, NY: Cambridge University Press.

Biber, D. (2006). *University Language: A corpus-based study to spoken and written registers* [Studies in Corpus Linguistics 23]. Amsterdam, The Netherlands: John Benjamins.

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics, 14*(3), 275–311.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, England: Longman.

Biber, D., Conrad, S. & Cortes, V. (2004). 'If you look at ...': Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.

Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263–286.

Biber, D. & Conrad, S. (2009). *Register, genre and style.* New York, NY: Cambridge University Press.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5–35.

Bondi, B. & Scott, M. (Eds.), 2010. *Keyness in texts*. Amsterdam, The Netherlands: John Benjamins.

Brooks, L. & Swain, M. (2009). Languaging in collaborative writing: Creation and response to expertise. In A. Mackey & C. Polio (Eds.) *Multiple perspectives on interaction in SLA* (pp. 58-89). Mahwah, NJ: Lawrence Erlbaum.

Chen, Y. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.

Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum Origins and functional analysis developments. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (pp. 197-216). Amsterdam, The Netherlands: John Benjamins.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4)*, 397–423.*

Cortes, V. (2002). *Lexical Bundles in Published and Student Academic Writing in History and Biology* (Ph.D. dissertation). Northern Arizona University.

Fernàndez-Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and
individual work. *Journal of Second Language Writing*, *21*, 40–58.

Gass, S. (2003). Input and interaction. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 224-255). Oxford, UK: Blackwell Publishers.

Granger, S. & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27-49). Amsterdam, The Netherlands: John Benjamins.

Housen, A. Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurements and research. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1-20). Amsterdam, The Netherlands: John Benjamins.

Hunston, S. & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English.* Amsterdam, The Netherlands: John Benjamins.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing, 21*(4), 390-403.

Lantolf, J. (2011). The sociocultural approach to second language acquisition: Sociocultural theory, second language acquisition, and artificial L2 development. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 24-47). London, UK: Routledge.

Leone, P. (2010). General spoken language and school language: Key words and discourse patterns in history textbooks. In B. Bondi & M. Scott (Eds.), *Keyness in texts* (pp. 238-248). Amsterdam, The Netherlands: John Benjamins.

Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 learning*. Oxford, England: Oxford University Press.

Mackey, A. (2007). Interaction as practice. In R. Dekeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 85-110). New York, NY: Cambridge University Press.

McDonough, K., Crawford, W. & De Vleeschauwer, J. (2014). Summary writing in a Thai EFL university context: Students' use of source texts. *Journal of Second Language Writing*, *34*, 20-32.

McDonough, K., Crawford, W. & De Vleeschauwer, J. (2016). Thai EFL learners' interaction during collaborative writing tasks and its relationship to text quality. In M. Sato & S. Ballinger (Eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda* (pp. 185-208). Amsterdam, The Netherlands: John Benjamins.

Muldering, J. (2008). Using keywords analysis in CDA: Evolving discourses of the knowledge economy in education. In B. Jessop, N. Fairclough, & R. Wodak (Eds.), *Education and the Knowledge-Based Economy in Europe* (pp. 149-170). Rotterdam: Sense Publishers.

Norris, J. & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578.

Ohta. A. (2000). Rethinking interaction in SLA: Developmentally appropriate assistance in the zone of proximal development and acquisition of L2 grammar. In J.P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 51-78). Oxford, England: Oxford University Press.

Parkinson, J. & Musgrave, J., (2013). Development of noun phrase complexity in the writing English for Academic Purposes students, *Journal of English for Academic Purposes*, *14*, 48-59.

Reppen, R., Fitzmaurice, S., & Biber. D. (2002). *Using corpora to explore linguistic variation.* Amsterdam, The Netherlands: John Benjamins.

Scott, M. (2012). WordSmith Tools version 6, Stroud: Lexical Analysis Software.

Scott, M. & Tribble, C. (2006). *Textual Patterns: Keyword and corpus analysis in language education*, Amsterdam, The Netherlands: Benjamins.

Storch, N. (2013). *Collaborative writing in L2 classrooms*. Buffalo, NY: Multilingual Matters.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing, 14*(3)*,* 153-173.

Storch, N. & Wigglesworth, G. (2007). Writing tasks: Comparing individual and collaborative writing. In M.P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 157-177). Buffalo, NY: Multilingual Matters.

Swain, (2010). Talking-it-through: Languaging as a source of learning. In R. Batstone (Ed.), *Sociocognitive perspectives on language use and language learning* (pp. 112-130). Oxford, UK: Oxford University Press.

Taguchi, N., Crawford, W., & Wetzell, D. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, *(47)*2, 420-430.

Tribble, C. (2000). Genres, keywords, teaching: Towards a pedagogic account of the language of project proposals. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 75-90) Frankfurt, Germany: Peter Lang.

Watanabe, Y. & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research, 11*(2), 121-142.

Wigglesworth, G. & Storch, N. (2009). Pairs versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing, 26*(3), 445-466.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing, 21*(4)*,* 321-333.

Zhang, C. (2013). Effect of instruction on ESL students' synthesis writing. *Journal of Second Language Writing, 22*(1), 51–67.