

Comparing the Lexical Features of EAP Students' Essays by Prompt and Rating

Maxime Lavallée & Kim McDonough

Previous research has shown that high frequency lexical items, such as AWL words and formulaic expressions, may differentiate between texts written by expert and novice writers (Chen & Baker, 2010; Hancioğlu, 2009), and that lexical features related to breadth, depth, and accessibility differentiate among texts from L2 writers of different proficiency levels (Crossley & McNamara, 2009, 2012; Crossley, Weston, McLain Sullivan, & McNamara, 2011). The current study compared the essays written by EAP students in response to either a cause or an effect writing prompt. As part of their EAP writing class, the students (N = 94) had two weeks to read six source texts and take notes to prepare for an integrative-writing exam. Students' essays were assessed by three raters using a holistic rubric, and five lexical features of their essays were analyzed: percentage of AWL word use, content word frequency, word familiarity, imaginability, and lexical diversity. The results indicated that responses to the effect prompt were rated significantly lower than cause essays, contained more frequent and familiar words, and had a lower percentage of AWL words. However, there was no significant correlation between essay ratings and lexical features. Potential explanations for the findings and pedagogical implications are discussed.

Des recherches antérieures ont révélé que les items lexicaux à haute fréquence, tels la liste des mots académiques et les formules rigides, peuvent varier selon que le texte soit écrit par un expert ou un débutant (Chen & Baker, 2010; Hancioğlu, 2009), et que les éléments lexicaux liés à l'envergure, la profondeur et l'accessibilité varient dans les textes écrits par des auteurs L2 de compétences différentes (Crossley & McNamara, 2009, 2012; Crossley, Weston, McLain Sullivan, & McNamara, 2011). La présente étude a comparé des rédactions écrites par des élèves d'anglais académique où la tâche d'écriture évoquait des causes ou des effets. Le cours d'anglais académique exigeait que les élèves (N=94) lisent, en deux semaines, six textes originaux et qu'ils prennent des notes pour se préparer à un examen écrit intégratif. Trois évaluateurs se sont appuyés sur une rubrique globale pour analyser cinq éléments lexicaux : pourcentage de l'utilisation des mots de la liste des mots académiques, fréquence des mots lexicaux, capacité à évoquer des images mentales et diversité lexicale. Les résultats démontrent que les évaluations des rédactions basées sur la tâche d'écriture évoquant des effets étaient nettement inférieures aux évaluations des rédactions basées sur la tâche d'écriture évoquant des causes. Ces premières comptaient également plus de mots familiers et répandus, et moins de mots figurant dans la liste des mots académiques. Les auteurs présentent des hypothèses pour expliquer ces résultats, ainsi que les implications pédagogiques de leur recherche.

It is widely acknowledged that acquiring lexical proficiency in a second language (L2) is a difficult process, especially for students who need to acquire the vocabulary knowledge necessary for academic reading and writing tasks (Chen & Ge, 2007; Cobb & Horst, 2004; Nation, 2001; Simpson-Vlach & Ellis, 2010). Although English for academic purposes (EAP) students tend to be familiar with high-frequency words of “general” English and the technical vocabulary of their own disciplines, they are often less familiar with cross-disciplinary academic lexical features—for instance, the frequent vocabulary items uncommon in other types of writing, such as fiction (Coxhead, 2000; Coxhead & Nation, 2001; Xue & Nation, 1984), and other aspects of lexical proficiency, like word breadth, depth, and accessibility (Meara, 2005). Lexical proficiency plays a particularly important role in academic writing because EAP students are typically evaluated based on how well they express their comprehension and interpretation of a subject through a written text (Douglas, 2013; Nation, 2008; Roessingh, 2006).

Due to the key role of lexical proficiency in academic writing, an important question is how to conceptualize it. One approach is to identify the word families or expressions that frequently occur in academic texts across disciplines, such as the *University Word List* (Xue & Nation, 1984), the *Academic Word List* (Coxhead, 2000), and the *Academic Formulas List* (Simpson-Vlach & Ellis, 2010). For example, Coxhead’s Academic Word List (AWL) consists of 570 word families that account for 10% of the vocabulary used in academic texts. Numerous studies have investigated the occurrence of AWL words in academic texts from specific disciplines ranging from applied linguistics to medicine, and have found that AWL words account for approximately 7–12% of the total words (see Coxhead, 2011, for a summary). Although the AWL has been criticized due to the uneven distribution of its words across disciplines (Hyland & Tse, 2009) and the imprecise boundaries between general and academic words (Hancioğlu, Neufeld, & Eldridge, 2008), researchers have suggested that knowledge of AWL vocabulary is a core component of the lexical proficiency necessary to read and write academic texts (Cobb & Horst, 2004; Vongpumivitch, Huang, & Chang, 2009).

Although much research has been done to determine how frequently AWL words occur in a variety of academic texts, fewer studies have investigated their occurrence in EAP student writing or explored whether AWL use is associated with text quality. Research with L1 English students who passed a university test of entrance-level writing competence (Douglas, 2013) revealed 6.7% use of AWL words in their essays. A comparative study of the abstracts written by students at universities in countries where English is spoken as a first language and English L2 writers at a university in Cyprus (Hancioğlu, 2009) found that the L2 abstracts had a lower rate of AWL use (12.95%) and used fewer word families (388) than the abstracts written by students in English-speaking countries (14.76% and 564 word families). Taken together, these studies suggest that AWL use may help differentiate between texts composed

by writers with varying levels of experience with academic writing. Just as formulaic language use has been shown to capture differences in writers' background and text quality (Chen & Baker, 2010), AWL use may also be associated with variation in the quality of texts written by EAP students.

In addition to the conceptualization of lexical proficiency as use of frequent academic words and phrases, a second approach considers measures of vocabulary size, depth of knowledge, and accessibility of core lexical items (Meara, 2005). Vocabulary size captures how many words a learner knows (e.g., lexical diversity measures), while depth of knowledge reflects how well a word is known (e.g., semantic relatedness or word associations). Accessibility refers to how quickly words can be retrieved or processed, and is based on judgements of word concreteness and familiarity. Using computational tools that generate lexical profiles, such as Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004), researchers have identified which lexical features differentiate texts written by L1 writers at varying grade levels (Crossley, Weston, McLain Sullivan, & McNamara, 2011; McNamara, Crossley, & McCarthy, 2010) and texts written by L2 writers from diverse proficiency levels (Crossley & McNamara, 2012; Crossley, Salsbury, & McNamara, 2011; Crossley, Salsbury, McNamara, & Jarvis, 2010). The L2 studies showed that lexical diversity and word frequency, both measures of vocabulary size, and word hypernymy (a measure of breadth) had strong relationships with ratings of lexical proficiency, while lexical diversity, word familiarity, frequency, meaningfulness, and imagability (i.e., ease of constructing a mental image) had strong relationships with ratings of text quality. In essence, L2 writers at higher proficiency levels use more imagable, unfamiliar, and infrequent words and show greater lexical diversity than lower proficiency writers (Crossley, Salsbury, et al., 2011).

In addition to language background and proficiency, researchers have also explored whether other attributes help account for variation in the lexical features of texts composed by L2 writers. Working from an assessment framework, researchers have explored the relationships among prompt characteristics, essay ratings, and linguistic features of L2 texts. Cho, Rijmen, and Novák (2013) found that there was a relationship between ratings of integrated-writing essays from the TOEFL iBT (Internet-based Test) and characteristics of the reading and lecture prompt materials. Similarly, researchers have shown that lower proficiency L2 writers may struggle to understand the source materials provided for integrated-writing exams and may not be able to select the appropriate or relevant information to use in their essays, ultimately affecting the overall quality of their texts (Gebriel & Plakans, 2013; Sawaki, Quinlan, & Lee, 2013; Wolfersberger, 2013). Whereas some integrated-writing research has shown a positive relationship between lexical sophistication (type/token ratio) and text ratings (Cumming et al., 2005), other studies have found that average word length (another measure of lexical sophistication) was not associated with essay ratings (Gebriel & Plakans, 2013; Sawaki et al., 2013).

To summarize, previous research has shown that high frequency lexical items, such as AWL words and formulaic expressions, are characteristic of academic texts across disciplines (Coxhead, 2011; Simpson-Vlach & Ellis, 2010), but that novice L1 writers and L2 writers may use such words and expressions at lower rates than more expert academic writers (Chen & Baker, 2010; Hancioğlu, 2009). Furthermore, lexical features related to breadth, depth, and accessibility also differentiate among L2 texts from writers of different proficiency levels (Crossley & McNamara, 2009, 2012; Crossley, Salsbury, et al., 2011). However, these studies did not analyze integrated-writing tasks, but instead used independent writing tasks such as free writing and narratives. Although integrated-writing assessment studies have demonstrated that lexical features differentiate essays at rating levels (Cumming et al., 2005), other research studies have reported no difference in lexical diversity features based on essay rating (Gebriel & Plakans, 2013). Furthermore, studies of prompt characteristics have shown that prompt variation in terms of the difficulty of source texts impacts the text ratings (Cho et al., 2013). However, the assessment studies were carried out under test conditions in which L2 writers had limited time to read and comprehend the sources before writing.

The current study examines the relationships among the lexical features of EAP students' texts, essay ratings, and writing prompts, but focuses on an integrated-writing test in which students had time to read and prepare source texts prior to writing their essays. To investigate these relationships, two research questions were addressed. The first question was "Does the writing prompt affect the text quality and lexical features of EAP students' essays?" Due to the previous research that demonstrated variability in EAP student writing based on prompt characteristics (Cho et al., 2013), we predicted that there would be differences in the ratings and lexical features of essays written in response to different prompts. The second research question was "Is there a relationship between the text quality and lexical features of EAP students' essays?" Based on the previous research that demonstrated that lexical features were predictive of general proficiency levels (Crossley, Weston, et al., 2011) and differentiated among essay ratings (Cumming et al., 2005), we expected that lexical features associated with academic writing would have a positive relationship with essay ratings.

Method

Participants

The participants were 94 students (51 women, 43 men) taking an EAP writing class at an English-medium university in Montreal, Canada. Their mean age was 22.1 years ($SD = 2.7$), and they came from a variety of first language backgrounds, which included Chinese (55), French (18), Arabic (11), Farsi (3), Ukrainian (2), Bengali (1), Spanish (1), Romanian (1), Russian (1), and Vietnamese (1). They were enrolled in undergraduate degree programs in the

Faculties of Business (46), Arts and Sciences (24), Engineering (23), and Fine Arts (1). They had already partially met the university's English proficiency requirement for admission by obtaining a TOEFL iBT score between 75 and 89 or an IELTS score of 6.0 to 6.5 (or equivalent), but were required to take the EAP writing class based on their performance on the university English placement test. The participants were recruited from 8 of the 14 EAP writing classes offered during the semester of data collection, which were taught by different instructors using the same course outline, textbook, and exams.

Materials and procedure

The students were required to complete three integrated-writing exams as part of their 13-week EAP class. The exams were written as part of the required assessment component of the class, following the procedures devised and implemented by the EAP program. The text types targeted in the three integrated-writing exams were summaries (week 4), cause and effect essays (week 9), and argumentative essays (week 13). The current study focused on the cause and effect essays, which had the most comparable distribution of essays written in response to different prompts. In other words, the number of students who selected the cause prompt was nearly equal to the number of students who selected the effect prompt.

Approximately two weeks prior to each exam, the students received a reading list with the references for six readings in the course textbook. Approximately half of the readings were core texts that had been the focus of comprehension and vocabulary activities as part of the EAP class, while the other readings were in a supplemental section of the textbook that had not been the focus of instructional activities. The readings were taken from newspapers, popular magazines, websites, and scholarly books or articles, and ranged in length from 918 to 2,057 words. To help prepare for the exam, students were given a note-taking template that provided limited space (one page) for students to write down the complete reference, key words, main idea, and key supporting ideas from each source text.

On the day of the examination, the students received two writing prompts, both of which were related to the content of the source texts. Students selected which prompt they wished to address in their essays. The first prompt was to discuss the negative effects of human activities on wildlife (i.e., the Wildlife effects prompt), while the second prompt was to discuss the causes of ecosystem damage (i.e., the Ecosystem cause prompt). For both prompts, students were instructed to support their ideas by using the information from the readings that they had summarized on the note-taking templates. The students had three hours to write their cause and effect essays by hand, during which time they had access to a paper-based monolingual English dictionary and their notes. The instructors checked the students' notes to ensure that no additional information was brought into the exam.

Analysis

The students' handwritten essays were typed, verified, de-identified, and saved as Microsoft Word files by research assistants. The electronic files were submitted to Coh-Metrix (www.cohmetrix.com) to obtain measures of the first four lexical features listed below, and to the Compleat Lexical Tutor (Cobb, 2007) for the AWL measure (www.lextutor.ca).

1. *Lexical diversity*. A measure of lexical breadth, specifically how many different words are found in a text, lexical diversity indicates the range of vocabulary used by a writer. It has been associated with L2 writing quality (Crossley & McNamara, 2012; Jarvis, 2002) and judgements of writers' lexical proficiency (Crossley, Salsbury, et al., 2011). The specific measure used was the mean textual lexical diversity (MTLD), which is not affected by text length (for more information about the calculation and validation of MTLD, see McCarthy & Jarvis, 2010).

2. *Word frequency*. Another measure of lexical breadth, word frequency reflects how often specific words in texts occur. Previous L2 research has shown that more proficient writers use less-frequent words (Crossley, Salsbury, et al., 2011; Grant & Ginther, 2000; Meara & Bell, 2001; Reid, 1986, 1990). Coh-Metrix uses information from four corpora of English to calculate word frequency scores for specific categories of words, with frequency of content words selected as the specific measure (see Graesser et al., 2004, for more information about the calculation of word frequency).

3. *Imagability*. A measure of word accessibility, imagability is an index in the MRC Psycholinguistics database (Coltheart, 1981) that reflects human judgements about how easy it is to construct a mental image of a word. Previous research has shown that higher proficiency writers use words that are less imaginable (Crossley & McNamara, 2012; Crossley, Salsbury, et al., 2011).

4. *Word familiarity*. A measure of word accessibility, word familiarity is based on human judgements about how frequently a word appears in print. Unlike word frequency, which is based on actual frequency counts in multiple corpora, word familiarity is assessed by ratings from the MRC Psycholinguistics database (Coltheart, 1981). Previous L2 studies have reported that higher proficiency writers use words that are less familiar (Crossley & McNamara, 2012; Crossley, Salsbury, et al., 2011).

5. *Percentage of AWL words*. The AWL (Coxhead, 2000) contains 570 word families that account for an average of 10% of the total words in academic texts (see Coxhead, 2011, for a review of recent studies), and approximately 7% of the total words in the essays written by incoming university students (Douglas, 2013).

The TOEFL integrated-writing rubric (http://www.ets.org/s/toefl/pdf/toefl_writing_rubrics.pdf) was used to rate the students' cause and effect essays. The rubric was modified slightly by removing all references to lectures because these students had access to written sources only. Both research-

ers pilot-tested the rubric using cause and effect essays written in response to the different prompts, and compared their ratings. After resolving any disagreements and identifying benchmark essays for each score in the rubric, the first researcher rated all 94 essays. Rating guides that elaborated on the rubric descriptors were created. The first researcher trained two ESL instructors using the rating guides, the benchmark essays, and the source texts. After the raters completed the training and practiced rating essays that were not included in the dataset, they independently coded all 94 essays. Interrater reliability between the researcher and the two independent raters, assessed using a two-way mixed average-measures intraclass correlation coefficient, was .76. The mean ratings of all three raters were used for subsequent analyses.

Results

The dataset consisted of 94 integrated-writing cause and effect essays that were written using the same source texts, but addressed different prompts. Approximately half the students ($n = 49$) self-selected the Ecosystem cause prompt, and the other students ($n = 45$) chose to address the Wildlife effects prompt. Table 1 provides an overview of the lexical features of the source texts, along with the descriptive statistics for the essay ratings and the lexical features of the essays. The lexical features of the source texts did not differ dramatically from those of the students' essays, but their word frequency and familiarity scores were lower, while their lexical diversity and AWL percentage were higher.

Table 1
Features of Source Texts and Essays

| | <i>Source Texts</i> ($n = 6$) | | <i>Ecosystem Cause</i> ($n = 49$) | | <i>Wildlife Effects</i> ($n = 45$) | |
|-------------------|------------------------------------|-----------|--|-----------|---|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Length | 1,408 | 502 | 541 | 111 | 515 | 120 |
| Essay rating | n/a | n/a | 3.37 | .61 | 2.85 | .49 |
| Imagability | 421.64 | 21.75 | 418.89 | 13.82 | 425.94 | 14.84 |
| Word frequency | 2.05 | .11 | 2.15 | 10 | 2.21 | .10 |
| Lexical diversity | 100.73 | 28.48 | 90.03 | 28.69 | 86.74 | 23.10 |
| Word familiarity | 561.25 | 5.74 | 570.67 | 5.41 | 575.09 | 6.18 |
| % AWL words | 9.20 | 3.95 | 7.64 | 1.55 | 6.69 | 1.72 |

The first research question asked whether the writing prompt impacted the text quality and lexical features of the EAP students' essays. As shown in Table 1, the Ecosystem cause essays had higher ratings than the Wildlife effect essays, and had lower imagability, frequency, and familiarity scores. They also had greater lexical diversity and a higher percentage of AWL words than the Wildlife effect essays. Independent-samples *t*-tests (equal variance assumed) were used to compare the ratings and lexical features of the essays written in response to two prompts. To account for multiple comparisons, the initial alpha level of .05 was divided by the number of variables to be compared (6). This resulted in an adjusted alpha level of .008 (.05/6). As shown in Table 2, the differences in essay rating, word frequency, word familiarity, and AWL percentage were all statistically significant, with effect sizes ranging from small to medium (.40 to .70) and approaching large (1.00) based on general benchmarks for applied linguistics research (Plonsky & Oswald, 2014).

Table 2
Statistical Comparison by Prompt Type

| | 95% CI | <i>t</i> | <i>p</i> | Cohen's <i>d</i> |
|-------------------|---------------|----------|----------|------------------|
| Essay rating | .29, .74 | 4.49 | .001 | .94 |
| Imagability | -12.96, -1.22 | 2.40 | .019 | .49 |
| Word frequency | -.10, -.02 | 2.77 | .007 | .60 |
| Lexical diversity | -7.44, 14.02 | .61 | .544 | .13 |
| Word familiarity | -6.79, -2.40 | 3.69 | .001 | .76 |
| % AWL words | .24, 1.60 | 2.83 | .006 | .58 |

The second research question asked whether there was a relationship between text quality and lexical features. As shown in Table 3, most relationships between the lexical features and essay ratings were in the expected directions. Imagability, word frequency, and word familiarity decreased as essay ratings increased, which reflects the tendency for more proficient writers to use less imagable, less frequent, and less familiar words. However, Spearman correlations indicated that the relationships were not statistically significant. Whereas the occurrence of AWL words showed the expected positive relationship with essay ratings, it also failed to reach significance. The only unexpected relationship was for lexical diversity, which decreased as essay ratings increased. However, due to the wide confidence intervals and the failure of any of the correlation coefficients to reach statistical significance, the data do not provide evidence for an overall relationship between text quality and these lexical features.

Table 3
Correlation Coefficients Between Essay Ratings and Lexical Features

| | r_s | 95% CI | p |
|-------------------|-------|-------------|------|
| Imagability | -.119 | -.330, .092 | .253 |
| Word frequency | -.081 | -.292, .119 | .463 |
| Lexical diversity | -.163 | -.354, .025 | .117 |
| Word familiarity | -.097 | -.311, .104 | .352 |
| % AWL words | .026 | -.187, .234 | .803 |

Discussion

To summarize the findings, the essays written in response to the Ecosystem cause prompt were rated more highly, contained fewer frequent and familiar words, and had a greater percentage of AWL words than the essays written in response to the Wildlife effects prompt. Thus, the findings confirm those of previous studies, which found that prompt characteristics can affect the quality of EAP students' written texts (Cho et al., 2013; Cumming et al., 2005; Gebril & Plakans, 2013). However, there was no correlation between the essay ratings and the lexical breadth, word accessibility, or AWL usage in the students' texts, which does not support the findings of previous research (Crossley, Salsbury, et al., 2011). Methodological differences may help account for the divergent findings. Whereas Crossley, Salsbury, and McNamara (2011) examined the texts written by students at diverse proficiency levels, the EAP students in the current study were less varied in their global proficiency levels and had placed into the same EAP writing class. As a result, the variation in these students' essay scores may have captured differences in their academic writing abilities, rather than variation in their global English proficiency. Furthermore, the written texts in Crossley, Salsbury, and McNamara were rated using the SAT rubric, which includes explicit language about vocabulary use in the descriptors. The holistic rubric used here may not have emphasized vocabulary enough for the ratings to capture differences in the lexical features of the students' essays. Finally, the participants in Crossley, Salsbury, and McNamara (2011) wrote unstructured, independent-writing texts for research purposes, whereas these EAP students had two weeks to prepare for an integrated-writing task that was written under high-stakes assessment conditions. It is possible that having time to read and prepare source texts prior to writing the essays led to reduced lexical variation compared to independent writing tasks.

An interesting question raised by the findings is why the essays written in response to the Wildlife effects prompt received lower ratings and had dif-

ferent lexical features than the Ecosystem cause essays. In order to identify possible explanations for the divergent ratings and features, the effect essays were examined in more detail. One possible reason for their lower ratings is that two thirds of the Wildlife effects essays (31 of 45) focused on causes for the destruction of ecosystems, rather than the effects that arise after ecosystems are destroyed. For example, one student wrote a paragraph to support the topic sentence "Farming is another example of how overusage of natural resources leads to the destruction of animal life." Rather than elaborate on what happens after ecosystems have suffered damage, this student explained how farming contributed to the decline in ecosystem health. Similarly, another student devoted a paragraph of her essay to explaining how human activity caused the destruction of the Aral Sea, but she did not provide any information about the subsequent effects of that destruction on wildlife. Because the rubric emphasized the relevance of the ideas presented to the writing prompt, Wildlife essays in which students focused on causes rather than effects received lower ratings.

An additional factor that may have contributed to the lower ratings is the use of personal information and prior experiences in the Wildlife effects essays. Because the exam was an academic integrated-writing task, students were instructed to draw upon source text information to support their ideas, as opposed to writing about their own experiences or beliefs. Nevertheless, nearly one third of the students who chose the Wildlife effect prompt (14 of 45) used personal experiences or information unrelated to the source texts to support their ideas, which had a negative effect on the essay ratings. For example, one student claimed that a second reason for ecosystem damage (which is a cause rather than an effect) was human wars, and referred to the use of nuclear bombs to support the idea. Similarly, another student argued in a Wildlife effects essay that tourism has created negative effects for wildlife, even though none of the source texts were related to that topic. Conversely, few students who chose the Ecosystem cause prompt (5 of 49, or 10%) relied upon their own experiences to support their main points. The students' use of personal experiences, rather than source text information, may have also impacted the lexical features of their essays, particularly their use of more frequent and familiar words, and their lower AWL percentage.

In order to explore whether the source texts may have played a role in the differences between the Wildlife effect and Ecosystem cause essays, we examined whether the source texts were used with equal frequency by students who responded to different prompts. Overall, there were no differences in the source texts used by students who responded to different prompts. With the exception of one source text that was referred to by only eight students (seven of whom wrote cause essays), all the source texts were used frequently and equally by students who responded to different prompts. Although the infrequently cited source text was the shortest (871 words), it had the highest percentage of AWL use (14.53%), which may have contributed to its less

frequent use by the students. We also examined whether the content of the source texts was applicable to both prompts. In general, all six source texts tended to emphasize the causes of environmental problems, as opposed to the effects of environmental damage on nature, wildlife, or people.

The orientation of the source texts toward causes, rather than effects, also carried over to the students' notes, which they compiled from the source texts prior to writing their essays under examination conditions. Although a complete analysis of the students' notes was beyond the scope of the current study, we did examine the ones completed and submitted (50 of 94 students) to determine if their notes contained more information related to the causes of environmental problems, which had been the focus of the source texts, than the effects of ecosystem damage. The orientation toward causes was reflected in the resource sheets, as every student wrote down multiple causes for environmental problems when they took notes about the source texts. In contrast, fewer students (only 20%) included specific information about effects of environmental degradation on people, plants, or animals.

Having identified differences in the ratings and lexical features of essays written in response to cause and effect prompts, the findings raised potential implications for EAP writing instructors. First, in contexts where integrated-writing tasks are used, the selection of source texts may play an important role in the students' essays. When multiple prompts are used for the same source texts, it is important to ensure that an equal amount of information presented in those texts is applicable to either prompt. In the current study, the majority of the source text information was oriented toward causes, which may have made it more difficult for students who selected the effect prompt to integrate source information into their essays. Second, the difficulty of the source texts should be considered, so that students do not avoid using texts because they perceive them as too difficult. Simply providing "easy" texts, however, is not advisable, as EAP students are expected to read academic sources (textbooks and journal articles) in their content classes. Numerous comparative studies (Biber & Reppen, 2002; Hyland, 1994; Miller, 2011) have shown that ESL textbooks do not reflect the grammatical, lexical, and discourse features of the academic texts that EAP students are expected to read in their disciplines. Furthermore, authentic texts of the cause and effect analytic genre have more sophisticated linguistic and discourse features than the features described in ESL and EAP textbooks (Moreno, 2003). As a result, it may be preferable to provide source texts that are comparable to academic texts in terms of language and discourse features, but to scaffold EAP students' reading comprehension and note-taking in order to ensure that the content of the source text is available for use during an integrated-writing task.

Similar to the findings of research with incoming university students (Douglas, 2013), these EAP writers included approximately 7% AWL vocabulary in their essays, which is lower than the occurrence of AWL words in the texts of more expert writers (Hancioğlu, 2009) and in academic textbooks

and journal articles (Coxhead, 2000). The findings suggest that EAP students might benefit from instruction designed to help them use AWL words in their own writing. Several overviews of vocabulary instruction (e.g., Coxhead & Byrd, 2007; Folse, 2011) have highlighted strategies for helping EAP students increase their knowledge and use of AWL words. For example, rather than use annotated readings or excerpts, it may be helpful to provide EAP students with whole texts whose length and language features are comparable to the texts they are reading and writing about in their academic classes. Besides working with EAP students to facilitate comprehension of authentic whole texts, attention can be directed toward the study of academic language (AWL vocabulary, grammatical structures, and discourse patterns) and work with strategies for including these features in their own written texts. Particularly in EAP writing classes, providing feedback and opportunities for revision of lexical features may help students acquire the ability to use more of the lexical features associated with academic writing.

Although the findings shed light on the impact of writing prompts on EAP students' text quality and lexical features, several limitations should be acknowledged that may limit the study's generalizability. As pointed out previously, the EAP students here represented a relatively homogeneous group in terms of their standardized proficiency scores and their assignment to the same EAP class based on a university placement exam. Since previous research has found evidence for a relationship between lexical features and global proficiency ratings, future research should identify which lexical features are most useful for discriminating among essay ratings for students within specific proficiency levels. Recent learner corpus research (Cobb & Horst, 2015; Crossley, Cobb, & McNamara, 2013) has highlighted a variety of measures for assessing L2 speakers' lexical proficiency that are predictive of human ratings of text quality.

An additional limitation is the unequal distribution of essays across prompt type and text ratings. Two thirds of the essays that received low ratings (lower than 3) were written in response to the Wildlife effects prompt. Conversely, almost 90% of the highly rated essays (3.67 and up) were written in response to the Ecosystem cause prompt. Due to the unequal distribution of high and low scoring essays across prompt type, the relationship between essay ratings and lexical features may have been obscured by prompt effects. Future research should further clarify the relationship between essay ratings and lexical features, using datasets in which students respond to the same prompt or the distribution of essays across rating levels is similar for all prompts. In light of the potential impact of rubric type on essay ratings, future studies could benefit from the use of analytic rubrics with descriptors specifically about language use, especially for research that examines relationships between linguistic features and human judgements of text quality.

Finally, although the students' resource sheets were examined for evidence that they oriented to causes rather than effects when taking notes from

the source texts, a more thorough analysis of their notes may provide more insight into the role of the contribution of the source text to their essay ratings. Future studies should explore what information EAP students extract from source texts, and how they record this information for future use in their own writing. Although numerous studies have documented whether EAP students copy verbatim from source texts (for recent research, see the special issues of the *Journal of English for Academic Purposes*, 2013, and the *Journal of Second Language Writing*, 2012), fewer studies have examined the process by which EAP students take notes about source texts and then use their notes (as opposed to the source texts) during integrative-writing tasks. Taking a broader perspective on the use of source texts (including selection, comprehension, note-taking, and incorporation into writing tasks) would provide greater insight into students' decisions about what information is important and how they record that information for future use. Furthermore, qualitative analysis to identify the factors that influence EAP students' choice of writing prompts and how they use source text versus personal experience in their essays would broaden our understanding of how they understand and carry out integrative writing tasks. By addressing these limitations in our future research, we aim to further clarify the complex relationships among EAP students' use of source text information, the lexical features of their written texts, and ratings of text quality.

Acknowledgements

This research was supported by the Concordia University Summer Research Award granted to the first author, and the research grant from the Canada Research Chairs program awarded to the second author. We would like to thank Matthew Lazenby and Jonathan Brouillette for their help rating the essays, and Randy Appel, Suzanne Cerreta, Vanessa Copeland, Jennifer Foote, Yeji Han, and Paula Kielstra for their assistance with data collection and handling.

The Authors

Maxime Lavallée is a recent graduate of the Teaching English as a Second Language Specialization program at Concordia and a graduate student in Second Language Education at McGill University. His research and teaching interests involve English for Academic Purposes at the postsecondary level.

Kim McDonough is a Professor and Canada Research Chair in applied linguistics in the Department of Education at Concordia University. Her research interests include interaction, task-based language teaching, and psycholinguistic approaches to second language acquisition.

References

- Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24(2), 199–208.
- Chen, Q., & Ge, G.-C. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes*, 26(4), 502–514.
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49.

- Cho, Y., Rijmen, F., & Novák, J. (2013). Investigating the effects of prompt characteristics on the comparability of TOEFL iBT™ integrated writing tasks. *Language Testing*, 30(4), 513–534.
- Cobb, T. (2007). *The compleat lexical tutor*. Retrieved from <http://www.lextutor.ca/>
- Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15–38). Amsterdam, Netherlands: John Benjamins.
- Cobb, T., & Horst, M. (2015). Learner corpora and lexis. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 185–206). Cambridge, UK: Cambridge University Press.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A, 497–505.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly*, 45(2), 355–362.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16(3), 129–147.
- Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252–267). Cambridge, UK: Cambridge University Press.
- Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41(4), 965–981.
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18(2), 119–135.
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115–135.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2), 243–263.
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2010). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561–580.
- Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28(3), 282–311.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5–43.
- Douglas, S. R. (2013). The lexical breadth of undergraduate novice level writing competency. *Canadian Journal of Applied Linguistics*, 16(1), 152–170.
- Folse, K. (2011). Applying L2 lexical research findings in ESL teaching. *TESOL Quarterly*, 45(2), 362–369.
- Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly*, 10(1), 9–27.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193–202.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123–145.
- Hancıoğlu, N. (2009). *Incorporating corpus data into an advanced academic thesis writing course* (Unpublished doctoral dissertation). Eastern Mediterranean University, North Cyprus.
- Hancıoğlu, N., Neufeld, S., & Eldridge, J. (2008). Through the looking glass and into the land of lexico-grammar. *English for Specific Purposes*, 27(4), 459–479.

- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes*, 13(3), 239–256.
- Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies*, 9(2), 111–129.
- Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57–84.
- McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86.
- Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics*, 26(1), 32–47.
- Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5–19.
- Miller, D. (2011). ESL reading textbooks vs. university textbooks: Are we giving our students the input they may need? *Journal of English for Academic Purposes*, 10(1), 32–46.
- Moreno, A. L. (2003). Matching theoretical descriptions of discourse and practical applications to teaching: The case of causal metatext. *English for Specific Purposes*, 22(3), 265–295.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912.
- Reid, J. (1986). Using the writer’s workbench in composition teaching and testing. In C. Stansfield (Ed.), *Technology and language testing* (pp. 167–188). Alexandria, VA: TESOL.
- Reid, J. (1990). Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 191–210). Cambridge, UK: Cambridge University Press.
- Roessingh, H. (2006). BICS-CALP: An introduction for some, a review for others. *TESL Canada Journal*, 23(2), 91–96.
- Sawaki, Y., Quinlan, T., & Lee Y.-W. (2013). Understanding learner strengths and weaknesses: Assessing performance on an integrated writing task. *Language Assessment Quarterly*, 10(1), 73–95.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512.
- Vongpumivitch, V., Huang, J.-Y., & Chang, Y.-C. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33–41.
- Wolfersberger, M. (2013). Refining the construct of classroom-based writing-from-readings assessment: The role of task representation. *Language Assessment Quarterly*, 10(1), 49–72.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215–229.