A Lexical Profile Analysis of a Diagnostic Writing Assessment: The Relationship between Lexical Profiles and Writing Proficiency

by

Rose Katagiri

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Arts in

Applied Linguistics and Discourse Studies

Carleton
University
Ottawa, Ontario

Abstract

Effective use of vocabulary contributes to academic language proficiency and academic success (Douglas, 2013). Sophisticated vocabulary use relates to increased quality of writing (Laufer & Nation, 1995; Kyle & Crossley, 2015). However, it is unclear which lexical sophistication characteristics contribute to writing quality assessments. Furthermore, previous studies focused on general assessments, rather than English for Specific Purposes diagnostic assessments which aid in early intervention for academic support. The present study investigates the relationship between vocabulary sophistication indices and writing scores ($N = 353$) on a post-entry university diagnostic test for engineers (Fox & Artemeva, 2017). Multiple lexical sophistication approaches were compared to writing scores and differences in lexical profile characteristics of successful and unsuccessful students were compared. Results indicated that samples of successful writing have a higher presence of tokens, types, lexical stretch, academic vocabulary and formulaic language. The findings have pedagogical implications for remedial writing instruction of engineering students.

Keywords: lexical sophistication, lexical frequency profiling, diagnostic assessment, academic vocabulary, discipline-specific vocabulary, formulaic language, academic writing, English for Specific Purposes

Table of Contents

Abstract ii

Table of Contents iii

List of Tables vi

List of Figures vii

List of Acronyms viii

Chapter 1: Introduction 1

1.0 Research context and gaps 1

1.2 Guiding Research Questions 3

1.3 Organization of the thesis 4

Chapter 2: Literature Review 5

2.0 Introduction 5

2.1 Importance of Vocabulary in Academic Writing 6

2.2 Lexical Sophistication 9

2.3 Lexical Frequency Profiling 10

2.4 Lexical Frequency Profile and Lexical Stretch (Douglas, 2010, 2013, 2015) 17

2.5 Tokens, Type, and Type-Token Ratio (TTR) 21

2.6 Specialized Vocabulary 22

2.7 Tool for the Automatic Analysis of Lexical Sophistication 33

Chapter 3: Methodology 39

3.0 Introduction 39

3.1 Materials 40

3.2 Research Question 1: Holistic scores and vocabulary scores 45

3.3 Research Question 2: Lexical Frequency and Lexical Stretch 48

3.4 Research Questions 3a and 3b: Specialized Vocabulary 57

3.5 Research Question 3c: Academic Formulaic Language 61

3.6 Tool for Automatic Analysis of Lexical Sophistication (TAALES) 65

Chapter 4: Discussion 68

4.0 Introduction 68

4.1 Research Question 1: The relationship between holistic and vocabulary scores 68

4.2 Research Question 2: Lexical Frequency and Lexical Stretch 72

4.3 Research Question 3a: Academic Vocabulary 87

4.4 Research Question 3b: Discipline-Specific Vocabulary 92

4.5 Research Question 3c: Academic Formulaic Language 94

Chapter 5: Conclusion 100

5.0 Introduction 100

5.1 Pedagogical Implications 100

5.2 Limitations and Future Research 107

5.3 Conclusion 109

References 110

Appendices 124

Appendix A Distribution of Holistic Test Scores According to Test Versions.........124

Appendix B Occurrences of 3-word Academic Formulas for the at-risk group 125

Appendix C Occurrences of 4-word Academic Formulas List for the at-risk group...126

Appendix D Occurrences of 5-word Academic Formulas List for the at-risk group...127

Appendix E Occurrences of 3-word Academic Formulas Lists for the not-at-risk group 128

Appendix F Occurrences of 4-word Academic Formulas Lists for the not-at-risk group 130

Appendix G Occurrences of 5-word Academic Formulas Lists for the not-at-risk group 131

List of Tables

Table 1 Number of total samples and tokens according to at-risk vs. not-at-risk group…43

Table 2 Average number of types and tokens according to at-risk and not-at-risk group..44

Table 3 Descriptive statistics for analytic vocabulary scores according to at-risk and not-at-risk group 46

Table 4 Descriptive statistics for holistic scores according to at-risk and not-at-risk group 47

Table 5 Mann-Whitney U test results between at-risk and not-at-risk group… 48

Table 6 Correlations between lexical profiling measures and holistic scores… 52

Table 7 Independent *t*-test results between at-risk and not-at-risk for token and types…..55

Table 8 Mann-Whitney U test results between at-risk and not-at-risk group for six measures.. 55

Table 9 Descriptive statistics for AWL token percentages according to at-risk and not-at-risk group… 58

Table 10 Average percentage of AWL tokens and types in at-risk and not-at-risk texts….59

Table 11 Descriptive statistics of the EEWL token percentages according to at-risk and not-at-risk group 59

Table 12 Average percentage of EEWL tokens and types in at-risk and not-at-risk texts...60 Table 13 Descriptive statistics for overall AFL occurrences according to at-risk and not-at-risk group 63

Table 14 Total frequency of occurrences for the three-, four-, five-word occurrences AFL ………………………………………………………………………......................64

Table 15 Correlations between TAALES Measures and Vocabulary Analytic Scores 66

List of Figures

Figure 1 Example of not-a-risk writing exercise that at-risk students can complete 104

Figure 2 Example of at-risk writing exercise that at-risk students can complete............. 105

List of Acronyms

AFL	Academic Formulas List
AWL	Academic Word List
BNC	British National Corpus
CAEL	Canadian Academic English Language Assessment
CELPIP	Canadian English Language Proficiency Index Program
COCA	Corpus of Contemporary American English
DGIT	Discipline-specific Graph Interpretation Task
EEWL	English Engineering Word List
ESL	English as a Second Language
HFV	High-Frequency Vocabulary
IELTS	International English Language Test System
LFP	Lexical Frequency Profile
LFV	Low-Frequency Vocabulary
MAWL	Medical Academic Word List
MFV	Mid-Frequency Vocabulary
NAWL	Nursing Academic Word List
SAFE	Self-Assessment for Engineers
TAALES	Tool for the Automatic Assessment of Lexical Sophistication
TOEFL	Test of English as a Foreign Language
TTR	Type-Token Ratio

Chapter 1: Introduction

1.0 Research context and gaps

In 2017, the enrollment in Canadian engineering undergraduate programs was 82,480 and this number continues to grow (Engineers Canada, 2017). In order to succeed as an engineer, students are required to not only develop knowledge of engineering related subjects, but to also develop linguistic skills- such as those needed for academic writing or public speaking- in order to communicate effectively (Budinski, 2001; Winsor, 2013). Writing is particularly important for engineering students' future careers because "Many engineers spend 40% of their work time writing, and usually find the percentage increases as they move up in the corporate ladder" (Beer & McMurrey, 2014, p. 3). However, few engineering schools [universities] offer adequate (if any) courses in engineering communication (Beer & McMurrey, 2014).

Researchers have explored the 'first-year experience' as there has been a significant number of first-year undergraduate students who continue to drop out of their program (e.g., Browne & Doyle, 2010; Fox, Haggerty, & Artemeva, 2016). Although there is an increasing number of universities that offer academic support, this support is generic and does not consider discipline-specific pedagogical initiatives (e.g., Fox et al., 2016; Fox, von Randow, & Volkov, 2016). Fox (2005), as well as Meyer and Land (2003) suggest that discipline-specific literacies and academic resources are important factors for program retention and completion.

In response to this, Fox and Artemeva (2017) developed the Self-Assessment for Engineers (SAFE) in response to generic resources and the need for the development of specialized disciplinary and professional varieties of English (e.g., Fox et al., 2016; Fox et al., 2016). SAFE is an English for Specific Purposes (ESP) diagnostic assessment administered to first-year undergraduate engineering students at a Canadian university. SAFE aims to provide students with a meaningful diagnosis for an indication of academic risk through authentic engineering tasks and academic literacies. This diagnosis, in turn, enables students to seek individualized discipline-specific academic support. Based on the test results, students may fall under one of three categories indicating their need for academic support: at-risk, somewhat at-risk, and not-at-risk. Students who fall in the at-risk category are of focus as they may need early intervention and academic support which in turn may prevent future failure and encourage student retention, and academic success (Fox et al., 2016). At-risk students are invited to visit a discipline-specific academic support learning center. To better inform the academic support at-risk students receive, the developers of SAFE (Fox & Artemeva, 2017) have tasked the author of this thesis to investigate the lexical characteristics that contribute to discipline-specific writing proficiency.

SAFE seeks to diagnose discipline-specific academic literacies. Academic writing ability is viewed as an essential skill for the achievement of success in university studies and in the future workplace (Geiser & Studley, 2002; Powell, 2009; Beer & McMurrey, 2014). There are various factors that contribute to the quality of academic writing such as rhetoric (Hyland, 2008), cohesive devices (Crossley, Kyle, & McNamara, 2016), syntax (Crossley, Cai, McNamara, 2012), genre (Cheng, 2007), formulaic language (Appel & Wood, 2016), and the length of composition (Appel & Wood, 2016; Douglas, 2015).

One important factor that has been shown to correlate significantly with writing proficiency scores is the use of sophisticated vocabulary (the size and range) of a writer's lexicon (e.g., Douglas, 2013; Laufer & Nation, 1995; McNamara, Crossley, & McCarthy, 2010; Nation, 2001). Students who produced low-proficiency writing have been found to use less sophisticated vocabulary (Douglas, 2015; Goodfellow, Lamy, & Jones, 2002; Laufer & Nation, 1995). The lexical characteristics that contribute to writing proficiency can be investigated using a method called *lexical frequency profiling* which in brief refers to describing the lexical content of a text according to the occurrences of different kinds of lexical characteristics. Previous studies have explored lexical characteristics that contribute to writing quality such as frequency and use of academic vocabulary (Goodfellow et al., 2002; Laufer & Nation, 1995), use of vocabulary from test prompts (Appel & Wood, 2016), use of formulaic language (e.g., Ädel & Erman, 2012; Appel & Wood, 2016), and indices such as the number of tokens and types (Douglas, 2015).

Despite previous research establishing the importance of vocabulary in writing for general assessments, the relationship between vocabulary and holistic writing scores for an English for Specific Purposes *diagnostic* assessment for discipline-specific literacies (such as engineering) has yet to be established. In particular, what remains unclear is the impact of lexical sophistication characteristics such as word frequency, academic vocabulary, discipline-specific vocabulary, and use of formulaic language on writing proficiency scores.

1.2 Guiding Research Questions

The present study investigates the relationship between vocabulary and holistic writing scores from a discipline-specific diagnostic assessment using lexical sophistication measures. In doing so, this study aims to investigate and compare the lexical characteristics of at-risk and not-at-risk student writing from a discipline-specific task. The guiding research question is:

1. What is the relationship between holistic writing scores and vocabulary usage for an engineering-specific writing task?

In order to answer the overarching question, the present study completed a corpus analysis of not-at-risk and at-risk student writing from SAFE. The SAFE rubrics' vocabulary and holistic scores were used to establish the relationship between the respective two variables. Then, four approaches to lexical sophistication were utilized to understand and differentiate the lexical characteristics of not-at-risk and at-risk writing. By characterizing the writing of not-at-risk and at-risk use of vocabulary, we can better approximate the students' needs for diagnosis and influence the assistance given to at-risk students which can lead to a successful academic career.

1.3 Organization of the thesis

The present study is organized into 5 chapters. Chapter 2 provides an overview of the importance of vocabulary in the quality of writing. As well, there is a discussion of four approaches to lexical sophistication used in the literature and their relationship to writing proficiency. Next, Chapter 3 describes the SAFE diagnostic process and corpus. Then, for each research question, the methodological processes and the results are presented together four times as there were four approaches used. Following this, Chapter 4 presents a discussion of the results according to each research question. Finally, Chapter 5 provides a conclusion which includes pedagogical implications, limitations of the present study, and directions for future research.

Chapter 2: Literature Review

2.0 Introduction

This chapter reviews the literature and methods associated with the main areas of investigation in this study. Section 1 discusses the importance of academic writing and is further expanded to include vocabulary's influence in academic writing scores. Section 2 outlines four approaches to operationalizing lexical sophistication and applications in previous empirical studies. The first approach reviewed is the Lexical Frequency Profile (Laufer & Nation, 1995), and within this approach is Lexical Stretch (Douglas, 2010, 2013, 2015). The second approach is the study of the occurrence of specialized vocabulary which includes academic and discipline-specific vocabulary. The third approach is the study of the use of formulaic language in writing. Finally, the Tool for the Automatic Assessment of Lexical Sophistication 2.0 (TAALES; Kyle & Crossley, 2015) will be reviewed. In this thesis, these approaches will be applied to operationalize or support sophisticated vocabulary used in writing and their relationship with successful and unsuccessful writing scores.

1.0 Academic Writing

Writing successfully is viewed as an essential skill for the achievement of success in university studies and in the future workplace (Beer & McMurrey, 2014; Geiser & Studley, 2001; Powell, 2009). Furthermore, academic writing may be a predictor of academic success for Grade Point Average (GPA; Elder, Bright, Bennet, 2007; Harrington & Roche, 2014). However, academic writing is a challenging skill to develop for both first and second language learners alike (e.g., Wood, 2015). This has prompted researchers to investigate the writing characteristics of successful and unsuccessful academic writing and use the results to inform pedagogical approaches to help improve unsuccessful writing (e.g., Appel & Wood, 2015; Crossley, Roscoe, & McNamara, 2014; McNamara et al., 2010). *Unsuccessful writing* in this thesis will refer to writing that has not obtained an adequate score according to the standards of the assessment for which it was produced.

Studies have acknowledged that academic writing is a complex task (e.g., Crossley et al., 2014; Raimes, 2002). Writing requires a variety of skills be able to piece together words and communicate in a logically coherent manner while monitoring grammar and selecting the appropriate vocabulary according to the purpose of writing. This is supported by Raimes (2002) who identifies writing as a complex task in which writers must generate ideas, present them, and be critical to both the ideas and structure of the piece.

Previous studies have examined linguistic features that contribute to the quality of writing using holistic writing scores as a measurement of success. Crossley et al. (2014) established that successful writing can employ a combination and variety of linguistic features. In the literature, the linguistic features found to characterize writing proficiency include: rhetoric and genre (e.g., Bruce, 2008; Hyland, 2006; Hyland & Tse, 2007; Swales, 1998), cohesive devices (e.g., Crossley, Kyle, & McNamara, 2016), formulaic language (e.g., Appel & Wood, 2016; Paquot, 2018), lexical diversity (e.g., González, 2017), sophistication of vocabulary (Kyle & Crossley, 2015; Kyle et al., 2017) and the length of composition (e.g., Appel & Wood, 2016; Douglas, 2015; McNamara et al., 2010), and many more. Specifically, the focus of this thesis will be on the lexical characteristics that contribute to holistic writing scores on a discipline-specific ESP-writing task.

2.1 Importance of Vocabulary in Academic Writing

Many researchers agree that vocabulary plays an important role in the development of writing and language skills in general (e.g., Coxhead, 2012; Douglas, 2013; Nation, 2001; Paquot, 2010). This is because vocabulary is the building block of language. Furthermore, vocabulary has been recognized as an important factor that may predict educational success

(Harrington & Roche, 2014) and contributes to academic language proficiency (Milton & Treffers-Daller, 2013; Pinchbeck, 2017) in both first and second language contexts (Webb & Nation, 2017). Most crucially, previous studies have identified that vocabulary use has an influence on the holistic evaluation of academic writing quality for various types of assessment. Astika (1993) investigated the components of an analytic scoring rubric on holistic writing scores using the English as a Second Language (ESL) Composition Profile developed by Jacobs, Zingraf, Wormuth, Hartfiel, and Hughey (1981) which featured Content, Organization, Vocabulary, Language use, and Mechanics as the sub-scores, hereby referred to as *analytic scores*. A total of 210 non-native writing samples from four different task topics were sampled from the Writing Sample Test at the University of Hawaii and assessed using the ESL Composition Profile (Jacobs et al., 1981). A multiple regression analysis for the analytic scores of the writing features revealed that vocabulary (83.75%), contributed the most to the variance in total scores, hereby referred to as the *holistic score*. The other analytic features contributed far less, ranging from .29% to 8.06%. This result indicates that vocabulary is an important underlying variable for writing assessments.

In a similar study, Espinosa (2005) also found a significant strong positive correlation between vocabulary proficiency and writing ability for two groups: Group A (n = 129) whose vocabulary knowledge was measured as very poor to fair and Group B (n = 55) whose vocabulary knowledge was measured as average to excellent. The results indicated a significant correlation between vocabulary proficiency and writing proficiency for Group A ($r = .90$, $p < .001$) and Group B ($r = .88$, $p < .001$). This study provides further evidence of the relationship between lexical proficiency and holistic writing scores.

The above studies both used Jacobs' et al. (1981) rubric criteria which has been established to have concurrent validity; test scores being highly correlated with those of the Test of English as a Foreign Language (TOEFL) and the Michigan Test Battery. Thus, further establishing that vocabulary use is an important part of writing assessment.

Roessingh (2008) likewise emphasizes the role of vocabulary in writing quality. Roessingh identified analytic vocabulary scores as an underlying variable for writing quality based on the analysis of texts from the Alberta English 30 Diploma exam for a mixed sample of native (n = 31) and non-native writers (n = 48). The texts were examined using the six analytic score criteria: Thought and Detail, Writing Skills, Organization, Matters of Choice (vocabulary) and Mechanics and Grammar. Roessingh was able to establish a moderately significant correlation between holistic writing scores and analytic vocabulary scores, $r = .59$, $p < 0.01$. Vocabulary scores explained 35% of the variance in holistic writing scores. The study also found that low vocabulary scores negatively impacted other analytic scores. On the other hand, high vocabulary scores positively impacted the other analytic scores. Roessingh suggested that vocabulary measures may be a predictor of writing quality, where more diversity of vocabulary is associated with higher writing scores.

In a study using a different form of assessment, Lee, Gentile, and Kantor (2009) examined 930 non-native writing samples for two separate prompts on a computer-based TOEFL writing assessment. Significant correlations between holistic scores and analytic vocabulary scores were found. There were significant, moderate positive correlations ($r = .50$ and $r = .44$, respectively) for two different prompts with vocabulary explaining 25% and 20%, respectively of variance in holistic scores.

The studies described above have demonstrated that vocabulary has an influence on various forms of writing assessment. What is not yet clear is the impact of analytic vocabulary scores on a discipline-specific diagnostic test. This thesis aims to address this gap. The relationship between analytic vocabulary scores and holistic writing scores will be investigated in addition to the lexical sophistication characteristics of the vocabulary used in writing.

2.2 Lexical Sophistication

The effective use of vocabulary has been found to be a contributing factor in academic language proficiency and a key element of academic outcomes (Bialystok, Luk, Peets, & Yang, 2010; Milton & Treffers-Daller, 2013). Previous studies have explored the lexical factors that contribute to holistic writing scores such as word frequency and use academic vocabulary (Goodfellow et al., 2002; Laufer & Nation, 1995), syntactic complexity (McNamara et al., 2010) use of vocabulary from test prompts (Appel & Wood, 2016), and indices such as the number and diversity of words (McNamara, Crossley, & McCarthy, 2010; Jarvis, 2002; Yu, 2009). A frequently used measure is the lexical richness or *sophistication* (size and range) of vocabulary used by a writer which has been linked to increased quality of writing (e.g., Brynildssen, 2000; Laufer & Nation, 1995; Smith, 2003). The use of sophisticated vocabulary has been shown to correlate significantly with writing quality scores (e.g., Engber, 1995; Douglas, 2013; Laufer & Nation, 1995; McNamara et al., 2010).

A large and growing body of literature has investigated the role of lexical sophistication in writing. Although the exact definition of lexical sophistication has not been determined (Kyle & Crossley, 2015), there have been some influential lexical sophistication researchers who have attempted definitions. Laufer and Nation (1995) defined lexical sophistication as “the degree to which a writer is using a varied and large vocabulary” (p. 307). In other words, the breadth and depth of lexical knowledge available to writers (Meara, 1996, 2005; Read, 1998). These definitions demonstrate that a variety of indices can be considered as measures for lexical sophistication. Lexical sophistication can be operationalized using a variety of methods. This may be a result of researchers continuing to find additional indices that contribute to lexical sophistication and writing scores.

Four methods for measuring lexical sophistication will be used for the analysis of student writing in this thesis and their relationship to writing proficiency assessment will be discussed. The approaches are: 1) Lexical Frequency Profiling (Laufer & Nation, 1995); 2) academic and discipline-specific vocabulary; 3) academic formulaic language, and 4) Tool for Analysis of Lexical Sophistication (Kyle, Crossley, & Berger, 2017). Such measures of lexical sophistication may provide a reliable way to identify how linguistically at-risk (unsuccessful) writers differ from not-at-risk (successful) writers for remedial instruction. This is important considering how writing and vocabulary may provide a reliable way of identifying writers who could be linguistically at-risk, with classification rates of 75-80% being reported for identifying such students based on writing and vocabulary (Harrington & Roche, 2014). Furthermore, the vocabulary characteristics of not-at-risk writing may be able to inform at-risk writing and therefore pedagogical recommendations can be provided based on levels on proficiency.

2.3 Lexical Frequency Profiling

Lexical frequency profiling[1] (LFP) has been a commonly used method to assess the sophistication of a text. This typically involves calculating the proportion (frequency of use) of a vocabulary characteristic used in a text. There have been two repeatedly used methods for the operationalization of lexical sophistication from a frequency perspective: 1) frequency word lists (e.g., Douglas, 2015; Higginbotham & Reid, 2019; Laufer & Nation, 1995; Morris & Cobb, 2004) in which the frequency of words is derived from vocabulary frequency lists; and 2) corpus-derived frequency counts (McNamara, Crossley, & Roscoe, 2013) in which the frequency of words is derived from a large collection of texts. Lexical sophistication in this thesis is operationalized through the frequency word list method. Based on the explanation provided by

[1] In the literature, the term 'lexical frequency profiling' has also been used by Laufer and Nation (1995) to profile texts with the word list recourses they had at the time. See subsequent discussion.

Laufer and Nation (1995), lexical items that occur more frequently in a corpus are deemed less sophisticated, while less frequent or rare words are more sophisticated. This measurement is based on frequency lists that define vocabulary items as sophisticated or not based on where they fall in the frequency list. Then, the proportion of occurrence for the vocabulary item is calculated in relation to the examined text. The most commonly used frequency word list in applied linguistics research was the General Service List (GSL; West, 1953) which has 2000 of the most commonly occurring word families in English. In this thesis, *word families* are defined as the word stem plus the inflectional and derivational forms (Nation, 2001, 2013). Another list, the University Word List (UWL) contains 836-word families that do not occur in the GSL but are common in academic texts (Xue & Nation, 1984). However, with the advancement of corpus linguistic technology, new and more comprehensive lists have been developed since the GSL and UWL.

Presently, one of the most commonly used word frequency lists is the BNC-COCA 25. Nation (2012) organized the most frequently occurring vocabulary in English into 25 vocabulary lists, each composed of 1,000-word (1K) families of decreasing frequency according to the British National Corpus (BNC; Leech, Rayson, & Wilson, 2002) and Corpus of Contemporary American English (COCA; Davies, 2010), hereby referred to as the BNC-COCA (Nation, 2012). The BNC contains 100+ million words while the COCA contains 450+ million words. Within the BNC-COCA 25, the 1K list consists of 1,000-word families, which includes the most commonly used vocabulary, i.e., the least sophisticated vocabulary. The 25K list includes the 1,000 lowest frequency word families (rare words), i.e., the most sophisticated vocabulary.

For LFP, the lists outlined above (or any list for that matter) can be used in conjunction with corpus analytic software to analyze a text, from which inferences can be made. With the advancement of corpus technology, some corpus analytic software have become available online, Web Vocabprofile (Cobb, n.d.; Heatley & Nation, 1994), or as a downloadable software, such as AntWord Profiler (Anthony, 2014). These software tools allow for any type of word list to be analyzed for frequency of occurrence in the desired texts. The present study uses AntWord Profiler because of its ability to analyze a greater number of texts individually in comparison to VocabProfile which can only analyze one large piece of text at a time.

The LFP has been used in studies that examine factors that correlate with writing quality and the closest to a standard analysis of lexical sophistication (Douglas, 2010, Laufer & Nation, 1995; Lemmouth, 2008). Previous literature has found a relationship between lexical frequency and holistic writing scores (Crossley, Cobb, & McNamara, 2013; Crossley et al., 2014; Laufer & Nation, 1995; McNamara et al., 2010; McNamara, Crossley, Roscoe, Allen, & Dai, 2015). However, more recent studies have reported that lexical frequency may not be as strong of a predictor for writing proficiency as once thought (Kyle & Crossley, 2015; Kyle et al., 2017).

In their seminal article, Laufer and Nation (1995) recommended a measure of lexical sophistication called the Lexical Frequency Profile (LFP) in an attempt to compensate for the short comings of other lexical sophistication measures. The LFP was calculated by examining the proportions of the 1,000 and 2,000 most frequent words, University Word List (UWL), and Not in List words used by a writer. Laufer and Nation measured lexical sophistication using the LFP for two writing compositions (200-350 words) by one L2 writer from each of the three groups of increasing proficiency levels (n = 65) during a timed (60 minutes), in-class independent writing task. The options for writing prompt topics were argumentative and pertained to general controversial issues which did not require expert knowledge. The results suggested that LFP may be a valid and reliable predictor of productive language use and proficiency. There was a trend for lower proficiency students to use more higher frequency words (1K lists), i.e., less sophisticated words, in comparison to the higher proficiency writers who used fewer words from those lists. However, the difference was significant for only the 1K list in comparison to the 2K list. As well, the lower proficiency writers used fewer UWL words and Not in List words in comparison to higher proficiency writers. In other words, higher proficiency writers tended to use more academic vocabulary and rarer words, both of which are both considered sophisticated because they occur infrequently in English. The findings suggest that LFP may help predict holistic writing quality and assess the relationship between vocabulary knowledge and use.

Since Laufer and Nation (1995) proposed the LFP, a number of studies have looked at lexical sophistication in learners' written text by adapting different frequency lists. However, a majority of these studies examined general [2] writing topics whereas the present thesis examines discipline-specific topics. Nevertheless, the studies have shown that the occurrence of specific words from different frequency lists relate to holistic writing scores or course grades (e.g., Li, 1997; Lemmouh, 2008; Morris & Cobb, 2004). It should also be noted that the task and topic have been shown to have an impact on the kind of vocabulary that is used (Arnaud, 1992; Read, 2000). Li (1997) measured the LFP for non-native writers and the relationship between holistic scores for timed compositions. The results indicated that the LFP correlated with holistic scores assigned by teachers. Furthermore, LFP was capable of differentiating between high- and low-proficiency essays, but not mid-proficiency texts. Goodfellow, Lamy, and Jones (2002) used LFP to examine 36 texts written by non-native writers with the purpose of using LFP as part of an automatic feedback tool for students. They found a weak negative correlation between 1K vocabulary and essay scores (r = -.35), as well as a moderate positive correlation between the 2K list words used and essay scores (r = .42). Taken together, these studies indicate there is a link between LFP and writing scores. It seems that as writing scores increase, the use of 1K words decreases while the use of 2K words increases. In other words, as writing scores increase, more

[2] 'General' is the term used throughout this thesis to refer to non-discipline specific writing assessments.

sophisticated vocabulary is used. However, these studied only examined non-native language speakers' writing.

In a recent study, Higginbotham and Reid (2019) investigated the LFP of essays averaging 2000 words written by advanced non-native speakers of English ($N = 472$) from the faculties of: humanities and social science (n = 307), law (n = 116), and science and engineering (n = 49). The essays topics were decided by the faculties and pertained to a common theme about globalization. The findings indicated there was a statistically significant negative correlation between the use of 1K list vocabulary and essay scores ($r = -.27$), while there was a non-significant negative correlation ($r = -.11$) between the use of 2K list vocabulary and essay scores for the entire sample ($N = 472$). Of relevance to the present study, is the science and engineering group because this study will be examining the LFP of engineering student writing. Within the science and engineering essays, there was a significant negative correlation between use of the 1K list vocabulary and essay scores ($r = -.37$) while there was a lack of correlation ($r = -0.01$) for the 2K list. These findings are somewhat similar to previous research. Goodfellow et al. (2002) also found a significant negative small correlation for use of the 1K list and writing scores, but a moderately positive correlation for the use of the 2K list words and writing scores.

A possible reason for the differences between the two previously mentioned studies is that Goodfellow et al. (2002) examined shorter texts written about a general topic (the life of Quebecois firefighters), while Higginbotham and Reid (2019) examined longer texts written about globalization customized according to the academic discipline. It could be suggested that general topics elicit high-frequency vocabulary in writing because general topics pertain to everyday life. On the other hand, more discipline-specific topics require more specific vocabulary which is associated with low-frequency vocabulary i.e., sophisticated vocabulary because they occur in limited contexts.

Higginbotham and Reid's (2019) study is important as it addressed the shortcoming of previous LFP studies that only examined short texts of approximately 300 words (Laufer & Nation, 1995; Morris & Cobb, 2002). Higginbotham and Reid argued it was important that the LFP be used on longer length texts because shorter texts do not give students enough opportunity to demonstrate their knowledge. Furthermore, there may be differences between the LFP and writing scores depending on the disciplines and task topics, as seen above when comparing the use of 2K lists and writing scores.

In addition to the LFP having a relationship with holistic writing scores, LFP has also been shown to differentiate between writing proficiency levels (Morris & Cobb, 2004; Muncie 2002). Morris and Cobb (2004) used LFP to examine 300-word writing samples from 122 TESL trainees with varying language backgrounds whose written English proficiency skills were considered good to highly skilled. A small but statistically significant negative correlation between 1K word usage and course grades ($r = -.34$, $p < 0.01$) was discovered. However, the researchers did not examine the use of words beyond the 1K word list as previous studies had. This should have been considered because, as Laufer and Nation (1995) suggested when exploring the LFP for more advanced learners, a "finer distinction should be made above the basic vocabulary [1K and 2K lists]" (p. 311). In Morris and Cobb's study, the participants were considered to be good to highly skilled, and therefore more likely to produce more advanced vocabulary beyond the 1K list, i.e., infrequent vocabulary.

A further analysis by Morris and Cobb (2004) found that native-speaker TESL trainees had higher grades and used fewer high-frequency words as well as more academic words rather than low-frequency words. In comparison to non-native speakers who obtained lower grades and, used more high-frequency and fewer academic vocabulary. This suggests that less successful writers may use more high-frequency (or less sophisticated) words in their writing, while successful writers do the opposite. Furthermore, LFP can be correlated with writing scores. Morris and Cobb recommended that LFP may be a good predictor of academic performance and could differentiate between proficiency levels, if used in combination with more traditional forms of entrance assessment.

Muncie (2002) examined the written redrafts on the topic of friendship written by 20 Japanese non-native intermediate writers in an English composition course. The results indicated there was a statistically significant difference for the LFP between the redrafts, with increases in infrequent words and decreases in frequent words below the 1K word list. Muncie used the redrafting process to measure proficiency with the assumption that the first draft would be less lexically sophisticated as students were just starting the course and the final draft becoming more lexically sophisticated at the end of the course. Although Muncie found differences between the drafts in terms of the LFP, the LFP was not correlated to writing scores. Nevertheless, Muncie's study suggests that the LFP may be capable of measuring differences in writing proficiency.

Taken together, the results of these studies indicate that the LFP may be a good indicator of characterizing writing proficiency and there may be a relationship between vocabulary frequency and holistic writing scores. However, the above-mentioned studies tended to focus on non-native English speakers and the writing topics examined using LFP were typically general which did not require discipline-specific knowledge. This is an important gap in the literature as vocabulary use varies according to task and topic (Arnaud, 1992; Read, 2000). Furthermore, diagnostic discipline-specific literacies in conjunction with academic support are a key factor for retention and program completion (Fox, 2005; Meyer & Land, 2003). The present study thus addresses these gaps by examining a mix of native- and non-native speakers' writing from an English for Specific Purposes diagnostic test for discipline-specific literacies—specifically for engineering. This study also seeks to examine the relationship between LFP and holistic writing scores by characterizing unsuccessful (at-risk) and successful (not-at-risk) texts based on studies that have compared writing compositions of different proficiency levels (Laufer & Nation 1995; Muncie, 2002). This is considering how LFP is widely accepted as an indicator of writing quality and an indicator of lexical sophistication (Higginbotham & Reid, 2014; Lemmouh, 2008).

2.4 Lexical Frequency Profile and Lexical Stretch (Douglas, 2010, 2013, 2015)

Traditional lexical profiles, such as the type developed by Laufer and Nation (1995), as well as those used by other researchers include the proportion of the most common word lists: 1K, 2K, University Word List (UWL) or its successor, the Academic Word List (AWL), and Not in List words (e.g., Laufer & Nation, 1995; Morris & Cobb, 2004; Muncie 2002). Over time, researchers have added different measures or word lists in their approaches to lexical profiling or measuring lexical sophistication to provide a more in-depth understanding of the vocabulary characteristics of student writing (Douglas, 2015). For example, Douglas (2015) examined the relationship between LFP and test scores for the spoken and written portions of the Canadian English Language Proficiency Index Program (CELPIP)-General Test. A total of 200 samples were compiled for both the spoken (211,602 tokens) and written (70,745 tokens) texts. Because of the nature of the CELPIP-General test, there was little information provided about the topic of the written task. Instead, it was stated that students completed two writing tasks in which they were given 60 minutes to write an email or respond to an opinion survey.

Douglas (2015) used VocabProfiler (Cobb, n.d.) to calculate eight lexical measures: number of tokens and types, proportion percentages of High-Frequency Vocabulary (HFV), Mid-Frequency Vocabulary (MFV), and Low-Frequency Vocabulary (LFV). Douglas (2015) categorized the 1K and 2K lists into one category labeled ‘High-Frequency Vocabulary (HFV)’ based on Horst’s (2013) finding that a learner knowing the 2,000 most frequent vocabulary can provide coverage of over 80% of the words encountered. The 3,000-10,000-word families were categorized as ‘Mid-Frequency Vocabulary (MFV)’ based on Schmitt and Schmitt’s (2014) definition of mid-frequency vocabulary. Then according to Douglas (2010), the 11,000 to 25,000-word families and off-list words were considered ‘Low-Frequency Vocabulary (LFV)’. Previous LFP literature considered words in the 1K word list as high-frequency while those in the 2K word list were considered less frequent (Goodfellow et al., 2002; Laufer & Nation, 1995; Muncie, 2002). Also, words beyond the 2K list were considered Not in List and could be words anywhere from the 3K list to the 25K list and beyond. Douglas’ (2015) distinction between HFV, MFV, and LFV allows for a greater understanding of the kinds of vocabulary used by writers. Furthermore, it addresses Laufer and Nation’s (1995) recommendation of providing a finer distinction beyond basic vocabulary (1K and 2K lists) for more advanced writers.

Douglas (2015) included three additional measures, referred to as ‘Lexical Stretch’ (Douglas, 2010, 2013, 2015):

1) the lowest frequency list used by the writer to cover 98% of the text. This is calculated by cumulatively adding the percentage of each frequency list to reach 98% coverage of a text, then the lowest frequency list is recorded;
2) the top frequency list used by the writer; and
3) the total number of frequency lists used in a text.

Douglas claimed these measures can provide a measurement for the gap of lexical knowledge and a better understanding of sophisticated vocabulary used by different writing proficiencies. For example, a text may have used words from the 1K and 2K lists (HFV) to reach 92% coverage of the text, words from MFV (3K and 5K lists), and LFV (14K and 15K lists) to provide 5% and 1%, respectively, to cover 98% of the text. The top frequency list would thus be 15K and the number of lists used was six (1K, 2K, 3K, 5K, 14K, and 15K).

In regard to 98% coverage, Douglas does not provide an explanation for the number 98% in regard to production. Previous literature has identified the three prominent lexical coverage figures: 95%, 98%, and 100%. These numbers typically refer to comprehension or word learning—more specifically, they refer to the percentage of coverage cumulatively provided by word lists (1K, 2K, 3K, etc.) from the BNC-COCA 25 in order to cover 95%, 98%, and 100%. The idea of lexical coverage follows the logic that: the larger the proportion of highly frequent words in a text, the easier the text will be to comprehend. The coverage figure of 95% has been cited as providing adequate or minimal comprehension (Laufer & Ravenhorst-Kalovski, 2010), while 98% provides ideal coverage (Laufer & Ravenhorst-Kalovski, 2010; Nation 2006). However, different lexical coverage numbers have been reported for different types of texts such as spoken discourse (e.g., Nation, 2006), test items (e.g., Webb & Paribakht, 2015), television programs (e.g., Rodgers & Webb, 2011), and academic essays (Douglas, 2013). In terms of LFP, 98% lexical coverage may not be as applicable to the productive aspect of writing.

The results from Douglas' (2015) indicated a significant strong positive correlation between CELPIP-General Writing Test levels and an increase in the: number of tokens (r = .54), diversity of words (r = .77), and percentage of mid-frequency vocabulary used (r = .73). In comparison, for HFV there was a strong negative correlation (r = -.73) while there was a small positive correlation for LFV (r = .28). In addition, increases in Lexical Stretch measures (98% coverage, top list accessed, and number of lists accessed) were found to increase with test scores. Douglas thus suggested that the use of LFV may not be the only differentiating factor to distinguish higher proficiency writers. Instead, MFV may also have an impact on writing proficiency test levels (Schmitt & Schmitt, 2014).

Douglas' (2015) finding that the HFV correlates negatively with writing scores is consistent with previous research (e.g., Goodfellow et al., 2002; Laufer & Nation, 1995, Morris

& Cobb, 2004). It is important to note, however, that Douglas' (2015) study is difficult to compare with previous LFP research because previous studies have typically measured the 1K and 2K lists as separate and words Not in the Lists as sophisticated vocabulary. Nevertheless, the general idea of HFV negatively correlating with proficiency appears to hold true. Douglas (2015) research is important for several reasons as it provides additional information as to which word lists writers use and their influence on writing proficiency. Furthermore, previous research had identified anything past the 2K word list as sophisticated. Douglas shed light on which words beyond the 2K word lists have a stronger relationship with writing levels. Lastly, while previous studies generally focused on LFP for instructional writing class tasks, Douglas investigated LFP for a standardized language proficiency test.

While Douglas correlated LFP to CELPIP-General Writing Test levels, a correlation with exact writing scores was not obtained. The present study will replicate and build off of Douglas (2015) by examining the relationship between exact writing scores, LFP, and Lexical Stretch. In addition, this study will address the gap in LFP studies that have not applied LFP to an ESP-written diagnostic assessment using holistic scores to further understand the lexical profiles of at-risk and not-at-risk writers. This includes amalgamating Laufer and Nation's (1995) overarching concept of LFP and Douglas' (2010, 2013, 2015) Lexical Stretch measures. The present study will use the frequency definitions outlined by Douglas (2015) and compare the results because both studies examine written assessments outside of the classroom. This is in response to the advancement of words lists since Laufer and Nation (1995) have created the LFP. Moreover, a refined definition for high-, mid-, and low-frequency vocabulary allows for comparability of a writing tests (SAFE vs. CELPIP) and an updated understanding of what kind of vocabulary is being used in at-risk and not-at-risk writing.

2.5 Tokens, Type, and Type-Token Ratio (TTR)

Commonly observed measures in lexical analysis research have also included counts of tokens, types and type-token ratio (TTR). ‘Tokens’ are the total number of running words in a text. Previous research has found that the length of composition is related to writing scores (e.g., Douglas 2015; McNamara et al., 2010; McNamara et al., 2013; Reid, 1990). Douglas (2015) found a significant moderate positive correlation for the number of tokens and proficiency levels for the CELPIP-General Test. Banerjee, Franceschina, and Smith (2007) found for the writing module of the International English Language Test System (IELTS), test takers who obtained low scores used fewer tokens and had less diversity of words (also known as ‘types’) in comparison to students who scored higher on IELTS. Frase, Faletti, Ginther, and Grant (1998) found that an increase of tokens in writing was associated with increased test scores on the TOEFL Test of Written English. They found a range of correlations (between $r = .65$ to $r = .82$) for writers of varying language backgrounds. Similarly, Kaplan, Wolff, Burstein, Lu, Rock, and Kaplan (1998) indicated that there was a correlation of $r = .80$ between text length and writing scores. This suggests that the length of texts, or number of tokens may be an important contributor to test scores.

‘Types’ are the total number of different words in a text according to the word families in the BNC-COCA 25 lists (Nation, 2001) which includes the stem of a word and, the inflectional and derivational forms. A closely related measure is type-token ratio (TTR) in which the total number of types in a text is divided by the number of tokens in the text. The total number of types is calculated according the inflectional and derivational forms within the word families in the BNC-COCA lists. Previous research has suggested that higher proficiency writers use more variation or diversity of words as indicated by the TTR (Cumming et al., 2005, 2006; Engber, 1995; McNamara, Crossley, & McCarthy, 2010). McNamara et al. (2010) found that high lexical variation in student writing had a significant small positive correlation with holistic writing scores ($r = .20, p < .01$) with 4% of the variance being explained by TTR. Engber (1995) found moderately positive correlations between holistic writing scores and TTR for writing which included errors ($r = .45, p < .01$), errors that were fixed ($r = .57, p < .01$). However, it is important to note that TTR has been found to be unstable for shorter text lengths (Laufer & Nation, 1995; McCarthy & Jarvis, 2010).

Although the above-mentioned traditional measures have been used in written assessment contexts, there is a lack of studies that have applied these measures to an ESP-diagnostic assessment on a discipline-specific topic such as engineering. This study addresses this gap.

2.6 Specialized Vocabulary

Another method used in the literature to characterize student writing is the study of academic and discipline-specific language (e.g., Biber, 2009; Coxhead & Nation, 2001). Knowledge of academic vocabulary is vital for writing appropriately and effectively for a specific subject area (Corson, 1997). A component of academic writing is the ability to demonstrate knowledge of the writing community through use of vocabulary (Bruce, 2008; Craswell & Poore, 2012; Paquot 2010). Vocabulary can be divided into: general (high frequency), academic (i.e., university), technical (hereby referred to as 'discipline-specific'), and other (low-frequency) (Nation, 2001).

General vocabulary are words that are used and encountered frequently. Such words can be found in West's (1953) General Service List (GSL) of the most widely used 2,000-word families in English, or Nation's (2012) 1K-2K frequency lists, which cover approximately 80-90% of most texts.

Academic vocabulary is frequent and specific in academic discourse which may cover 8-10% of academic texts. Examples include the previously mentioned list, the UWL. The most prevalent and frequently used academic list is the Academic Word List (AWL; Coxhead, 2000). Developed by Coxhead (2000), the AWL comprises of 570-word families derived from a corpus of 3.5 million running words of academic texts from the four sub-corpora: arts, commerce, law, and science, which are further divided into seven subject areas. The AWL occurs in roughly 10% of academic texts. However, Gardner and Davies (2014), who developed the Academic Vocabulary List (AVL), found that the AWL covered 7.2% of the words in the COCA's academic texts, while their top 570-word families in their list provided almost twice the amount of coverage at 14 %. The AVL has gained popularity and is being used by academic vocabulary researchers (e.g., Csomay & Prades, 2018). The AVL comes from a 120-million-word corpus and is claimed to better depict academic vocabulary than the AWL because of the size of the list, which contains 3,000 core academic words (as lemmas). It is important to note that the AVL is not as pedagogically useful for students considering how it includes 3,000 words for students to learn, in contrast to the AWL's 570 words.

Nation (2013) reclassified the above vocabulary categories to include mid-frequency vocabulary and, combined academic and discipline-specific vocabulary under a new category, *specialized vocabulary*. Nation's (2013) vocabulary terminology of specialized vocabulary will be used in this thesis because it represents the combination of academic and discipline-specific vocabulary.

2.6.1 Academic Vocabulary

The use of academic vocabulary in student writing has been investigated by several researchers and is measured as part of LFP (e.g., Csomay & Prades, 2018; Douglas, 2013; Goodfellow et al., 2002; Laufer & Nation 1995; Morris & Cobb, 2004; Nadarjan, 2011). Using LFP, Laufer and Nation (1995) found that as language proficiency increased, the use of academic words also increased. The lowest proficiency group used on average 4% of academic words in their writing while the highest proficiency group used 10%. Douglas (2013) measured the occurrence of the AWL in 120 successful first-year university students' writing for the Effective Writing Test. Douglas likewise found that the AWL occurred 6.74% on average in successful student writing.

Morris and Cobb (2004), meanwhile, found there was a statistically significant correlation between academic word usage and course grades (r = .37) for 122 TESL trainees. They examined statistically significant differences between native writers and non-native writers use of academic vocabulary and concluded that native writers obtained better grades and used over 5% of AWL words in their writing. Students who had dropped out of the program had an average AWL usage of under 5%. Similarly, Higginbotham and Reid (2019) investigated the relationship between the AWL and essay scores through an analysis of 472 essays averaging 2000 words from three faculties: humanities and social science, law, and science and engineering. The results showed a significant positive relationship between the use of AWL words and essays scores (r = .21). For each discipline, the correlations were as listed: humanities and social science, r = .17; law, r = .22; science and engineering, r = .26. Interestingly, the highest correlation out of the three faculties was for the science and engineering writing. This is relevant to the present study which will also be examining the correlation between the use of academic vocabulary and writing scores but only for engineering writing and shorter text lengths.

On the other hand, some researchers found no correlation between the use of academic vocabulary and writing scores. Nadarajan (2011) investigated the relationship between academic word use and holistic scores for 387 essays written by 129 native and non-native English-speaking students enrolled in an English Composition class. Three essays about general topics containing around 100-130 words were analyzed for each student. For native speakers' essays, one out of three essays demonstrated a statically significant correlation between AWL word usage and holistic scores (*r* =.30), although not for the remaining two essays. For non-native speakers' essays, there was no statically significant relationship between the AWL and holistic scores. However, essays that scored high overall typically had a range of 4.98% to 5.41% of AWL words; students who obtained higher scores were generally found to have high proportions of AWL words in their writing. In comparison, students who obtained lower scores used fewer AWL words. Nadarajan's finding that higher scoring texts used approximately 5% of AWL words is similar to Morris and Cobb's (2004) findings. Both studies used general topics and texts that were approximately 300 words or less. Furthermore, texts that contained less than 5% of AWL words were considered at-risk of failing. Laufer and Nation (1995) similarly found that their lowest proficiency group used 4% of academic words for similar text lengths.

In contrast, Csomay and Prades (2018) found there was no statistically significant correlation between the percentage of academic vocabulary used and holistic essay scores (*r* =. 071); for essays averaging 1000 words from a variety of text types such as response papers, comparative analysis, exploratory synthesis, argument synthesis, rhetorical analysis, and editorials. In addition, it was found that low writing proficiency students (*M* = 10.25) used on average more academic vocabulary than high writing proficiency students (*M* = 8.60). This is surprising given that academic vocabulary has been found to occur more frequently in higher proficiency writing, as noted in the studies above. Similarly, Goodfellow et al. (2002) analyzed 36 French second language writing texts (averaging 385 words) about the life of Quebecois firefighters. The results suggested there was no relationship between academic word use and grades obtained.

The above research found there was no relationship between academic vocabulary and writing proficiency. These results contradict Higginbotham and Reid (2019) who found positive strong correlations for the AWL in three different specific disciplines. However, Higginbotham and Reid analyzed texts averaging 2000 words, which is double the number of words in Csomay and Prades' (2018) texts and more than five times the number of words in Goodfellow et al.'s (2002) texts. Goodfellow et al.'s (2002) results may be because general topics do not elicit or require academic vocabulary, unlike texts written for different specific disciplines.

In sum, research findings regarding productive academic vocabulary used in writing is mixed in terms of academic vocabulary's impact on overall writing quality. There is not enough consistent evidence to come to a reliable conclusion. Most importantly, most of the research in the area has examined second language learners rather than holistic proficiency. In this thesis, the investigation of the relationship between students' use of academic vocabulary and writing scores will be examined for a mixed sample of native and non-native writers but focuses on the differentiation between at-risk and not-risk writing.

2.6.2 Discipline Specific Vocabulary

Discipline-specific word lists (also known as 'technical' word lists) include words that are infrequent and used in specific disciplines; thus, they vary according to context. Researchers have made numerous discipline-specific vocabulary lists using the high frequency of occurrence in a specific discipline area. The aim of these lists is to provide learners with vocabulary that would be useful based on the likelihood of encountering the vocabulary in their studies. Two examples of discipline-specific wordlists are: the Medical Academic Word List (MAWL) with 623-word families developed by Wang, Liang, and Ge's (2008) and Yang's (2015) Nursing Academic Word List (NAWL) containing 676-word families in the nursing discipline.

Of relevance to the present study is Hsu's English Engineering Word List (EEWL; 2014a). The EEWL contains 729 of the most frequently occurring word families derived from a corpus of 4.57 million running words from engineering textbooks from 20 engineering sub-disciplines. The EEWL occurs in 14.3 % of engineering texts. A similar list is Ward's (2009)

Basic Engineering List (BEL) which contains 299 words derived from a 1 million-token corpus of foundation engineering textbooks from a broad range of topics. The EEWL will be used for the present study as it originates from a larger and more representative corpus.

Research has primarily focused on academic vocabulary in student writing (e.g., Csomay & Prades, 2018; Morris & Cobb, 2004), but not the use of discipline-specific words in student writing. One case-study was found by Lessard-Clouston (2012) who compared the use of academic and discipline-specific words in the writing of native and non-native writers enrolled in a theology graduate program. Discipline-specific words were operationalized as the percentage of theology words in writing. The results showed no difference in the use of academic and discipline-specific vocabulary between native and non-native writers. Notably, this study did not differentiate between writing proficiency levels nor was there an investigation of the relationship between the use of discipline-specific words and holistic scores. In addition, the results are difficult to generalize because of the small sample size and specific topic of theology.

To address the lack of studies that examined the productive use of discipline-specific words in student writing, the present study examines discipline-specific word use in a large engineering student writing corpus and the relationship with holistic writing scores, as seen with academic word use studies. In addition, this study will observe the productive use of the ever-growing number of discipline-specific wordlists in the literature. As a prominent second language researcher, David Wood, once said about the overabundance of corpus derived frequency lists, "*Another* list!" (D. Wood, personal communication, September 2018). Thus, demonstrating the lack of studies that examine the productive use of discipline-specific lists in student writing. This is important considering how some scholars (e.g., Hyland & Tse, 2007) have questioned the usefulness of general academic vocabulary lists using corpus derived findings because vocabulary use varies according to discipline with respect to frequency and meaning. This is supported by Fox, Hargety, and Artemeva (2016) who point out the need for academic discipline-specific literacies, which includes content and vocabulary.

In short, in order to succeed in university, students must develop academic vocabulary (Shanahan & Shanahan, 2008); the same could be said for discipline-specific vocabulary. In order to acknowledge this gap, the present study will investigate and compare the use of academic and engineering vocabulary to better understand the kind of vocabulary that characterizes engineering writing. SAFE is an ESP-based writing task focused on engineering; therefore, it is expected that all writers will use academic vocabulary and discipline-specific vocabulary (i.e., engineering vocabulary). This is important because Hyland (2004) proposed the assessment of academic writing should consider the text based on the specific context, register, and discourse community. This suggests that appropriate vocabulary should be used according to the type of text.

2.6.3 Academic Formulaic Language

When examining the lexical characteristics of writing, researchers have recently begun to shift their investigation from solely single words to include multi-word sequences, hereby referred to as *formulaic language* (e.g., Crossley et al., 2012; Kyle et al., 2017). Formulaic language can be an umbrella term used to characterize fixed and semifixed expressions of different lengths in spoken and written discourse (Wood, 2015). The present study will follow the famous definition provided by a prominent formulaic language researcher, Alison Wray (2002):

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p. 9)

Formulaic language use has been noted in the literature for supporting faster processing of discourse because of the whole storage of the word sequences that are ready to be retrieved cognitively (Nattinger & DeCarrico, 1992; Wray, 2002). Therefore, resulting in the production of longer and more accurate sequences of language in comparison to a word-by-word composition. In the context of writing, this may result in the increased production speed of sentences and accuracy of texts.

Wray and Perkins (2000, p. 3) observed that the formulaic language phenomena had been labeled with over 40 different terms. These terms include examples such as: 'lexical bundles', 'multi-word units', 'idioms', 'collocations', 'n-grams'. Such differences in the terminology demonstrates the variability of what word sequences are considered and identified as formulaic language.

Identification of formulaic language has typically followed a *frequency-based* approach in which researchers set a minimum cut-off for the frequency of occurrence as well as a range (i.e., how many individual texts the formulaic sequence occurred in). Using the set criteria, a representative corpus is analyzed for the occurrence of formulaic sequences quantitatively. Analytic software such as AncConc (Anthony, 2018) can be used to aid in the process of finding formulaic language in text(s) by examining the concordances. However, the use of the frequency-based approach alone has been questioned by some researchers (e.g., Simpson-Vlach & Ellis, 2010) because such an approach results in long lists of incomplete or meaningless recurrent word sequences which may not actually be formulaic but just occur frequently. These long lists are then collapsed using the researcher's intuition as the criteria of what list of sequences should be considered formulaic or not, but this method is "methodologically tricky and open to claims of subjectivity" (Simpson-Vlach & Ellis, 2010, p. 490). The frequency-based approach is used often because it is a simple method for identifying frequently occurring word sequences.

To address the methodological flaws of the commonly used frequency-based approach, Simpson-Vlach and Ellis (2010) complied the Academic Formulas List (AFL). The frequency-based approach was used as a starting point. Then, an innovative combination of quantitative and qualitative criteria, corpus statistics and linguistic analyses, psycholinguistic processing metrics, and instructor insights were used to determine pedagogically useful formulas. The lists are ranked according to the formula teaching worth (FTW) which is informed by a statistical index of word cohesion — mutual information (MI) and frequency, the combination of these two provide instructors with an idea for prioritizing formulaic sequences to use and teach.

The resulting AFL contains 607 recurrent word combinations from a corpus of 2.1 million words each for academic speech and academic writing. The 607 academic formulas compose of three separate lists: 1) Written AFL, 2) Spoken AFL, and 3) Core AFL. The Written and Spoken AFL each separately contain 200 academic formulas that are commonly found in the respective academic registers. The Core AFL contains 207 formulas which are common in either academic spoken or written language. The AFL items were also categorized according to Biber, Conrad, and Cortes' (2004) functional taxonomy. Of interest to the present study, is the Written AFL which was derived from an academic writing corpus containing sub-corpora from the subjects: humanities and arts, social sciences, natural sciences/ medicine, and technology and engineering. The Written AFL will be used for the present study because the study examines academic student writing.

Use of formulaic language has been linked to lexical sophistication for writing (Crossley et al., 2012 as cited in Kyle & Crossley, 2015). However, Kyle and Crossley (2015) had investigated the AFL use in student writing and found that the AFL (and AWL) sub-list indices were “not important indicators of lexical sophistication” (p. 775). It was speculated that the respective measurements may be predictive of sophisticated texts for longer texts, as many of the texts in their study were under 100 words. Additionally, Kyle and Crossley (2015) analyzed free writing which may not have required the use of academic language such as the AWL and AFL.

Previous research has found that the use of formulaic language in student writing has a connection with writing proficiency. Although much of the research pertaining to formulaic language use in student writing has focused on comparing native and non-native speakers, relatively few studies have examined the use of formulas among proficiency levels. Ädel and Erman (2012) compared a large corpus of writing from advanced Swedish language learners of English to native English speakers writing at a British university. It was found that native speakers used a greater range of formulaic sequences in their writing. Chen and Baker (2010) evaluated formulaic language in native expert, native student, and non-native student writing. The findings indicated that with increased writing proficiency, there was an increase in the use of formulaic sequences in academic writing. Similarly, when Li and Volkov (2018) investigated the written texts of student writing for the CELPIP-General Writing task, they found higher proficiency levels produced more formulaic sequences in their writing. Hyland (2008, 2012) notes that formulaic language can aid in the composition of logical and coherent texts, the lack or misuse of formulaic language may indicate writing as low-proficiency or non-native.

However, some studies have found the opposite. Appel and Wood (2016) examined a corpus of student writing from the Canadian Academic English Language Assessment (CAEL) to compare high- and low-proficiency non-native speaker writing. The findings indicated that lower-level students tended to use more formulaic language in their writing. Similarly, Staples, Egbert, Biber, and McClair (2013) investigated a corpus of writing from the Test of English as a Foreign Language (TOEFL) within three proficiency levels, and similar to Appel and Wood

(2016), the findings indicated that lower proficiency writers tended to use more formulaic language while higher proficiency writers used less formulaic language. The authors of these studies suggested that lower proficiency non-native writers may rely on formulaic language to support their limited vocabulary and store formulaic sequences. This method was compared to the “teddy-bear principle” (Hasselgren, 1994) in which writers hold on to singular or multiple word sequences they are comfortable using.

The literature on formulaic language in student writing tends to focus on comparing native and non-native speakers’ use of formulaic language in writing. The present study instead focuses on comparing formulaic language use in the writing proficiency levels of at-risk and not-at-risk writing to shed light on the lack of research focusing on formulaic language and writing proficiency as indicated by Wood (2015). Furthermore, as Flowerdew (2012) points out, previous formulaic language studies rely on different types of corpora as well as use various software for analysis resulting in a controversial area of research. Hence, it is not surprising that previous formulaic language findings vary.

The present study does not derive a list of frequently occurring formulas from a corpus as previous research has. Rather, the present study investigates the occurrence of a corpus derived representative list of academic formulas in at-risk and not-at-risk writing. The present study will quantify and compare the occurrence of the AFL within at-risk and not-at-risk student writing. This is important because David Wood (2015), a prominent formulaic language researcher, emphasizes that formulaic language is a “foundation of successful academic writing skills because they comprise the basic elements of academic discourse and are specific to particular disciplines, registers, and genres” (p. 103). Additionally, the ability to use vocabulary accurately and appropriately in writing demonstrates the writer belongs to a particular community, in this case, the use of formulaic sequences in academic writing.

2.7 Tool for the Automatic Analysis of Lexical Sophistication

To address the limitations of previous lexical analysis software and measures of lexical sophistication, Kyle and Crossley (2015) developed the first version of the Tool for the Automatic Analysis of Lexical Sophistication (TAALES), a freely available and downloadable application that measures 135 indices of lexical sophistication. These novel indices are related to word frequency, range, bigram and trigram frequency (*n*-grams), academic language, and psycholinguistic word information (see Kyle & Crossley, 2015 for more information about the indices). Some of the main purposes of TAALES is to aid in the measurement of lexical sophistication, text analysis, and language assessment.

Within the same study, Kyle and Crossley (2015) validated TAALES by applying the indices to explain the variance in human judgements of written and spoken proficiency. The writing samples analyzed came from a corpus compiled by Crossley, Salsbury, McNamara, and Jarvis (2011). This corpus included texts from 180 non-native English writers grouped according to TOEFL scores from beginning (n = 60), medium (n = 60), and high (n = 60) levels. These were combined with undergraduate native speaker writing compositions (n = 60), resulting in a corpus of 240 writing samples. Using a stepwise multiple regression, the lexical sophistication indices explained 47.5% of the variance in holistic scores for written proficiency. This suggests that lexical sophistication may contribute to holistic writing scores. The results indicated that the word frequency index was not a predictor variable for scores (i.e., more advanced writers do not necessarily produce more infrequent words). Rather, other indices may play a larger role. These findings contradict previous research (e.g. Crossley et al., 2011; Laufer & Nation, 1995) which found frequency of vocabulary to be related to holistic scores.

Instead, the largest contributing index found by Kyle and Crossley (2015) was word range (25.9%), which was negatively correlated to writing scores. Kyle and Crossley suggested that words not used in a wide variety of contexts are considered more sophisticated than words that are used in a variety of contexts. In other words, rare or context specific words are considered sophisticated. Addtinally, *n*-grams correlated strongly and explained 22.3% of the variance in holistic scores. Hence, suggesting that bigrams and trigrams frequency may be related to lexical sophistication (e.g., Crossley et al., 2012). This is related to the concept of formulaic language as discussed earlier.

In 2017, the successor for TAALES called TAALES 2.0 was expanded by Kyle, Crossley, and Berger (2017) to include an additional 316 new lexical sophistication indices. In order to validate TAALES 2.0, the indices were used to model and predict holistic scores and word choice scores using a multiple regression analysis. The same previously used corpus collected by Crossley et al. (2011), which contained 240 writing sample and holistic scores, was used for analysis. The multiple regression analysis results indicated that TAALES 2.0 indices could explain 58% of the variance in writing scores. The indices that contributed the most were related to *n*-grams and word range.

Kyle et al. (2017) found that texts which included more bigrams and trigrams in addition to a higher percentage of frequent trigrams were associated with higher writing proficiency scores. The second largest factor was word range. This indicates that texts with register specific words were associated with higher writing scores. Thus, suggesting that register specific words may be a contributing factor to holistic writing scores. These findings are consistent with Kyle and Crossley (2015) who found that TAALES indices could explain 47.5% of the variance in holistic scores and suggested that *n*-grams and word range contribute the most. A surprising result Kyle et al. (2017) found was that word frequency only explained 3.6% of the variance in the writing scores. This finding suggests that frequency of vocabulary may not have as strong of a relationship with holistic scores as previous literature found (e.g., Crossley et al., 2011; Laufer

& Nation, 1995). Instead, other lexical sophistication indices may contribute to writing proficiency.

In a more recent study, Kim, Crossley, and Kyle (2018) reduced 424 lexical sophistication measures into 12 macro-components of lexical sophistication using a Principle Component Analysis. These 12 macro-components contained various micro-components of lexical sophistication and were used to predict human ratings for two written proficiency level corpora: 1) the Yonsei English Learner Corpus (YELC; Rhee & Jung, 2014), which is a corpus of 3,031 texts written by Korean high school graduates and; 2) the previously mentioned Crossley et al. (2011) corpus which includes 240 written samples. The results indicated that higher proficiency levels were associated with use of bigram and trigrams, use of advanced content words, and word range (i.e., words that are used in fewer contexts or have less range).

The results from the above-mentioned TAALES studies provide insight into lexical sophistication indices beyond the traditional lexical frequency indices mentioned. The literature further suggests that lexical sophistication is not based solely on frequency, but rather lexical sophistication is a multidimensional phenomenon. The general trend seems to be that word range and *n*-grams are the largest predictors of holistic writing proficiency.

The above studies typically used a variety of measures from TAALES and completed a regression based on the numerous indices in an attempt to predict holistic scores. The present study will also use TAALES 2.0 but will only use selected indices that pertain to the lexical characteristics examined in the present study. This is to help triangulate or support the findings for the use of academic singular words and formulaic language. The reason for this is because previous research using LFP or TAALES are mixed as to whether academic words (Kyle & Crossley, 2015) and formulaic language (Kyle et al., 2017; Kim et al., 2018) contribute to writing scores. Furthermore, no studies to date have used lexical sophistication indices from

TAALES to triangulate traditional lexical profiling findings.

2.6 Overview of the present study

This chapter has reviewed the literature concerned with the relationship between lexical sophistication and writing proficiency. First, academic writing and the importance of vocabulary were discussed. Section 2 explored the four lexical sophistication approaches to characterize the writing profiles in the present research. First, the literature addressed Lexical Frequency Profiling (Laufer & Nation, 1995) with a focus on the relationship between word frequency and holistic writing scores. Within this approach, Douglas' (2015) adaptation of the LFP and Lexical Stretch (Douglas (2010, 2013, 2015) was introduced. In addition, three commonly used measurements, tokens, types, and TTR in lexical analysis research were explained for their significance. Section 2.6 focused on the previous literature associated with the relationship between specialized (academic, discipline-specific, formulaic language) vocabulary used in a text and the link to writing proficiency. Finally, TAALES was introduced and reviewed for its relevance and findings in written corpora.

Lexical sophistication has evolved and continues to be redefined as a multidimensional phenomenon and is linked to writing proficiency. Various indices such as word frequency, tokens, types, academic vocabulary, discipline-specific vocabulary, formulaic language, and many more as identified by TAALES, have been found to contribute to the lexical characteristics of successful and unsuccessful writing. Empirical evidence has shown that writing context and topics also have an impact on these characteristic and as a result, the analytic and holistic writing scores.

Yet, it should be pointed out that, until now, most of the empirical research concerning lexical sophistication characteristics appears to focus on one of the mentioned lexical sophistication measures and their relationship to writing proficiency scores for non-native speakers in a classroom or general standardized assessment setting. Meanwhile, the context of a diagnostic assessment for a specific discipline such as engineering and the impact of vocabulary on holistic scores has yet to be explored. Furthermore, a distinction between successful and unsuccessful writing has not been made. This is important as awareness of the importance of discipline-specific literacies assessments is increasing due to an understanding of the value in tailoring assessment to suit the specific requirements of a given discipline. Furthermore, diagnostic assessments need to be investigated because they provide insight into how a student would perform in their field and if they are at-risk of not succeeding academically. In addition to the context, there is a serious disparity in the research of discipline-specific vocabulary use and relationship with holistic scores. Additionally, previous research has had varied conclusions made for academic vocabulary and formulaic language used in a text and the connection to holistic writing scores. Finally, a number of researchers have actually suggested that their lexical sophistication results be correlated with exact writing scores rather than general levels (Douglas, 2015; Kim et al., 2018).

This study attempts to fill these gaps by investigating engineering student writing from an ESP-diagnostic writing test using lexical sophistication measures and support from TAALES. Three key research questions (with sub-questions) are asked below to explore the lexical characteristics of at-risk and not-at-risk student writing from a discipline-specific task that requires students to use accurate engineering content and language. The research presented here attempts to establish the important relationship between vocabulary and holistic writing scores. In addition, the present research attempts to identify differences in the presence of these lexical characteristics in the writing of at-risk and not-at-risk students using holistic writing scores. The main research questions investigated are:

1. What is the relationship between holistic writing scores and analytic vocabulary scores on a diagnostic test?

2. What is the relationship between multiple lexical frequency profiling indices and holistic scores obtained for writing proficiency on a diagnostic test?

3. What is the relationship between holistic writing scores on a diagnostic test and specialized vocabulary use identified by:

   a. academic vocabulary?

   b. discipline-specific vocabulary?

   c. academic formulas?

The answers to these questions will inform writing pedagogy from a lexical perceptive. Furthermore, in the context of the SAFE test, the characteristics identified from the not-at-risk writing will help inform academic support for at-risk students. These findings can be generalized to less proficient writers in academia as well. Finally, there are assessment implications as such findings can inform assessment procedures by human raters and contribute to the concept of automatic assessment.

Chapter 3: Methodology

3.0 Introduction

The aim of this thesis is to conduct a lexical analysis to understand and compare the vocabulary characteristics of writing for students who are considered at-risk (i.e. in need of academic support) and not-at-risk (i.e. not in need of academic support) according to the SAFE (Self-Assessment for Engineers). In order to investigate the research questions, as listed in the previous chapter, a corpus was collected and analyzed. To start the investigation of the first research question, a correlation analysis was conducted to explore the relationship between holistic scores and analytic vocabulary scores. The reason for this is to observe the degree to which vocabulary usage contributes to the holistic assessment score. To answer the second research question, several lexical frequency profiling measures were compared to holistic scores for not- at-risk and at-risk writing. These measures included: tokens, types, type-token ratio, frequency of vocabulary, and Lexical Stretch (Douglas, 2010, 2013, 2015). Research question three concerned correlating and comparing the usage of general academic vocabulary, discipline-specific vocabulary, and academic formulaic language, respectively, to holistic scores for not-at-risk and at-risk writing. General academic and discipline-specific vocabulary were addressed using the lexical software tool, AntWordProfiler (2014). AntWordProfiler is a freely available lexical analysis software tool which allows for comparisons of word list(s) occurrences to the investigated corpora. Next, the occurrences of academic formulaic language were examined using AntConc 3.5.7 (Anthony, 2018). AntConc is a freely available corpus analysis toolkit for concordance and text analysis. Finally, TAALES 2.0 (Kyle et al., 2017) will be mentioned throughout this chapter in the relevant sections as support and triangulation of other lexical sophistication indices.

3.1 Materials

3.1.1 Self-Assessment For Engineers (SAFE) (Fox & Artemeva, 2017)

The Self-Assessment For Engineers (SAFE) (Fox & Artemeva, 2017) is an English for Specific Purposes (ESP) based diagnostic writing task that aims to inform first-year engineering students of their level of writing, enabling them to seek academic help if needed. Based on their holistic score, students are classified as 'at-risk' (holistic scores of 60 or less), 'somewhat at-risk' (between 61-79), or 'not-at-risk' (80 or above). Afterwards, only students who are considered at-risk are informed by email in which they are recommended to attend a personal academic support consultation and provided with advice (intervention) on how to improve their writing.

Fox and Artemeva (2017) developed SAFE in a multiphase process informed by applied linguistic scholars and professional engineers. This was in response to the findings by Fox, Haggerty, and Artemeva (2016) who observed a significant number of students drop out of their first-year undergraduate programs. This prompted the need for a diagnostic procedure to provide early intervention and academic support. The SAFE diagnostic procedure requires students to attend a first-year engineering lecture, watch an engineering innovation video, attend a lab to review the topic of the engineering task, then write responses to authentic engineering tasks. Previous literature suggests that discipline-specific rather than general academic literacies diagnoses are key factors that contribute to retention and program completion (Fox, 2005; Meyer & Land, 2003).

In a two-phase study, Fox and Artemeva (2017) developed an indigenously drawn ESP-based diagnostic writing task and rubric for SAFE. Phase 1 involved an analysis of a generic DELNA (Diagnostic English Language Needs Assessment) writing task and analytic rubric for its efficiency for identifying post-entry undergraduate engineering students who may need academic support. Using the results of Phase 1, Phase 2 produced a discipline-specific writing task which consisted of using the generic DELNA writing task to assess the developed task. Finally, the findings from the previous phases were combined to design and assess the practicality of the newly developed ESP-based writing task and analytic rubric.

Of interest for the current study is the Discipline-specific Graph Interpretation Task (DGIT). Students are required to interpret graphs displaying theoretical and actual data on an engineering innovation and to provide insight. The topic of innovation differs for each year of administration but are designed to be parallel tasks targeting the same writing skill. The topic of DGIT Version 1 involved car innovations, hereby referred to as 'DGIT-Car', while Version 2 concerned energy innovations, hereby referred to as 'DGIT-Energy'. [3]

3.1.2 SAFE Rubric

A discipline-specific ESP-based analytic rubric developed by Fox and Arteva (2017) is used to assess the completed SAFE tests. The indigenously drawn rubric was informed by a multiphase study for engineering writing expectations including accurate content, engineering disciplinary rhetorical expectations, language, and logic. Specifically, there are 15 separate analytic scores that divide evenly to target assessment of three main areas: Language, Rhetoric, and Logic. For each analytic criterion, a score between 0-9 could be obtained; 0 being the lowest score while 9 being the highest. A holistic score is calculated by adding the scores received for each of the 15 analytic scores. As previously mentioned, based on the holistic score, students are classified as 'at-risk' (holistic score of 60 or less), 'somewhat at-risk' (holistic score between 61-79), and 'not-at-risk' (holistic of 80 or above).

[3] Because SAFE is currently being used in the assessment of entering engineering students, it is considered proprietary and full details of the various tasks cannot be disclosed.

3.1.3 Human Ratings

Using the SAFE rubric, assessment of the completed SAFE tests is conducted by graduate students studying applied linguistics and discourse studies at a large university in Canada (See Section 3.1.2 for an explanation of rubric layout). These students referred to as “raters”, are trained during two separate sessions using test samples from previous years. During the training session, raters are introduced to the rubric with an explanation of the 15 analytic categories and expectations of student writing. In addition, raters are provided with examples of student writing that varied in final assessment scores and classifications of at-risk, somewhat at- risk, and not-at-risk provided by previous raters. Raters are first asked to rate the examples overall based on the formerly listed classifications. A group discussion is conducted for the reasoning of the overall decision with comparisons to the assessment classification provided by previous raters. Next, raters are asked to reconvene at a second training session with 10 marked writing samples using the SAFE rubric’s 15 analytic scores, provide a holistic score, and select one of the three classifications. During the second session, raters are paired and asked to switch their student writing samples to provide their own assessment of the writing sample for comparison. This results in two separately marked rubrics for each writing sample which are then compared and discussed for logic behind the responses indicated.

During the actual marking process, the ‘divide and conquer’ method is used in which raters read all the tests and divide them according to at-risk, somewhat at-risk, and not-at-risk based on the training standards. Then, all the at-risk tests are marked using the SAFE rubric. A selected number of somewhat at-risk and not-at-risk tests are marked using the rubric because the goal of SAFE is to identify students who are at-risk to provide academic support and intervention. In regard to the interrater reliability, a previous analysis of the responses provided by the raters indicated there was an exact agreement of 80% for interrater reliability.

Furthermore, there was 90% interrater reliability for classification agreement using the SAFE rubric (Fox & Artemeva, 2017).

3.1.3 Corpus Compilation

A total of 358 first-year engineering student SAFE tests were randomly selected and anonymized from the SAFE test bank collection of 2016 DGIT-Car (n = 179) and 2017 DGIT-Energy (n = 179). These samples represented roughly 15% of the test taker population for each year. Within each version, the tests were further divided according to the holistic test score obtained. Students who obtained 80 and above were categorized as 'not-at-risk' while students who obtained 60 and below were categorized as 'at-risk'. The decision to not include the 'somewhat at-risk' students in the analysis was made in order to see a clear distinction between the lexical characteristics of unsuccessful and successful writers. This resulted in two groups: 1) the not-at-risk group which included the Car not-at-risk (n = 85) and Energy not-at-risk (n = 85) and; 2) the at-risk group which included the Car at-risk (n = 94) and Energy at-risk (n = 94), (see Appendix A for a scatter plot of the distribution of holistic scores). However, five writing samples were removed from the at-risk group as they were either blank or off-topic. This resulted in the not-at-risk group (n = 170) and at-risk group (n = 183) which formed a corpus of 59,844 running words with a total of 353 writing samples. See Table 1 for the final details about the overall corpus.

Table 1

Number of total samples and tokens according to at-risk vs. not-at-risk group

Version	At-Risk	Tokens	Not-at-Risk	Tokens	Total Tokens
Car	90	11,756	85	16,313	28,069
Energy	93	14,740	85	17,041	31,781
Total	**183**	**26,496**	**170**	**33,354**	**59,844**

Each hand-written text sample was typed and saved as a separate Word (.doc) file and labeled according to the student case number assigned, task version, then the file was inserted in a categorization master file (i.e. at-risk or not-at-risk). For example, a file would be saved as Case#_SI in the At-risk Energy masterfile. After all the samples were typed, the SAFE corpus was cleaned, and the individual files were converted to TXT files in order for the lexical software tools to be able to analyze them. See Table 2 for the average tokens and types according to the at-risk and not-at-risk group.

Table 2

Average number of types and tokens according to at-risk and not-at-risk group

	At-Risk	**Not-At-Risk**
n	183	170
Tokens	144.4	196.2
Types	77.17	101.7

During the cleaning process, spelling errors, connected speech, and hyphenated words were corrected to match the spelling found in the BNC/COCA word lists. For example, in Case 100, 'assummed' was changed to 'assumed'. This was required in order for AntWordProfiler (Anthony, 2014) to accurately process the lexical coverage figures. Otherwise, 'assumed' would have appeared in the 'Not in List' categories or appear less frequent than Nation's (2012) 25,000 most frequent word families. Hence, resulting in the word frequency counts to be incorrect. In addition, contractions such as: *I'm, it's, can't, won't, that's, doesn't,* and *don't*, were changed to *I am, it is, cannot, will not, that is, does not,* and *do not,* respectively. Symbols such as °C were changed to 'Degrees Celsius' in TXT flies as they cannot be properly analyzed by AntWordProfiler. These changes occurring during the cleaning process accounted for approximately .10 % of the tokens in the study.

3.1.4 Distinctive Features of the SAFE Corpus

The present study uses a corpus of first-year engineering student writing from the graph interpretation task from two parallel test versions (Car and Energy). The SAFE corpus has distinctive features that allow for direct pedagogical recommendations to be made because the texts are from a diagnostic task which allows for meaningful insight into students' writing proficiency. One feature of the SAFE corpus is that it includes texts from specialized engineering topics. This is important because the corpora used in the previous literature typically included texts from non-specialized topics which the layman could answer (Douglas, 2015; Goodfellow, Lamy, Jones, 2002; Laufer & Nation, 1995). Another feature of the SAFE corpus is that the texts are responses to short answer questions used during assessments in later engineering exams and replicate real-word analysis writing. As stated previously, SAFE is informed by applied linguistic scholars and professional engineers. The SAFE tasks are meant to elicit authentic writing that engineering students will eventually be required to write in their future university courses and workplaces. Thus, adding to the pedagogical significance of the analysis of such texts.

3.2 Research Question 1: Holistic scores and vocabulary scores

3.2.1 Correlation Analysis: Analytic Vocabulary and Holistic Score

To answer the first research question of whether there is a relationship between analytic vocabulary scores and holistic test scores, a correlation analysis was conducted using the Pearson product moment correlation coefficient (Pearson *r*). The cut-off for significance of $p \leq .05$ will be used throughout this thesis. The holistic scores were obtained by adding each analytic criterion score together. A Microsoft Excel sheet (hereby referred to as mastersheet) was created to record the 15 analytic scores, holistic scores, and overall classification necessary for academic support from the SAFE rubric for each student. Of importance to the present study, was the analytic criteria evaluating students' appropriate vocabulary usage because this study seeks to examine the relationship between the vocabulary scores and overall (holistic) performance on the diagnostic test. After the analytic scores and holistic scores were recorded for 353 students, a correlation analysis was conducted between the two variables using JASP (Version 0.9.2.0; JASP Team, 2018) for the overall corpus. JASP is a freely available, user-friendly statistical analysis tool that works using CSV files. This analysis was conducted to examine the degree to which vocabulary usage contributes to holistic scores.

The effect size was calculated in order to determine the strength of the relationship between holistic scores and vocabulary scores (Creswell, 2012). Plonsky and Oswald's (2014) recommendations for selecting the cut-off for correlations in Applied Linguistic research were followed accordingly throughout this thesis: for a small effect $r = .25$, medium $r = .40$, large $r = .60$ (p. 889). These cut-offs will be used throughout the thesis.

3.2.2 Correlation Results

The first set of correlation analyses concerned examining the relationship between vocabulary scores and holistic scores. There was a statically significant large positive correlation between vocabulary scores ($M = 5.30$, $SD = 2.46$) and holistic scores ($M = 74.16$, $SD = 30.23$), $r(351) = .84$, $p < .001$. This indicates as students' vocabulary scores increased, holistic scores also increased. Thus, establishing the importance of the vocabulary score's contribution to the holistic test score. Table 3 and 4 provide the descriptive statistics for the vocabulary and holistic scores for each group examined, respectively.

Table 3

Descriptive statistics for analytic vocabulary scores according to at-risk and not-at-risk group

	Vocabulary Scores	
	At-Risk	**Not-At-Risk**
n	183	170
Mean	3.434	7.312
Median	3.000	8.000
Std. Deviation	1.682	1.333
Minimum	0.000	2.000
Maximum	8.000	9.000

Table 4

Descriptive statistics for holistic scores according to at-risk and not-at-risk group

	Holistic Scores	
	At-Risk	**Not-At-Risk**
n	183	170
Mean	47.44	102.9
Median	49.00	103.0
Std. Deviation	9.196	14.39
Minimum	15.00	80.00
Maximum	60.00	132.0

3.2.3 Independent *t*-test between at-risk and not-at-risk writing

After establishing that there was a relationship between vocabulary and holistic scores, a further investigation into whether there were differences in vocabulary and holistic scores between the at-risk and not-at-risk group was conducted. It was important to determine such differences to ensure that the two groups are not statistically the same. An independent samples *t*-test was conducted using JASP. The independent samples *t*-test establishes whether or not the mean difference is due to the treatment or sampling error (Lowie & Seton, 2012). However, the results showed the two groups were not normally distributed according to Levene's Test for

Equality of Variance for an independent samples *t*-test. Therefore, a Mann-Whitney U test was conducted.

3.2.4 Independent *t*-test Results

The results from the Mann-Whitney U test indicated that there was a significant difference between the at-risk and not-at-risk groups for the holistic and vocabulary scores. For the holistic scores, the not-at-risk group (*Mdn* = 103) obtained significantly higher holistic scores than the at-risk group (*Mdn* = 49), $U = 0$, $p <.001$, $r = - 1.00$. A similar result was found for vocabulary scores with the not-at-risk-group (*Mdn* = 8) scoring significantly higher than the at-risk group (*Mdn* = 3), $U = 1487$, $p < .001$, $r = -.904$ The results are displayed below in the order discussed (Table 5).

Table 5

Mann-Whitney U test results between the at-risk and not-at-risk group

	U	***p***	***r***
Holistic Grade	0	< .001	-1.00
Vocabulary Scores	1487.50	< .001	-.904

Overall, the results demonstrate that there was a significant difference between the at-risk and not-at-risk group for holistic and vocabulary scores. This means that the writing proficiency groups are different and therefore should be investigated further in regard to the vocabulary characteristics that contribute to the differences between the not-at-risk and at-risk group.

3.3 Research Question 2: Lexical Frequency and Lexical Stretch

3.3.1 Lexical profiles of at-risk and not-at-risk students

When examining the lexical profiles of student writing, nine lexical measures were compared to the holistic scores of writing proficiency obtained from the not-at-risk and at-risk writing samples using the Pearson product moment correlation coefficient (Pearson *r*).

Specifically, the number of tokens, types, and type-token ratio (TTR) were examined. In addition, a modified version of Laufer and Nation's (1995) Lexical Frequency Profile (LFP) analysis using the proportions of high-, medium-, and low-frequency words used by Douglas (2015) was examined. As well, Douglas' Lexical Stretch (2010, 2013, 2015) which records: the lowest frequency list accessed by a test-taker to cover 98% of the text, the lowest frequency list used, and the number of frequency lists accessed were analyzed. These analyses together allowed for a deeper understanding of what *kind* of vocabulary contributes to holistic scores.

3.3.2 Token, Type, Type-Token Ratio

To start the investigation of which vocabulary characteristics contribute to at-risk and not-at-risk writing, the number of tokens (the total number of running words), number of types (the total number of different words according to the word families in the BNC/COCA), and type-token ratio (TTR) were correlated to holistic scores.

The token and type values were obtained for each of the 353 writing sample using AntWordProfiler and transferred to the Excel mastersheet containing all the participants' SAFE rubric analytic and holistic scores. TTR was calculated by dividing the number of types in a text by the number of tokens then recorded for each writing sample. A high TTR value often implies that there is a large amount of lexical variation in texts whereas a low TTR implies the opposite (e.g., Cumming et al., 2005, 2006; Engber, 1995; McNamara et al., 2010).

3.3.3 Lexical Frequency Profile (Laufer & Nation, 1995)

The next analysis conducted on at-risk and not-at-risk writing was the Lexical Frequency Profile (Laufer & Nation, 1995) which is a measure of lexical sophistication that examines the proportion of differing vocabulary frequency levels and academic words in writing. Laufer and Nation (1995) claimed that the Lexical Frequency Profile (LFP) can differentiate between students of different proficiency levels and can be beneficial for deciding the lexical characteristics that impact judgements of writing quality. In other words, LFP may also be able to differentiate between the levels of not-at-risk and at-risk writing levels and determine factors that affect such judgements of writing.

To measure the LFP for the present study, Douglas' (2015) LFP categorizations as discussed previously were used as a measurement of lexical sophistication because Laufer and Nation's (1995) cut-offs do not consider the more refined word frequency categorizations such as mid- and low- frequency words that are presently relevant in vocabulary frequency research. In addition, Laufer and Nation did not find significant differences in the usage of the 2,000 most frequent words used between low- and high-proficiency groups even though there was a pattern for the less proficient groups to utilize more of the second 1,000-word families. In addition to Douglas' (2015) LFP categorization method, three additional measures called Lexical Stretch (Douglas, 2010, 2013, 2015) were used for the present study.

3.3.4 Lexical Stretch (Douglas, 2010, 2013, 2015)

Douglas (2010, 2013, 2015) developed three lexical measures referred to as Lexical Stretch which are additional measures for lexical frequency profiling. The first, was to measure the lowest frequency list used by the writer to cover 98% of a text. This was calculated by cumulatively adding the coverage percentages provided by the frequency lists until 98% coverage, then that frequency list number is recorded. For example, a not-at-risk writer may possibly use a lexical item from the 17K list, which is a low-frequency list or sophisticated vocabulary, with words from the 1K-17K lists representing 98% of the vocabulary output. In comparison, an at-risk writer may have only used vocabulary items from the 1-2K lists, which are higher frequency lists, covering 98 % of the vocabulary output.

The second Lexical Stretch measure was the lowest frequency list used overall by a writer in a written sample. For example, if a not-at-risk writer used words from the 17K list to reach 98% coverage, they may have also used a word from the 22K list. Meaning that the 22K list would be the lowest frequency list accessed and recorded. In contrast, an at-risk writer may only access the 1-2K lists to reach 98%, but also used words from the 5K list. This would mean the lowest frequency list accessed overall was the 5K list.

The third Lexical Stretch measure counts the number of frequency lists used in a writer's text. Douglas (2015) claims that this analysis may reveal proficiency gaps between lists used by a writer and can provide a better understanding of a writer's lexical stretch. For example, the lowest frequency list used by a writer may have been the 10K list without accessing the 4K, 6K, and 8K lists. Even though the top list accessed is 10K, the total number of lists accessed is four (e.g., 1K, 2K, 3K and 10K lists).

3.3.5 Correlation Analysis: Lexical Frequency Profile Measures and Holistic Scores

The nine lexical frequency profiling measures included in LFP and Lexical Stretch were calculated then recorded for each student. To start, each of the writing samples ($N = 353$) were processed through AntWordProfiler (Anthony, 2014) to obtain lexical coverage figures according to the proportion of frequency lists occurring in groups of the 1,000-word families. AntWordProfiler compared the writing samples to Nation's (2012) vocabulary lists which are composed of 1,000-word families of decreasing frequency in English. By understanding which word lists were used for each writing sample, a better understanding of what kind of vocabulary students at-risk and not-at-risk used can be revealed. The lexical coverage figures provided by AntWordProfiler were sorted according to the BNC/COCA 25 frequency lists with the associated cumulative percentages to cover each text. These cumulative percentages were recorded according to Douglas' (2015) categorizations of High-Frequency Vocabulary (HFV) defined as the 1,000-2,000 word list, Mid-Frequency Vocabulary (MFV) defined as the 3,000-10,000 word list, and Low-Frequency Vocabulary (LFV) defined as the 11,000-25,000 word list (including off-list items). Finally, the three Lexical Stretch measures were calculated and recorded for each student as outlined in Section 3.3.4.

After all the LFP and Lexical Stretch measures were recorded for each participant, the nine measures listed above were correlated with the holistic grade received using JASP. The reason for correlating the holistic grade rather than the vocabulary analytic score is because the purpose of this research is to understand which lexical characteristics are impacting overall writing proficiency. The correlation analysis was conducted using the Pearson product moment correlation coefficient (Pearson *r*) with the same effect size cut-offs used earlier in the study. All results are presented in the order discussed.

3.3.6 Correlation Results

Overall, the results for the nine lexical frequency profile measures calculated for the 353 students demonstrated that six of the nine measures were significantly correlated with the holistic writing scores obtained. The results for the comparisons of lexical measures and holistic scores are displayed in Table 6.

Table 6

Correlations between lexical profiling measures and holistic scores

Lexical Measure	**Pearson *r* (*N* = 353)**	***p***
Token	.49	< .001*
Type	.51	< .001*
TT/Ratio	-.190	< .001*
HFV (1K+2K)	-.049	.356
MFV (3K-10K)	.060	.260
LFV (11K-25K + off-list)	.018	.734
98 % list coverage	.16	.002*
Top lists accessed	.20	< .001*
# of lists accessed	.28	< .001*

Significant values of $p \leq .05$ are indicated using an asterisk

The first set of analyses examined the relationship between the number of tokens, types, and TTR with holistic writing scores. As seen in Table 6, there was a significant positive correlation for the number of tokens and types used in relation to the holistic writing score. For the token measure, there was a significant medium positive correlation between the number tokens in student writing and holistic scores, with 24% of the variance explained by the number of tokens in a text, $r(351) = .49$, $p < .001$, $R^2 = .24$. Similarly, there was a significant large positive correlation between the number of types in a text and holistic scores obtained, with 26 % of the variance explained by the number of types used, $r(351) = .51$, $p < .001$, $R^2 = .26$. Students who wrote more tokens and used more word types in their writing were associated with obtaining higher writing proficiency scores. In contrast, there was a significant small negative correlation between the TTR and holistic writing scores with 3.6% explaining the variance in scores, $r(351) = -.19$, $p < .001$, $R^2 = .036$. This suggests students who used more word variation in their writing were associated with obtaining lower writing proficiency scores.

The second set of analyses investigated the relationship between the proportion of coverage in each student writing sample for HFV, MFV, and LFV and holistic writing proficiency scores. Surprisingly, the results indicated there was a non-statistically significant, very small negative correlation between the proportion of HFV, $r(351) = -.049$, $p = .356$. Additionally, there was a non-statistically significant very small correlation between the proportion of MFV, $r(351) = .060$, $p = .260$, and LFV, $r(351) = .018$, $p = .734$, and holistic writing scores. The results indicate the percentage of coverage provided by HFV, MFV, and LFV in writing texts had no significant relationship to holistic writing scores. This suggests that word frequency in terms of high-, mid-, and low-frequency vocabulary may not contribute to writing proficiency scores.

The third set of correlation analyses concerned the relationship between the three measures for lexical stretch and holistic writing scores. For the first correlation concerned with the lowest frequency list used to reach 98% coverage of a writing sample, there was a statistically significant small positive relationship, $r(351) = .16$, $p = <.05$, $R^2 = .026$. Thus, suggesting that increases in the lowest frequency list used to reach 98% coverage were associated with increases in students' holistic writing score. For the relationship between the top list used by a student and holistic scores, a statistically significant small positive correlation was found, $r(351) = .20$, $p <.001$, $R^2 = .040$. Demonstrating 4% of the variance was explained by the top frequency list used in a piece of writing. Likewise, there was a statistically significant small positive correlation between the number of lists accessed in writing samples and holistic score, $r(351) = .28$, $p <.001$, $R^2 = .078$. This resulted in 7.84% of the variance explained by the number of frequency lists used by a writer. The results imply that increases in the top list used and the number of lists accessed by a writer somewhat contribute to increases in holistic scores obtained.

3.3.7 Independent *t*-test between at-risk and not-at-risk writing

After establishing the relationships between the lexical profiling measures and holistic scores, a further investigation into whether there were differences for such relationships between the at-risk and not-at-risk group was conducted. An independent samples *t*-test was conducted for the token and type variables because they are continuous variables. A Man-Whitney U test was conducted for the remaining variables because they were either non-normally distributed or, nominal or ordinal variables (Pallant, 2016) A Bonferroni correction was applied to the significance level to control the overall Type I error rate from inflating when multiple statistical tests are performed (Field, 2013; Tabachnick & Fidell, 2007). This was calculated by dividing the significance level by the number of significance tests performed. As stated previously, the significance level for this study is α = .05. There are 9 nine statistical measures, therefore, the adjusted significance level is .005 (.05/9).

3.3.8 Independent t-test Results: Tokens and Types

The results indicated there were significant differences between the at-risk and not-at-risk group for six of the nine measures, as shown in Table 7. Starting with the first measure, the results indicated that the average number of tokens produced by the not-at-risk group (*M* = 196.2, *SD* = 50.40) was significantly greater than at-risk group (*M* = 144.4, *SD* = 51.87), $t(351) = -9.49$, $p < .001$. Similar results occurred for the types measurement, the not-at-risk group (*M* = 101.7, *SD* = 22.13) used significantly more types on average in their writing than the at-risk group (*M* = 77.17, *SD* = 23.17), $t(351) = -10.16$, $p < .001$. These findings suggest that a characteristic of not-at-risk writing may be longer texts containing a variety of words while the opposite may be true for at-risk writing.

Table 7

Independent *t*-test results between at-risk and not-at-risk for token and types

Lexical Measure	**t**	**p**	***d***
Token	-9.49	< .001*	-1.01
Type	-10.16	< .001*	-1.08

Significant values of $p \leq .005$ are indicated using an asterisk

3.3.9 Independent t-test Results: TTR, HFV, MFV, LFV, Lexical Stretch

The results from the Mann-Whitney U test results are displayed in Table 8. The not-at-risk group (*Mdn* = 52.57) had a significantly lower type-token ratio in comparison to the at-risk group (*Mdn* = 55.30), $U = 18490$, $p = .002$. This suggests that at-risk writers may use a greater variety of words in relation to the number of words in a text.

Table 8

Mann-Whitney U test results between at-risk and not-at-risk group for six measures

Lexical Measure	**U**	***p***	***r***
TT/Ratio	18490	.002*	0.189
HFV %	16906	.159	0.087
MFV % 3-10	14408	.231	-0.074
LFV % (+ off list) 11-25	14127	.106	-0.092
98 % list coverage	11840	< .001*	-0.259
Top list accessed	11521	< .001*	-0.239
# of lists accessed	10843	< .001*	-0.303

Significant values of $p \leq .005$ are indicated using an asterisk

As for the frequency of vocabulary measures (HFV, MV and LFV), there were no significant differences between the not-at-risk and at-risk group. There were no significant differences between the not-at-risk group (*Mdn* = 84.52) and at-risk group use of HFV (*Mdn* = 85.23), $U = 16906$, $p = .159$. Similar results occurred for the use of MFV, the not-at-risk group (*Mdn* = 13.44) did not have a significantly higher percentage of MFV in their writing than the at-risk group (*Mdn* = 13.25), $U = 14408$, $p = .231$. Furthermore, for LFV, the not-at-risk group (*Mdn* = .40), did not use a significantly higher percentage of LFV than the at-risk group (*Mdn* = 0), $U = 14127$, $p = 0.106$. These findings suggest that the frequency of vocabulary does not have a relationship with holistic scores. As well, these findings may suggest that using these methods, we are unable to distinguish differences in vocabulary use between the two groups under study.

Moving on to the Lexical Stretch measures, the not-at-risk group (*Mdn* = 5) used a significantly lower frequency list to reach 98% coverage than the at-risk group (*Mdn* = 4), $U = 11840$, $p < .001$. This may suggest that not-at-risk writers use a larger span of vocabulary to cover 98% of the text in comparison to at-risk writers. As for the top frequency list, the not-at-risk group (*Mdn* = 9), used a significantly lower top frequency list than the at-risk group (*Mdn* = 6), $U = 11521$, $p < .001$. This implies that not-at-risk writers are using words from lower frequency lists in comparison to the at-risk writers. Finally, the not-at-risk group (*Mdn* = 6) used a significantly greater number of lists than the at-risk group (*Mdn* = 5), $U = 10843$, $p < .001$. This finding suggests that the not-at-risk group is using a greater variety of words from different frequency lists.

3.4 Research Questions 3a and 3b: Specialized Vocabulary

3.4.1 Correlation Analysis and Independent Samples t-test: Academic and Discipline-Specific Vocabulary Usage

In order to write successfully, it is recommended to use vocabulary specific to the genre such as academic and discipline-specific vocabulary (e.g., Bruce, 2008; Swales, 1998). One such lexical characteristic to examine is whether more successful students use more academic vocabulary and discipline-specific (engineering) vocabulary in their writing to accurately write to the genre for SAFE. Therefore, it may be predicted that students are who are not-at-risk use more academic and disincline-specific vocabulary in their writing in comparison to students who are at-risk. For this thesis, academic vocabulary was operationalized using the Academic Word List (AWL) (Coxhead, 2000) while discipline-specific vocabulary for engineering students was operationalized using the English Engineering Word List (EEWL) (Hus, 2014). To investigate the usage of academic and discipline-specific vocabulary in not-at-risk and at-risk writing, AntWordProfiler (Anthony, 2014) was used to analyze the SAFE corpus for occurrences of the respective two lists.

To investigate the occurrences of the word lists mentioned above, all of the TXT files for the SAFE corpus were uploaded to AntWordProfiler, then the AWL, which contains 570 word families, was uploaded with the 'Batch Process' which allowed for the production of the AWL tokens and types coverage percentages in relation to the number of overall tokens, individually for 353 students. As AntWordProfiler provides various lexical details about each text file in the output results, only the tokens and types coverage percentages were extracted. The same method was used to find the occurrences of the EEWL for each student.

A correlation analysis was conducted between the percentage of AWL words and holistic scores using Pearson product moment correlation coefficient (Pearson *r*). This analysis was also completed for the EEWL. These two analyses were completed to understand the relationship between the use of AWL and EEWL vocabulary in relation to holistic scores. Finally, independent samples *t*-tests were conducted to compare the differences between the at-risk and not-at-risk group usage of the tokens and types for the AWL and EEWL, respectively.

3.4.2 Correlation Results: Academic Vocabulary

The Pearson correlation results demonstrated that there was a significant small positive correlation between the percentage of academic vocabulary ($M = 7.68$, $SD = 2.99$) and holistic scores ($M = 74.16$, $SD = 30.23$), $r(351) = .201$, $p < .001$, resulting in the AWL explaining 4% of the variance in holistic scores. This indicates as students used more academic vocabulary, holistic scores also increased. Table 9 provides the descriptive statistics of the AWL token percentage according to the at-risk and not-at-risk group. The not-at-risk group on average used more academic vocabulary in their writing. This is regardless of the fact that there were fewer written samples for the not-at-risk group than the at-risk group. In other words, even though there were more written samples for the at-risk group, they still used fewer frequently occurring academic words than the not-at-risk group.

Table 9

Descriptive statistics for AWL token percentages according to at-risk and not-at-risk group

	AWL token %	
	At-Risk	**Not-At-Risk**
n	183	170
Mean	7.18%	8.21%
Median	7.14%	8.08%
Std. Deviation	3.03	2.86
Minimum	0%	2.24%
Maximum	14.67%	17.49%

3.4.3 Independent *t*-test Results: Academic Vocabulary

The average percentage of AWL tokens and types in the not-at-risk and at-risk group are presented below (Table 10). The not-at-risk group ($M = 8.21$, $SD = 2.86$), had a larger percentage of AWL tokens in their writing in comparison to the at-risk group ($M = 7.18$, $SD = 3.03$), $t(351) = -3.29$ $p \leq .001$. Additionally, the not-a-risk group ($M = 10.60$, $SD = 3.79$), used a significantly greater percentage of AWL types in their writing than at-risk students ($M = 9.09$, $SD = 3.52$), $t(351) = -3.87$ $p < .001$. Overall, the at-risk group consistently used fewer academic tokens and types in comparison to the not-at-risk group. It should be noted that these differences occurred despite the number of not-at-risk samples being fewer than the at-risk samples. This suggests that students who perform better on SAFE are using more academic vocabulary in their writing than students whose writing are considered at-risk.

Table 10

Average percentage of AWL tokens and types in at-risk and not-at-risk texts

	AWL occurrences	
	At-Risk	**Not-At-Risk**
n	183	170
Tokens	7.18%	8.21%
Types	9.09%	10.60%

3.4.4 Correlation Results: Discipline-Specific Vocabulary

The Pearson correlation results demonstrated that there was a non-significant positive correlation between the percentage of engineering vocabulary (M = 9.47, SD = 3.16) and holistic scores (M = 74.16, SD = 30.23), $r(351) = .10$, $p = 0.060$. This suggests that the use of discipline-specific vocabulary is not associated with holistic scores. Table 11 provides the descriptive statistics of the EEWL token percentage according to the at-risk and not-at-risk group. The at-risk group and not-at-risk group are producing similar amounts of engineering specific vocabulary.

Table 11

Descriptive statistics of the EEWL token percentages according to at-risk and not-at-risk group

	EEWL token %	
	At-Risk	**Not-At-Risk**
n	183	170
Mean	9.23%	9.72%
Median	8.94%	9.52%
Std. Deviation	3.30	2.98
Minimum	1.44%	2.66%
Maximum	19.64%	18.03%

3.4.5 Independent *t*-test results: Discipline-Specific Vocabulary

Moving on to the occurrence of the discipline-specific vocabulary in the not-at-risk and at-risk groups, the average occurrence of EEWL tokens and types in percentages are presented below in Table 12. At first glance, the not-at-risk writers are on average using more EEWL tokens and types in their writing in comparison to the at-risk students for each test. However, the results from the t-test indicated that the EEWL token usage was not significantly greater for the not-at-risk group (M = 9.72, SD = 2.98) in comparison to the at-risk group (M = 9.24, SD = 3.30), $t(351) = -1.43$, $p = .153$. These results imply that not-at-risk and at-risk students use a similar amount of discipline-specific vocabulary in their writing. Similar results occurred for the

EEWL types, the not-at-risk group (*M* = 11.21, *SD* = 3.45) did not have a significantly greater percentage of EEWL types in their writing than the at-risk group (*M* = 10.23, *SD* = 3.69), $t(351) = -2.58$, $p = .010$. These results imply that not-at-risk and at-risk students use a similar amount of discipline-specific vocabulary in terms of individual words and variety in their writing.

Table 12

Average percentage of EEWL tokens and types in at-risk and not-at-risk texts

	EEWL occurrences	
	At-Risk	**Not-At-Risk**
n	183	170
Tokens	9.24%	9.72%
Types	10.23%	11.21%

3.5 Research Question 3c: Academic Formulaic Language

3.5.1 Correlation Analysis: Academic Formulaic Language

Since the 1980s, research has found that higher writing proficiency has been connected to the use of formulaic language, especially in academic discourse (e.g., McCully, 1985). However, when Appel and Wood (2016), compared how high- and low-proficiency writers use of formulaic language in EAP student writing, the results indicated that higher proficiency writers used fewer formulaic language sequences than low-proficiency writers. This contrast prompted an investigation of the use of formulaic language in at-risk versus not-at-risk writing compositions. Specifically, the use of *academic* formulaic language as SAFE seeks to diagnose academic writing ability. Simpson-Vlach and Ellis (2010) created the Academic Formulas List (AFL) which includes 200 formulaic sequences from a corpus of 2.1 million words each for academic speech and academic writing. Specifically, the academic writing sub-corpora were derived from the subjects: humanities and arts, social sciences, natural sciences/ medicine, and technology and engineering. The logic behind using the Written AFL for the present study is that if these formulas are occurring frequently in academic writing, then in order for writing to be considered proficient, students would likely need to include these frequently occurring formulas in their writing to successfully write to the genre.

To analyze the differences in academic formulaic language between at-risk and not-at-risk writing samples, AntConc (Anthony, 2018) was used. AntConc is an open-source corpus analysis toolkit for concordance and text analysis. The software can be used to produce lists of frequently occurring word combinations/clusters, concordances, and collocates based on the corpus of focus. For the present study, the list of the 200 Written AFL was compared to the SAFE corpus to understand if academic formulas were being used in student writing.

To start, the Written AFL was modified to also include 'wildcards' by inserting an asterisk after words for the sequences that may have inflected forms. For example, the sequence 'the difference between the' was modified to 'the difference* between the'. This was completed to ensure that the regular forms and inflected were detected by AntConc.

Because the AFL contains word sequences ranging from three-, four-, and five-words, they needed to be separated and searched for individually. This was required as AntConc can only find recurrent word combinations in a corpus from a list based on a single number of word cluster sizes. For example, to find the three-word clusters, the 'cluster size' function would need to be set to a minimum and maximum of three. As well, only the three-word clusters from the AFL would be input.

Afterwards, the TXT masterfiles for each group was loaded to AntConc respectively, then under 'cluster/*n*-grams', the 'advanced' option was selected. Next, the modified AFL lists were loaded with the wildcards and 'use search term(s) from list below' was selected. This produced a list of the individual types of AFL formulas along with the frequency of occurrence of the academic formulas, and the range of occurrence in the corpus (how many separate writing samples the formulas occurred in) listed according to the rank of frequency. This resulted in three separate AFL formula lists (three-, four-, and five-words) each for the at-risk (See Appendix B, C, D) and not-at-risk group (See Appendix E, F, G). All of the AFL occurrences appearing in the SAFE corpus were verified for correct use in the context of writing using the concordance lines generated by AntConc.

Of importance, was the frequency of occurrence of the academic formulas. The 'concordance' function in AntConc was used to identify the frequency in which TXT file (i.e., student writing) the academic formula occurred in. The occurrence of each three-, four-, and five-word academic formula was recorded for each student. However, since the AFL has a fair bit of overlap, the clusters that AntConc produced needed to be sorted so the occurrences of the formulas were not counted twice. This was completed by using the 'concordance' function in AntConc. For example, the five-word cluster 'on the other hand the' may have appeared two times in one TXT file. However, there is a four-word cluster from the AFL, 'on the other hand', that also appeared and was counted by AntConc because 'on the other hand' occurs within 'on the other hand *the*'. Although they may seem the same, when using the concordance function the word after the four-word cluster may be different such as 'on the other hand, *rises*…' and therefore must be counted separately.

A correlation analysis was conducted between the frequency of AFL occurrences and holistic scores using the Pearson product moment correlation coefficient (Pearson *r*). The frequency of occurrence values for each test was included for all three-, four-, and five-word academic formulaic sequences.

The last step was to conduct an independent samples *t*-test to determine if there were differences in the use of academic formulas between the at-risk and not-at-risk groups. Because the formulaic sequences frequencies are categorical variables, a Man-Whitney U test was conducted in order to determine if there were significant differences between the not-at-risk and at-risk use of formulaic language in writing.

3.5.2 Correlation Results: Academic Formulaic Language

The Pearson correlation results demonstrated that there was a significant small positive correlation between academic formula use and holistic scores, $r(351) = .19$, $p < .001$ ($R^2 = .036$) This indicates as students used more academic vocabulary, holistic scores also increased. Table 13 provides the descriptive statistics of the AFL occurrences according to the at-risk and not-at- risk group.

Table 13

Descriptive statistics for overall AFL occurrences according to at-risk and not-at-risk group

	AFL Occurrences	
	At-Risk	**Not-At-Risk**
n	183	170
Mean	0.74	1.22
Median	0	1.00
Std. Deviation	1.18	1.381
Minimum	0	0
Maximum	5	6

3.5.3 Independent t-test results: Academic Formulaic Language

The results for the total number of three-, four-, and five-word academic formulaic occurrences for each group are presented in below (Table 14). As can be seen, the at-risk group consistently produced fewer three-, four-, and five-word academic formula than the not-at-risk group. For the overall use of the AFL, the Mann-Whitney U test found that the not-at-risk group (*Mdn* = 1) used significantly more academic formulaic language than the at-risk group (*Mdn* = 0) $U = 11874.50$, $p < 0.001$. These results suggest that the lack of formulaic language in writing may be a characteristic of at-risk writing for a diagnostic assessment.

Table 14

Total frequency of occurrences for the three-, four-, five-word occurrences AFL

	AFL Occurrences	
	At-Risk	**Not-At-Risk**
3-word formula	61	117
4-word formula	31	43
5-word formula	13	31
Total	105	191

3.6 Tool for Automatic Analysis of Lexical Sophistication (TAALES)

3.5.1 Correlation Analysis: TAALES

A further analysis of the vocabulary in student writing was conducted to support the findings of the present study. The Tool for Automatic Analysis of Lexical Sophistication 2.0 (TAALES; Kyle, Crossley, & Berger, 2017) was used to support the relationship between sophisticated vocabulary use and holistic scores obtained on the SAFE exam. As mentioned previously, TAALES is a freely available lexical analysis tool that provides various indices of lexical sophistication. TAALES indices were used to obtain lexical sophistication scores pertaining to the AWL (Research Question 3a) and the AFL (Research Question 3b). These selected indices were correlated to the vocabulary analytic scores because they provided support for results that were previously observed in this thesis for the relationship between holistic scores and the other lexical sophistication measures.

TAALES 2.0 works by loading the input text, selecting the desired index options, then processes the lexical indices in an output file into a CSV file. To start the analysis, all of the TXT masterfiles for each group were loaded in the 'input' text. Then, all the indices under 'academic language' were chosen to be included in the analysis. These indices included 'academic formulas list' and 'academic word list'. Next, the option of 'individual item output' was selected and processed into an output CSV file.

After the selected indices were generated, only select supporting indices that were associated with the research questions for each writing sample were used for analysis. The indices were selected in the output CVS file according to the research question. To support Research Question 3a, the 'All AWL Normed' index was selected. For Research Question 3c, the 'Written AFL Normed', 'All AFL Normed', 'Core AFL Normed', and 'COCA Academic Trigram Frequency' were selected. The trigram indices were the only indices available to support the 3-word formulaic sequences but not for longer word sequences. The above measures were then correlated to the vocabulary analytic scores using the Pearson product moment correlation coefficient (Pearson r) with the same effect size cut-offs indicated earlier in the study. The cut-off for significance was $p \leq .05$. Writing compositions that effectively used selected indices are predicted to have been rated as high-proficiency writing.

3.5.1 Correlation Results: Academic Word List (AWL) using TAALES

The results for the TAALES analysis are presented in Table 15. As can be seen, the results of the Pearson correlation indicated that there was a significant large positive relationship between the analytic vocabulary scores and AWL index, $r(351) = .57$, $p < .001$, 32% of the variance explained by the AWL. This suggests that the use of academic vocabulary may be associated with analytic vocabulary scores which also contribute to holistic scores.

3.5.2 Correlation Results: Academic Formulas List (AFL) using TAALES

As illustrated in Table 15, the Pearson correlation results indicated that there was a significant strong positive relationship between the COCA Academic Trigram Frequency index and analytic vocabulary scores, $r(351) = .43$, $p < .001$, $R^2 = .18$. This finding suggests that three- word sequences may support analytic vocabulary scores. As for both the Written AFL and the All AFL index, there was a significant medium positive relationship between the respective indices and analytic vocabulary scores, $r(351) = .15$, $p < .001$, $R^2 = .022$. This suggests that both written and spoken academic formulaic sequences contribute to analytic vocabulary scores. Finally, there was a non-significant small positive correlation between the Core AFL index and analytic vocabulary scores, $r(351) = .080$, $p = .066$. This finding suggests that the Core AFL may not be pedagogically relevant for written diagnostic assessments.

Table 15

Results for Pearson r Correlations between TAALES Measures and Vocabulary Analytic Scores

TAALES Measure	**Pearson r (N = 353)**	**p**
All_AWL_Normed	.57	< .001*
COCA_Academic_Trigram_Frequency	.43	< .001*
Written_AFL_Normed	.15	.002*
All_AFL_Normed	.15	.002*
Core_AFL_Normed	.080	.066

Significant values of $p \leq .05$ are indicated using an asterisk

Chapter 4: Discussion

4.0 Introduction

This chapter discusses the main research findings in response to each of the research questions. The results presented in the previous chapter are evaluated and compared to previous research studies. Section 4.1 discusses the relationship between holistic scores and vocabulary scores. Section 4.2 examines the relationships between nine lexical measures and holistic scores. Next, Section 4.3 and 4.4 reviews the use of academic and discipline-specific vocabulary in at-risk and not-at-risk writing. Academic formulaic language is also interpreted in terms of frequency of use in at-risk and not-at-risk writing in Section 4.5. Throughout the discussion, TAALES will be utilized to triangulate the results with the other measures used for lexical sophistication characteristics.

4.1 Research Question 1: The relationship between holistic and vocabulary scores

The aim of the first research question was to examine the relationship between holistic writing scores and vocabulary scores. This analysis helped to determine what impact analytic vocabulary scores had on holistic writing scores. Through correlation analyses, there was a confirmation of a significant large positive correlation between vocabulary scores and holistic writing scores ($r = .84$, $p < .001$). The large positive relationship between vocabulary scores and holistic scores demonstrated that higher vocabulary scores were consistent with increased writing proficiency scores.

These findings are consistent with previous studies (e.g., Astika, 1993; Espinosa, 2005; Lee et al., 2009) which found a range of correlations from $r = .44$ to $r = .94$, indicating that vocabulary scores are associated with writing scores. However, past research has typically examined general writing test prompts (e.g., Astika, 1993; Espinosa, 2005; Lee et al., 2009). Topics have included self-introductions and personal lifestyle questions or lists of options from which the writer selects a desired topic, but rarely discipline-specific topics. The present thesis adds to the growing body of literature which recognizes the importance of analyzing vocabulary use in discipline-specific writing—in this case, ESP-based writing tasks in engineering. This is important for several reasons. Discipline-specific literacies are gaining popularity in assessment and diagnosis contexts because of their ability to provide meaningful diagnostics and aid in the prospective impact of the individualized academic support students would receive (Fox & Artemeva, 2017). Prior (1998) emphasizes that discipline-specific literacies are "central to disciplinary enculturation [...] for foregrounding representations of disciplinarily, and for negotiating trajectories of participation in communities of practice" (p. 32). Furthermore, the importance of disciplinary writing tasks is recognized by the Measuring the Academic Skills of University Students approach (e.g., Bonanno & Jones, 2007). SAFE recognizes that academic literacy skills need to be diagnosed in order to provide useful and impactful feedback to augment individualized academic support which may aid in program retention. As seen from the results, vocabulary can be a part of this process.

Previous research has suggested that disciplinary literacies and academic support are a crucial factor for retention and program completion (Fox, 2005; Meyer & Land, 2003). Considering how the analytic vocabulary score on SAFE targets sufficient and effective vocabulary use, it is even more critical to consider vocabulary as a potential indicator and point of intervention for at-risk academic writing. This is reflected through the large correlation found in the present study and is important because vocabulary is an underlying factor that contributes to the development of writing and language skills in general (e.g., Coxhead, 2012; Douglas, 2013; Nation, 2001; Paquot, 2010).

The findings of the present research are consistent with previous studies which found vocabulary to be associated with writing quality scores using different types of assessment rubrics, such as the ESL Composition Profile (Jacobs et al., 1981) or TOEFL in the non-native speaker context. Despite previous research utilizing different grading rubrics and non-native language groups to examine the relationship between vocabulary and writing, the present study has also arrived at the same conclusion using an indigenously drawn rubric that was informed collaboratively by language and engineering experts (Fox & Artvema 2017). In addition, the SAFE rubric was able to draw this conclusion for native and non-native speakers who participated in the assessment. For example, Astika (1993) found that vocabulary accounted for the largest percentage of variance (83.75%) in holistic scores for writing compositions out of all the writing characteristics examined. Jacobs et al.'s (1981) rubric was used which included Content, Organization, Vocabulary, Language use, and Mechanics designed to assess ESL students, with Vocabulary contributing to 20% of the holistic mark (Jacobs et al., 1981). The SAFE rubric had three main components: Language, Logic, and Rhetoric each of which can be broken down into five separate analytic scores, with Vocabulary contributing 6%. The foci of the SAFE rubric and Jacobs et al.'s (1981) rubric were the same – to assess writing. However, Astika's (1993) focus was on the assessment of general language while the present study focused on language in addition to engineering content and understanding.

Similarly, for the TOEFL rubric which also focused on Content, Organization, Vocabulary, Language use, and Mechanics but with different numerical analytic scores, Lee, Gentile, and Kantor (2009) found large correlations of $r = .50$ and $r = .44$, respectively between vocabulary and holistic scores when comparing two prompts for a TOEFL computer-based writing assessment. Vocabulary was the highest variable to correlate with holistic scores out of the previously mentioned variables. Lee et al.'s (2009) sample also included 930 test-takers with 58 language backgrounds, all of which were non-native speakers. In comparison to the present study, which had 353 test-takers with a mix of native and non-native speakers. It has previously been reported that native and non-native speakers produce different writing profiles based on their own language experience (Morris & Cobb, 2004).

Considering the differences in sample sizes, which also contributes to the nature of the sample (different subgroups for first languages), a larger and more variable sample size impacts the strength of correlation (Goodwin & Leech, 2006). Nevertheless, the present study found a large positive correlation between vocabulary and holistic scores. It is therefore noteworthy that despite examining different assessment prompts (Car and Energy), rubrics with numerical component distributions and foci, and a mix of native and non-native speakers, the present study can support the importance of vocabulary use in academic writing.

Another finding in this study was that students who are at-risk typically had lower vocabulary and holistic writing scores regardless of the prompt. This adds to the argument of vocabulary being an important underlying variable for writing quality. A closer look at the scores for each DGIT confirmed this idea. For example, at-risk students who on average had low vocabulary scores (DGIT-Car and Energy, Mdn = 3) typically received a low holistic grade (Mdn = 52 and Mdn = 46, respectively). The same was found for the not-at-risk group for vocabulary scores on the DGIT-Car (Mdn = 8) and Energy test (Mdn = 7), which typically received a high holistic grade (Mdn = 107 and Mdn = 102, respectively).

This finding aligns with Espinos (2005) who found a large positive correlation between vocabulary proficiency (measured as good vs. poor) and writing proficiency. Roessingh (2008) also found that low analytic vocabulary scores negatively correlated with other analytic criteria scores when assessing writing quality. These studies further suggest that the kind of vocabulary used contributes to holistic scores.

Based on the result that indicated a significant difference in the holistic scores between at-risk (*Mdn* = 103) and not-at-risk students (*Mdn* = 49), it may be suggested that at-risk students' writing resembles non-native speakers' writing abilities. According to Raimes (1985), when comparing native and non-native writers, non-native writers require "more of everything" (p. 250), which includes more time to read, compose, think, edit, and review English writing texts. This is further supported by Eckstein and Ferris (2018), who compared native and non-native speakers' written texts. They found that there was a significant difference between the two groups' holistic scores – non-native speakers received lower writing scores in comparison to native speakers.

Roessingh (2008) confirms the relationship between vocabulary and writing scores for both native and non-native speakers. When comparing native and non-native speakers, it was found that the non-native speakers were lacking vocabulary and it was recommended that non-native speakers would benefit if they increased their vocabulary knowledge, as vocabulary also contributes to conveying thoughts and details in a writing composition. This concept may also apply to SAFE assessment students who may potentially be diagnosed as at-risk because of their lack of appropriate vocabulary. In other words, at-risk students may need help developing their vocabulary to improve their writing, which means that second language research studies focusing on writing may be able to inform at-risk writers regardless of their first language. A future study could include participants' demographic information and the number of languages spoken. These details would help shed light on the individual characteristics of at-risk students, thus aiding in the SAFE diagnostic and academic support process.

4.2 Research Question 2: Lexical Frequency and Lexical Stretch

The results in response to Research Question 2 will be discussed in the order they are listed in Table 6 (Section 3.3.6). The findings will primarily be discussed and compared in relation to Douglas' (2015) study, as the present study builds off of that study and uses his categorizations of vocabulary frequencies. Additionally, both the present research (N = 353) and Douglas' (N = 200) investigated writing assessments through a lexical lens for large samples.

4.2.1 Tokens

The present study found that there was a significant medium correlation between holistic scores and the number of tokens used in student writing (r = .49). The number of tokens explained 24% of the variance in writing scores. These findings are supported by previous research which found a range of large significant correlations between tokens and holistic writing scores (r = .65 to r = .82) for various forms of standardized language assessments, including IELTS and TOEFL (e.g., Banerjee et al. 2007; Douglas, 2015; Frase, et al., 1998; Kaplan et al., 1998; McNamara et al., 2010, 2013). Douglas (2015) found a significant medium positive correlation between the number of written tokens and CELPIP-General test levels (r = .54). The number of tokens explained 30% of the variance in holistic scores, only 6% more than the present study found.

Similarly, the present study adds to previous research findings of the relationship between tokens and writing proficiency derived from large-scale general English proficiency assessments (e.g., Banerjee et al., 2007; Douglas, 2015). The present study appears to confirm that the same is true for a discipline-specific diagnostic writing assessment.

The findings indicated that the not-at-risk group produced significantly more tokens than the at-risk group (see Table 7), suggesting that students who write more are more likely to obtain greater writing scores for SAFE, whereas students who write less are more likely to receive a lower writing score. This result may also suggest that not-at-risk students are able to better express their ideas when writing, resulting in more content and tokens. Meanwhile, at-risk students may have difficulty expressing their ideas, resulting in less content and fewer tokens.

This finding is interesting because during the SAFE process, there are is a video and lecture to support the task. During the written administration, the prompts specifically target engineers. As a result, students are typically equipped to answer the content of the discipline-specific questions. Despite this, at-risk students have been found to write less on the SAFE assessment. This could suggest students who are at-risk simply have poor communication skills or do not have the ability to express their ideas to answer and construct a successful writing piece. This is known to be a great challenge in writing according to Raimes (2002), a prominent writing researcher, who points out that writing is an intricate task, as writers must produce ideas, present them, and be critical to both the ideas and structure of the piece. In the future, SAFE could require a minimum word (token) count in order to determine if students are at-risk due to a lack of ideas or language, or both.

Considering the connection between an increase in the number of words and holistic scores, it could be suggested that not-at-risk students are using more syntactically complex sentences. It has been found that syntactic complexity is associated with token increases (Stockwell & Harrington, 2003). Similarly, McNamara, Crossley, and McCarthy (2010) compared native speaker writing judged as high- or low-proficiency and found that the high-proficiency writers used more syntactically complex sentences. However, McNamara et al. (2010) measured syntactic complexity by the number of words before the main verb, while the present study only used token counts. A future study may examine the number of words before the main verb in the SAFE test samples to confirm such a connection.

4.2.2 Types

Out of the six significantly correlated lexical profiling measures examined in the present study, the number of types used in the writing compositions for SAFE correlated the strongest with holistic writing scores. There was a significant large positive correlation between the number of types in a text and holistic scores obtained (*r* = .51), resulting in 26% of the variance being explained by the number of types used. At-risk students (*M* = 77.17, *SD* = 22.13) on average used fewer types than not-at-risk students (*M* = 101.7, *SD* = 23.17). Students who used more word types, or more variety of words in their writing, were associated with higher scores in comparison to students who did not.

These findings echo Douglas (2015) who found a significant large positive correlation between CELPIP-General Writing Test levels and the number of word types used in writing (*r* = .77). In his study, 59% of the variance could be explained by the number of types in a text. This suggests that an increase in CELPIP-General Writing Test levels is associated with increased vocabulary variety. Therefore, it may be said that more proficient writers use more variety or diversity of words in their writing. In the present study, this was observed when comparing the overall average number of types produced by at-risk (*M* = 76.60) and not-at-risk writers (*M* = 101.7).

In the context of SAFE, perhaps a salient feature of at-risk students is vocabulary repetition—or lack thereof. Repetition may indicate several things. First, at-risk students may not have adequate control over the vocabulary they produce. Although students may comprehend the meaning of a variety of words, they may not think to produce them or vary their vocabulary in their own writing, as not-at-risk students do. Second, unlike not-at risk students, at-risk students may not have sufficient vocabulary to employ or control. Thus, at-risk students may need to be informed about the importance of the variety of vocabulary in successful writing. Smith (2003) suggested that using an increased range of vocabulary supports the improvement of writing skills; SAFE could emphasize the value of using diverse vocabulary and synonyms when providing writing support to at-risk students.

4.2.3 Type-Token Ratio (TTR)

The current study found that there was a small negative correlation between the type-token ratio (TTR) and holistic scores (r = -.19). TTR explained 4% of the variance in holistic scores. A high TTR value typically indicates that there is a large amount of lexical variation in texts, whereas a low TTR implies the opposite. However, the present findings imply that as students' writing proficiency scores increased, there was a decrease in the variation of unique words used in comparison to the total number of words used. Douglas (2015) did not include the TTR measurement in his study therefore the result will be compared to other research. This is surprising considering that TTR is a common measurement in lexical analysis research (e.g., Cumming et al., 2005, 2006; Engber, 1995; McNamara, Crossley, & McCarthy, 2010), which has suggested that more proficient writers use more variation or diversity of words as indicated by TTR. For example, McNamara et al. (2010) found that high lexical variation in student writing had a significant small positive correlation with high writing proficiency scores (r = .20), where 4% of the variance was explained by TTR.

A potential cause for the discrepancy between the present study and previous work is the length of texts examined. For the present study, the average text length for the entire corpus (N = 353) was 169.4 tokens: the at-risk group averaged 144.4 tokens and the not-at-risk group averaged 196.2 tokens. In comparison, McNamara et al.'s (2010) low-proficiency group's text length was on average 700.11 tokens, and 748.65 tokens for their high-proficiency group's compositions. TTR is known to be sensitive and unstable for varying text lengths and is also impacted by shorter text lengths (Laufer & Nation, 1995; McCarthy & Jarvis, 2010). Thus, the present study may suggest that TTR may not be a good indicator of writing proficiency for shorter texts.

This conclusion would also help explain why there was a contradictory finding in the present study regarding the positive correlation between the types of words used and holistic scores, and the negative correlation between TTR and holistic scores (see Section 3.3.6). A reason for this could be because TTR examines the types in relation to the number of tokens while type solely examines the variety of words. Another potential reason for a negative correlation could be the type of topics for the DGIT. The two prompts in the present study are focused on specific engineering subjects: Car and Energy. SAFE focuses on a concentrated discussion of engineering topics, low diversity in vocabulary used may have been predicted because such discipline-specific words would be used to sufficiently answer the prompt (Frase et al., 1998). This is further reflected and supported in the previous results of the present study, which found no statistical difference between the usage of engineering vocabulary among the at-risk and not-at-risk group (see Section 3.4.5). In addition, engineering communication guidebooks have reiterated the importance of keeping language simple in their writing (e.g., Beer & McMurrey, 2014; Budinski, 2001; Winsor, 2013).

4.2.4 High-, Mid-, Low-Frequency Vocabulary

The present study used Douglas' (2015) criteria for high-, mid-, and low-frequency vocabulary and correlated the proportion of the categories to holistic scores for writing. The results were unexpected: the findings indicated a statistically non-significant, very small negative correlation between the proportion of HFV and holistic scores. Furthermore, there was a non-statistically significant, very small correlation between holistic writing scores and the proportion of MFV and LFV. Notably, there were statistically non-significant differences between the not-at-risk and at-risk group for frequency of vocabulary. This result implies that word frequency as measured by high-, mid-, and low-frequency vocabulary may not contribute to SAFE writing proficiency scores and may not be relevant for discipline-specific writing tasks. Most importantly, LFP may not be a suitable measure for characterizing writing proficiency as previous studies (e.g., Douglas, 2015; Laufer & Nation, 1995) have suggested in the context of a discipline-specific writing task but may be relevant in general language tasks, longer text lengths, and for non-native speakers.

The present LFP findings contradict previous studies that have found word frequency to be a predictor of writing scores (e.g., Douglas, 2015; Laufer & Nation, 1995). Douglas (2015) found a significant large negative correlation between HFV and CELPIP-General Writing Test levels ($r = -.73$), with 54% of the variance being explained by HFV for a standardized test. It is interesting that Douglas (2015) found such a large correlation because the CELPIP-General Test prompts require general knowledge in order to write an e-mail or respond to an opinion survey; these tasks would typically require general vocabulary or high-frequency vocabulary (Coxhead, 2006; Nation, 2001, 2008). Perhaps this difference was because Douglas included a range of more proficient test levels (3-12) texts but omitted the lowest levels (0-2), while the present study only included the highest and lowest levels. Omission of the lowest levels of 0-2 may have helped strengthen the correlation, as the range of data becomes shortened (Goodwin & Leech, 2006). Furthermore, levels 0-2 included non-native writers who may have relied on HFV because such individuals typically learn English in a frequency-based order — such as using Nation's BNC-COCA 25 lists (Nation, 2001; 2013). Thus, potentially resulting in the lack of occurrence in the examined texts because lower-level non-native speakers learn HFV first and therefore are more familiar and comfortable using such vocabulary (Nation, 2013).

In Douglas' (2015) study, for the relationship between MFV and writing scores, there was a significant large positive correlation ($r = .73$), with 53% of the variance being explained by MFV. This suggests that using mid-frequency vocabulary is connected to higher CELPIP-General Test levels. In contrast, the present study found no relationship between MFV and holistic writing scores (r = 0.060). This is rather surprising considering Douglas found such a large correlation and suggested that students seek to use more MFV rather than HFV for general writing tasks. Schmitt and Schmitt (2014) also claimed that having good control of MFV is an essential component of English language proficiency. Furthermore, MFV was the largest spanning vocabulary category used in the present study. MFV included the 3K list to the 10K list, (9,000 word families), suggesting that MFV could have been counted more because such a large range of vocabulary fall in this category. The present study did not examine the somewhat at-risk group (between at-risk and not-at-risk). Meanwhile Douglas' sample included more mid- to higher-proficiency levels, who may have been more likely to use MFV (Nation, 2013).

Whereas Douglas (2015) found a small significant small positive correlation between LFV and test levels (r = .28), the present study did not (r = .018) for holistic scores. This finding is interesting because SAFE is an academic discipline-specific writing task that requires some engineering knowledge. Thus, the nature of the SAFE prompts includes specialized vocabulary, such as academic and engineering specific vocabulary, which are known to be low-frequency items (Coxhead & Nation, 2001; Nation, 2013). When examining the correlation between holistic scores and other measures of low-frequency vocabulary, such as the AWL and EEWL, there was a very similar correlation for academic vocabulary (r = .20) to Douglas (2015), r = .28, but not for the EEWL. Douglas (2015) did not include the AWL, and so it is difficult to compare.

An examination of the data found there were some gaps, between 1-10 word lists in relation to the lowest word list used (i.e., 15K). These gaps could be attributed to the discipline-specific vocabulary, creating a larger gap because it is low- frequency vocabulary. Nevertheless, the findings suggest that LFV may be unnecessary to write a successful text for a discipline-specific task, but it is important to include academic vocabulary although it may be considered infrequent. This finding highlights the importance of the academic genre and using academic vocabulary in order to write appropriately and successfully through the use of vocabulary (Bruce, 2008; Craswell & Poore, 2012; Paquot 2010).

One reason for the difference between the use of MFV and LFV in Douglas (2015) and the present study could be attributed to the differences in sampling methods. As previously mentioned, Douglas (2015) had a total of 200 written samples composed of 20 texts from each CELPIP-General test levels 3 to 12. CELPIP-General test levels 0, 1, and 2 were omitted due to lack of data. However, a lack of data can be a quantitative indicator of those in the lower levels lacking the ability to write enough, which the researcher may have considered as insufficient data. In addition, Douglas (2015) correlated the word frequencies to general levels while the present study used more precisely calculated holistic scores as a method of correlation and differentiated between writing proficiency levels. This could mean that Douglas' (2015) correlation results may have been larger because of the omission of data from lower proficiency students with lower scores.

Another potential reason for differences in the findings for HFV, MFV and LFV between Douglas (2015) and the present study may be the different test prompt requirements. The CELPIP-General test focuses on general workplace and community contexts and the writing prompts are general (write an e-mail or respond to an opinion survey). On the other hand, SAFE is an academic discipline-specific writing task that requires some engineering knowledge. Thus, the nature of the SAFE prompts includes specialized vocabulary, such as academic and engineering-specific vocabulary, which are known to be infrequent (Coxhead & Nation, 2001; Nation, 2013). This may have prompted writers to use specialized vocabulary and rely less on HFV or a more even distribution of HFV, MFV, and LFV, not predictable by correlation. Perhaps other kinds of vocabulary used may contribute to holistic scores for SAFE, such as academic vocabulary, as discussed earlier. This idea is supported by Laufer and Nation (1995) who point out that there are a variety of factors that contribute to lexical choice in writing, including “familiarity with the topic, skill in writing, and communicative purpose” (p. 308).

Although previous studies have found that frequency may contribute to productive skills (e.g., Douglas, 2015; Laufer & Nation, 1995; McNamara et al., 2010), more recent research uses other lexical sophistication indices and software, such as TAALES (Kyle & Crossley, 2015; Kyle et al., 2017). These studies have found that word frequency may not contribute as strongly to holistic writing scores as previous research claimed (Crossley et al., 2013; Laufer & Nation, 1995); instead, the researchers suggested that word range and *n*-gram indices may be more central. Although many of their frequency indices correlated with holistic scores, frequency was not the best predictor in comparison to the other highly correlated indices for a regression analysis. This idea is further supported by the present findings of no significant differences found between the not-at- risk and at-risk groups for frequency of vocabulary used. In the context of SAFE, frequency of vocabulary may not matter for holistic writing scores. The use of rare words is also unnecessary for a discipline-specific diagnostic assessment. Lastly, LFP may not be a good measure of language proficiency for discipline-specific tasks or for a mixed sample of native and non-native speakers. Therefore, other vocabulary characteristics must be considered for at-risk students to improve their writing.

4.2.5 Lexical Stretch (Douglas, 2010, 2013, 2015)

The present study examined the relationship between holistic scores and three Lexical Stretch measures by Douglas (2010, 2013, 2015): 1) 98% list coverage, 2) top list accessed, and 3) the number of lists accessed. In the present study, there was a trend for a significant small positive correlation between the three measures and holistic scores. Additionally, significant differences between the not-at-risk and at-risk groups were found for all three measures.

4.2.6 98% list coverage

The present study found a statistically significant small positive correlation between holistic writing scores and the lowest frequency list used to reach 98% coverage of a written text (r = .16), explaining 2.6% of the variance in holistic writing scores. This finding does not resemble Douglas' (2015) who found a significant large positive correlation between the two mentioned variables (r = .57), with 32% of the variance explained in CELPIP-General test levels. Douglas hypothesized that a higher proficiency writer may use vocabulary from low-frequency lists to cover 98% of a text. For example, a high-proficiency writer may use vocabulary from the 15K list so the 1K-15K lists would represent 98% of the lexical output. In comparison, a lower proficiency writer may use vocabulary from higher frequency lists (such as the 3K list). Thus, this writer used exclusively the 1K-3K lists which would represent 98% of the lexical output. Although Douglas (2015) hypothesized this, he only reported correlations which would not help verify such differences between higher and lower proficiency writers. The present study conducted a Mann-Whitney U test which compared the differences between the not-at-risk and at-risk group for the list used to reach 98% list coverage. The findings may help shed light on the differences in correlation between the present study and Douglas' (2015).

The present study found that the not-at-risk group (*Mdn* = 5) used a significantly lower frequency list to reach 98% list coverage in comparison to the at-risk group (*Mdn* = 4). Although these differences were significant, there were no large stretches between the lists in the data that resulted in notable differences in terms of frequency as Douglas (2015) had hypothesized previously. For example, both the not-at-risk (5K list) and at-risk group (4K list) on average used mid-frequency vocabulary to reach 98% coverage. This lack of difference in the use of frequency lists used to reach 98% coverage may have been expected because of the low correlation between LFV and holistic scores, as discussed earlier. An interesting note is that an analysis of the two DGITs (Car and Energy) found that knowledge of the 5K list was required to reach 98% coverage of the text. This may indicate that the 98% list coverage measure could be influenced by the task topic.

Douglas (2015) remarked that the lowest frequency list used to reach 98% may indicate that writers are using less frequent words in order to "deploy a wider variety of frequency bands [lists]" (p. 59). However, based on the results of the present study, this claim may not be true for discipline-specific assessments considering that the previous results in this thesis indicated the frequency of vocabulary—whether high, mid, or low—did not correlate with holistic writing scores. The present study did find that writers deployed a wider variety of frequency lists as Douglas (2015) sought to measure, but this evidence may be reflected in the word types correlation found in the present study (r = .51). This may suggest that the 98% coverage Lexical Stretch measure is tapping into a similar measure to word types.

When Douglas (2010) first introduced the lowest frequency list used by a writer to cover 98% of text, he discussed it as 99% explaining that native writers used more lower frequency lists to reach 99% coverage in comparison to non-native writers who used more higher frequency lists to reach 99% coverage. In Douglas (2013, 2015), this number changed to 98% coverage but does not quite focus on the applications. Although, Douglas did, however, discuss the lexical coverage figures in the context of the literature for comprehension. Douglas did connect the 98% coverage suggested for reading comprehension (Laufer & Ravenhorst-Kalovski, 2010) to receptive skills but did not quite make a clear connection as to how this number is transferable to written production. A couple of questions arise from this change. First, what are the advantages of shifting 99% to 98? Since 98% coverage indicates approximately 1 in 50 words (depending on the length of text), this is a small amount of words to account for, even for a text length of M =

519 words (Douglas, 2010; Douglas [2013, 2015] did not mention the mean length of the texts). In this example, 98% is accounting for approximately 10 words out of a text (500/50).

A second question that arises is, what are the implications of selecting 98% but not 95% or 100% lexical coverage? Previous research used the lexical coverage figures of 95%, 98%, and 100% for different types of texts such as spoken discourse (e.g., Nation, 2006), test items (e.g., Webb & Paribakht, 2015), television programs (e.g., Rodgers & Webb, 2011), all with the focus of comprehension. Hence, it could be speculated that 98% coverage may not apply to written production but to the comprehension of discourse.

4.2.7 Top list accessed

The correlation analysis conducted for the relationship between holistic writing scores and the top list accessed revealed there was a significant small positive correlation ($r = .20$), explaining 4% of the variance in holistic scores. This finding seems to suggest that low-frequency word usage may contribute to holistic writing scores. However, based on the previous results of the study (Section 3.3.6) this is not likely because of the lack of a correlation for LFV and actually, the lowest out of the three examined frequencies. This finding is in comparison to Douglas (2015) who found a medium positive correlation ($r = .45$), explaining 20% of the variance in CELPIP-General test levels. A potential reason for these differences may be attributed to the prompt and what type of vocabulary is required to answer an engineering specific prompt. Douglas' test prompt pertained to general topics, while the present study examined two prompts from the two engineering topics, Car and Energy. Because these topics are specialized, they contain discipline-specific vocabulary that is low frequency and therefore lower frequency lists may be used as the top list accessed. This may mean that regardless of the proficiency level of the student, the prompt demanded use of certain low-frequency lists.

Another finding in the present study indicated there were significant differences for the top word list used between the not-at-risk (*Mdn* = 9) and at-risk group (*Mdn* = 6). The results showed that not-at-risk students used vocabulary from lower frequency lists. As Douglas (2015) suggested, “lower-frequency lexical choices often equate with a more precise lexical choice…” (p. 59). This may be true based on the results seen here; it could be suggested that not-at-risk students are making more precise lexical choices.

A further analysis was completed to determine if the top list accessed was influenced by the test prompts. Within the DGIT-Car, it was found that the top list words typically came from the 15K list. Interestingly, the only words that came from the 15K list were “Celsius” and “vertex”, which came from the prompt. A closer look at the average top list accessed for the DGIT-Car found that both the not-at-risk and at-risk groups accessed the 15K list. This may be an indication that the top list used in writing is influenced by the prompt topic, regardless of writing proficiency level, rather than an indication of more precise lexical choices.

A similar result was found for the DGIT-Energy. The top list used in the prompt was the 7K list, but this was only one word (“utmost”). Meanwhile, the next top list that occurred in the Energy prompt was the 6K list with two words (“juggle” and “optimize”). Interestingly, the words that related the most to the prompt came from the 5K list, which would be the third top list accessed. This list contained five words that related the most to the prompt topic (“altitude”, “equilibrium”, “turbulence”, “ascend”, and “fluctuate”). A closer look at the average top list accessed for the DGIT-Energy found that both the at-risk and the not-at-risk groups accessed the 5K list. Thus, the topic may have an impact on the top list used by writers regardless of the proficiency level. This may further suggest that rather than more precise lexical choices being made, the top list may be influenced by the topic for discipline-specific related task topics. This may be because such topics possibly contain lower frequency (MFV and LFV) vocabulary because they are specialized.

4.2.8 Number of lists accessed

The results of the present study indicated a statistically significant small positive correlation between the number of frequency lists used in a text and holistic scores obtained (r = .28). The number of frequency lists used by a writer explained 8% of the variance. Similarly, Douglas (2015) found there was a significant medium positive correlation between the number of lists accessed and CELPIP-General Writing Test levels (r = .58), explaining 34% of the variance in CELPIP-General Test levels. These findings together may suggest that the diversity of vocabulary used, or different types of vocabulary may have an impact on holistic writing scores and may characterize writing proficiency levels. Perhaps the Lexical Stretch measure of the number of lists accessed is also tapping into a similar measure such as word types or diversity of words. This is further supported in the earlier findings of this study which found that word type usage was the highest correlated lexical measure to correlate with holistic writing scores, r = .51. Nation (2001) also expressed that writing texts which receive higher ratings contain more diverse vocabulary use. Douglas (2015) likewise suggested, "a fuller and more balanced access through the frequency bands [lists] goes hand in hand with increasing rater judgments of general English language proficiency" (p. 58).

Interestingly, for the present study, of the three Lexical Stretch measures, the number of lists accessed correlated the highest with holistic writing scores. Douglas (2015) suggested that the number of lists accessed is a greater indicator of increased writing proficiency in comparison to LFV. This claim holds true based on previous research and the present research results. Furthermore, Douglas (2015) suggested that low-frequency vocabulary was insufficient to differentiate lexical characteristics between high- and low-proficiency writers. Such findings demonstrate that not-at-risk students understand the value of diversity in word choice in their writing. This may indicate that not-at-risk students, who are successful writers, have a larger lexicon in terms of variety of words, which in turn allows them to minimize repetition in their writing and provide diversity in the manner they convey their writing. Not-at-risk writers may also make more concise lexical choices, as they have a more diverse lexicon. On the contrary, at-risk writers may have a smaller lexicon from which to choose words; thus, they rely on similar words, which may result in the reliance and repetition of vocabulary from their limited lexicon. Douglas (2015) likewise suggests that successful writers may use hypernyms, hyponyms, and synonyms and pronouns. A future study could investigate the extent to which the listed variables contribute to writing proficiency and perhaps use TAALES for confirmation.

4.3 Research Question 3a: Academic Vocabulary

In answer to the third research question, the results demonstrated that increases in academic vocabulary usage were associated with increases in holistic writing scores – a significant correlation of $r = .20$ was found. These findings are consistent with previous research that found a range of correlations from $r = .17$ to $r = .37$ (Higginbotham & Reid, 2019; Morris & Cobb, 2004; Nadarajan, 2011). The present findings shed light on the mixed results in the literature which found no correlation between academic vocabulary used and overall holistic scores (Csomay & Prades, 2018; Goodfellow et al., 2002). The results of the present study are also consistent with studies which also examined academic vocabulary in native and non-native writing for shorter texts ranging from 300 words or less (Morris & Cobb, 2004; Nadarjan, 2011). This is comparable to the present study which had texts ranging from 350 tokens or less. This suggests that even for shorter texts; the use of academic vocabulary is important for writing proficiency for both native and non-native writers.

The most similar and comparable study to the present study is by Higginbotham and Reid (2019) who found a correlation of $r = .37$ between the occurrence of the AWL and writing proficiency scores for science and engineering student texts. The present study had similar results as Higginbotham and Reid (2019) despite the differences in the average text length and participants. Higginbotham and Reid examined texts averaging 2000 tokens from non-native writers while the present study examined texts averaging 169.4 tokens from a mix of native and non-native writers. In spite of these differences, both studies arrived at similar conclusions. Perhaps the present study may have found a larger correlation if the text lengths were longer because this would provide more opportunity for students to use academic vocabulary. SAFE is a discipline-specific task which simulates future writing tasks engineering students will encounter in their course work and future careers. This means that in the future, engineering students will eventually write longer texts for other instructional classes or their career. When combining the findings of Higginbotham and Reid (2019) and the present study, it may emphasize the importance of academic vocabulary use for longer texts in engineering that first-year undergraduate SAFE students may eventually write.

Another finding in the present study found that not-at-risk students used a significantly higher proportion of academic tokens ($M = 8.21$) and types ($M = 10.60$) than at-risk students ($M = 7.18$, $M = 9.09$, respectively). This suggests that a characteristic of successful writing may be the increased use in the diversity and number of academic words. On the other hand, at-risk writing may be lacking these characteristics in their writing. The present study adds to the previous literature that found a relationship between writing proficiency and academic word usage as the present study found significant differences in the use of academic tokens and types between writing proficiency groups. These results may suggest that not-at-risk students have a better understanding of what is expected for an academic writing task and the genre as they have included words that occur frequently in academia to enhance and elevate their writing while the at-risk students seem to not understand this accordingly. For SAFE, this suggests that at-risk students should be informed about the use of academic vocabulary in writing and how this can support their writing skills.

The findings from the present study are further supported by the TAALES results. TAALES found a significant correlation (r = .57) between vocabulary analytic scores and the AWL index. The AWL index predicted 32% of the variance in analytic vocabulary scores. This result further emphasizes the importance of academic vocabulary use which in turn also impacts holistic writing scores. This result contrasts previous TAALES research (Kyle & Crossley, 2015; Kyle et al., 2017) which found that the AWL index did not contribute a particularly important role in predicting non-native writing proficiency. This contrast may be due to the differences in corpora examined. The present study used a corpus of student academic writing while previous studies have used free writing or narrative essays. The AWL may not be as relevant for less formal assessments such as free writing and narrative essays, which could have been expected as the AWL was derived from academic texts. The present study may be able to verify that the AWL may indeed be useful to support writing proficiency.

Previous studies have found significant differences in the use of academic vocabulary between native and non-native writers with a trend for native writers to use more academic words and obtain higher scores than non-native writers (Morris & Cobb, 2004; Nadarajan, 2011). These previous findings may be comparable to the present study in that at-risk student writing may resemble non-native writing in comparison to the not-at-risk students, because the present study found that at-risk students consistently used fewer academic words than not-at-risk students. It is a logical idea deeming that at-risk and non-native writers are still learning to develop and control English academic writing skills, which is a skill that is difficult for both native and non-native writers to acquire (Douglas, 2013; Morris & Cobb, 2004; Wood, 2015). Thus, suggesting for SAFE, that second language research may be able to inform support for less successful writing for both native and non-native writers who are ultimately less proficient in writing.

Morris and Cobb (2004) suggested that using vocabulary profiling to identify at-risk students may have helped diagnose some of their participants who had dropped out of the instructional class in their study. These at-risk students who had dropped out had an AWL percentage of under 5% in their written texts. In the future, perhaps SAFE can also use vocabulary profiling techniques to examine first-year engineering student writing for academic vocabulary. Presently, SAFE uses the analytic scores to calculate a holistic score and the rater provides an impressionistic categorization for at-risk, somewhat at-risk, and not-at-risk students in order to diagnose a student's writing and provide early intervention for at-risk students. Based on the findings of the present study, together with Morris and Cobbs' results, the percentage of AWL in writing could be an indication of an at-risk student. A small percentage of AWL in at-risk student writing could indicate these students may struggle in their future as academic vocabulary is an important part of academic writing (Coxhead & Nation, 2001; Coxhead, 2012).

The findings of the present study are contradictory to previous studies that found no correlation between academic word use and writing proficiency (Csomay & Prades; 2018; Goodfellow et al., 2002). This may be attributed to the prompts and the types of tasks in previous studies. Goodfellow et al. (2002) found no correlation between the use of academic words and writing proficiency scores when examining texts written about a general topic (the life of Quebecois firefighters). A reason for the differences in correlations may be because the present study used engineering specific texts in the academic writing context which would require some academic vocabulary. Meanwhile, Goodfellow et al. (2002) focused on a general writing topic which would not elicit academic vocabulary.

Similarly, Csomay and Prades (2018) found no correlation between academic vocabulary use and writing scores. Surprisingly, Csomay and Prades found that lower-level writers (M = 10.25) on average used a significantly greater amount of academic vocabulary than the higher-level writers (M = 8.60). These results contradict the present findings which indicated that lower-level writers used significantly fewer academic vocabulary than higher-level writers. One reason for this difference may be the disproportionate samples sizes in Csomay and Prades' (2018) study which included 154 lower-level and 37 higher-level texts. These differences in samples sizes, particularly for the higher-level samples, means that there were less chances for the use of academic vocabulary to occur.

Another possible reason for the differences in findings between the present study and Csomay and Prades (2018) may be due to the variation in text types analyzed. Csomay and Prades study did not compare the same text types for lower- and higher-proficiency writing. For example, the lower-level samples included response, comparative, argumentative, and exploratory syntheses texts. While the higher-level samples included rhetorical analysis and editorial texts. These are quite different types of writing that may or may not require academic vocabulary and therefore may not even be comparable. Csomay and Prades (2018) concluded that academic vocabulary is "solely dependent on the type of text they are used in" (p. 110). Furthermore, Csomay and Prades suggested that text type and rhetorical purpose, rather than just topic selection, contribute more greatly to the percentage of academic words used. This idea may be true as the present study found that academic words occurred in both at-risk and not-at-risk texts for the same text type. However, the present study does provide evidence that academic word use may also be dependent on writing proficiency.

4.4 Research Question 3b: Discipline-Specific Vocabulary

The next component of the third research question concerned the use of discipline-specific vocabulary use in at-risk and not-at-risk writing. The findings indicated that there was a non-significant correlation between holistic scores and engineering vocabulary. In addition, the not-at-risk and at-risk group used a similar amount of engineering specific tokens and types. This suggests that the use of discipline-specific vocabulary is not a characteristic that differentiates between writing proficiency levels. As stated previously, the research in the area of discipline-specific vocabulary usage and holistic scores is limited. However, the results of the present study may be comparable to Lessard-Clouston's (2012) case study which found no difference in the use of discipline-specific (theology) vocabulary between native and non-native writers. These findings together suggest that the use of discipline-specific vocabulary does not matter for writing proficiency because native writers are meant to represent a higher proficiency writers while non-native writers represent lower proficiency writers. This finding may suggest that other lexical characteristics contribute to writing proficiency scores.

A further analysis of the data indicated that at-risk and not-at-risk students are using similar amounts of engineering vocabulary in their writing regardless of the topic. For example, within both the Energy- and Car-DGIT, the at-risk (M = 8.84, M = 9.63, respectively) and not-at-risk (M = 9.96, M = 9.48, respectively) are using similar amounts of engineering vocabulary. This was unexpected considering in order to write to a genre, discipline-specific vocabulary must be used for the text to be considered proficient (Bruce, 2008; Swales, 1998). As stated earlier, previous research on the use of discipline-specific vocabulary is limited to Lessard-Clouston's (2012) case study which found no differences between the use of discipline-specific vocabulary between native and non-native writers. Using the logic stated previously regarding the at-risk group possibly resembling non-native writers, the results from the present study may be consistent with Lessard-Clouston who found no differences. However, considering that non-native writers have challenges with discipline-specific vocabulary when entering university (Evans & Morrison, 2010; Wu & Hammond, 2011), a future study may take a closer look at the production of the discipline-specific vocabulary in native and non-native writing to determine if there is a pattern within low- and high-proficiency groups.

Another potential reason for why there were no significant differences found between the at-risk and not-at-risk writers may be that writers, in general, may intuitively use engineering words because that is what the task topics are about. SAFE is an embedded task with a lecture supplementing the writing task. Students may have already been exposed to the engineering vocabulary and clarified their meanings, so the vocabulary could be used or was required to complete the task. Even though the task topics in the SAFE corpus are technically about different topics, Car versus Energy, both at-risk and not-at-risk students were answering the engineering prompt. To further support this point, after Hsu (2014a) compiled the EEWL from a corpus containing 100 college textbooks across 20 engineering subject areas, the list was tested for validity. She applied the EEWL to an engineering college textbook corpus and a general English textbook corpus. It was found that the EEWL provided 14.3% in the engineering textbook corpus while the same list only occurred in 2.22 % of the English textbook corpus. This suggests that texts about a specific topic will undoubtedly elicit or use vocabulary that is required to discuss the topic. In this case, the EEWL provided similar coverage figures for all texts, because of the topic rather than a function of writing proficiency. This implies that other lexical variables such as academic vocabulary may be more important for writing proficiency. This may further suggest that a crucial difference between at-risk and not-at-risk students may be the lack of academic vocabulary use, rather than engineering vocabulary. This may explain why there were more salient differences in the AWL compared to the EEWL investigation.

Another possibility for the lack of differences in the groups is because the test prompts are genre specific to engineering, any writer whether at-risk or not-at-risk will inevitably use engineering vocabulary to conform to the genre discipline-wise, but not academically. Additionally, this may be attributed to the fact the writing samples came from first-year students who, in secondary school, may not have felt the need to pay attention to or consider how to write academically and instead focused on engineering concepts. This idea is echoed in communication textbooks specifically designed for engineering student writing, “Many engineers and engineering students dislike writing” (Beer & McMurrey, 2014, p. 1).

There does not appear to be any previous research that has looked specifically at discipline-specific vocabulary use in relation to holistic writing scores. The present study adds to the limited research concerning discipline-specific vocabulary and writing proficiency. Also, this study contributes to the actual application of discipline-specific lists which are recommended for students to use. When researching the literature on the use of discipline-specific vocabulary, the focus was mostly on methodologies of how to compile such vocabulary lists that could be used for pedagogical purposes. There was a lack of studies that focused on the actual use of these lists in student writing and their impact on writing scores which could be compared to the present study.

4.5 Research Question 3c: Academic Formulaic Language

The present study found a statistically significant correlation between the use of academic formulaic language and holistic writing scores (r = .19). Most importantly, the not-at-risk students used significantly more academic formulaic language in their writing than at-risk students. Thus, suggesting that the use of academic formulaic language is a characteristic of successful writing. This may indicate that not-at-risks students are writing appropriately to academic prose because formulaic language occurs frequently in academic discourse (Biber et al., 2004; Wood, 2015). Not-at-risk students are able to demonstrate that they are a "member" of the discourse community (Byrd & Coxhead, 2010), and "guide readers through signaling linkage of ideas, the writer's stance, or attitudes implicit in prose" (Wood, 2015, p. 134) through the use of formulaic language.

Not-at-risk writers seem to understand how to use the appropriate formulaic sequence according to the functions served and write appropriately to the task and discourse while at-risk writers may not (Hyland, 2008, 2012). At-risk writers may be lacking an important aspect of academic prose which includes academic formulaic language to help structure texts and aid in the ability to introduce, provide details, and elaborate on their ideas in a coherent and organized manner (Biber et al., 2004; Hyland, 2008; Wood, 2015).

Another finding in the present study was that the not-at-risk group consistently used more academic formulaic sequences of varying word lengths (three-, four-, and five-words) than the at-risk group (see Table 14). These finding may suggest not-at-risk students are using longer and more grammatically accurate sequences of language in comparison to not-at-risk students who may still be focused on word-by-word sequences (Nattinger & DeCarrico, 1992; Wray, 2002).

However, it is important to note that there may be other academic formulas occurring in the SAFE corpus that were not counted because they were not included in the AFL. Furthermore, a large portion of the AFL contains three-word sequences (144 of the 200 sequences). This means there may have been other four- or five-word sequences that were underrepresented in the present study's results. Hyland (2008) argues that there are not enough formulaic sequences common to multiple disciplines to constitute a core academic phrasal lexicon. The AFL is a combination of academic formulas that occur in various disciplines and therefore are not specific to engineering but do occur in academia.

Considering that the use and knowledge of formulaic sequences provides a production advantage when writing texts (Byrd & Coxhead, 2010; Simpson-Vlach & Ellis, 2010; Wray, 2002), not-at-risk students may be able to produce ideas and language faster because they have long and accurate sequences of language ready to use for their writing. Thus, aiding in the overall assessment of writing quality. It could be suggested that at-risk writers may be lacking the awareness and skills to use academic formulaic language (Hyland, 2008), or lacking specialized knowledge of academic genres (Coxhead & Byrd, 2007). This is an important observation because according to Coxhead and Byrd (2007), academic writing (including for SAFE) requires control of grammar and vocabulary, proper knowledge of academic genres, and vocabulary to a specific field.

Formulaic language has been said to aid in the readability of texts (Nattinger & DeCarrico, 1992; Wray, 2002), this may contribute to the holistic and analytic scores raters give when using the SAFE rubric. Not-at-risk writers who used more academic formulaic language may achieve higher scores because they are conforming to the "strict expectations of the types of words and structures [in academic writing]" (Wood, p. 103). For at-risk writers, the lack of formulaic language and conforming to academic expectations for structure may decrease the readability and processing of the text. Thus, the inadequacy of the texts may become more salient.

The findings of the connection between academic formulaic language and writing proficiency may be supported by the results from TAALES. The results from the TAALES AFL measures indicated that there were significant small positive correlations between the analytic vocabulary scores for the Written AFL Normed and All AFL Normed indices. These are consistent with Kyle and Crossley (2015) who found small positive correlations between AFL indices and holistic scores.

However, the present study found there was no significant correlation between the analytic vocabulary score and the Core AFL Normed. The Core AFL results may emphasize that only the Written AFL should be prescribed because the Core AFL includes the overlapping formulas for both spoken and written language. These findings suggest that the Written AFL should be prescribed to at-risk writers as it supports the finding of the present study which found at-risk writers to use significantly fewer academic formulas in writing in comparison to not-at-risk writers.

In addition, the findings of a significant medium positive correlation for the COCA Academic Trigram Frequency suggest that the length of the formulaic sequence may contribute to writing scores. However, it is difficult to make strong conclusions because TAALES does not provide indices for longer word sequences such as four-, or five-word sequences, or more. Nevertheless, the present thesis did find that the not-at-risk group consistently produced a greater number of academic formulaic sequences of varying word lengths (three-, four-, and five-words) than the at-risk group.

The findings of the present study are consistent with previous studies (Ädel & Erman, 2012; Chen & Baker, 2010; Li & Volkov, 2018) which indicated that higher proficiency writers use more formulaic language in their writing. Previous studies typically equated proficiency as native or non-native writers, while the present study used precisely calculated holistic scores to quantify proficiency. Hence, confirming the importance of formulaic language for writing proficiency. Previous studies have examined the use of formulaic language in email tasks or published academic texts in disciplines including arts, humanities, and sciences. The present study adds to the growing body of literature as the findings indicate formulaic language use is associated with higher proficiency levels for engineering writing as well.

When comparing the present findings to previous research, they are contradictory to some previous research which has found low-proficiency writers to use more formulaic language than high-proficiency writers (Appel & Wood, 2016; Staples et al., 2013). A possible explanation for these differences may be because of the sample demographics. Previous studies examined formulaic language in TOEFL and CAEL, both standardized assessments of proficiency in English for non-native writers and found that lower proficiency writers relied on information and structures found in the prompt materials. This was attributed to the 'teddy-bear principle' (Hasselgren, 1994) where writers recycle words or phrases that they feel confident and knowledgeable using. Moreover, it was suggested that students used formulaic language as a strategy to manage their limited vocabulary. The present study examined a mix of native and non-native writers which were only differentiated by writing proficiency scores. A future study could compare the use of formulaic sequences according to language demographics and examine the use of relying on prompt materials in written texts to confirm this idea.

Appel and Wood (2016) have also examined formulaic language usage according to proficiency levels. However, it is difficult to compare their results to the present study for several reasons. Appel and Wood's (2016) sub-corpus of low-proficiency writing (n = 339) included a considerable difference in the number of tests compared to the higher proficiency sub-corpus (n = 254). In comparison to the present study, which had a similar number of at-risk (n = 183) and not-at-risk (n = 170) texts. In addition, Appel and Wood examined four- to seven-word formulaic sequences derived from the learner corpora while the current study examined three-to five-word recurrent sequences which were already identified previously based on academic texts.

Additionally, Appel and Wood (2015) identified the occurrences of formulaic language using the frequency-based approach and subjective judgment to select word sequences that are considered formulaic (see Section 1.4 for more detail). The frequency-based approach has been questioned by some researchers because some formulaic sequences that are identified are "open to claims of subjectivity" (Simpson-Vlach & Ellis, 2010, p. 490). Therefore, the formulaic language frequency counts provided by Appel and Wood may not be comparable to the present study. This is because the present study used the AFL as a consistent objective measure for the use of formulaic language in not-at-risk and at-risk texts.

The findings from the present study of higher proficient writers using more formulaic language in their writing are consistent with previous studies (Ädel & Erman, 2012; Chen & Baker, 2010; Li & Volkov, 2018), with further support from the TAALES indices (Kyle, Crossley, & Berger, 2017). Hyland (2012) notes, "the absence of such clusters [formulaic language] reveal the lack of fluency of a novice or newcomer to that community" (p. 165). Thus, suggesting that at-risk writing may be characterized by the lack of academic formulaic language. This indicates that it would be pedagogically useful to teach or encourage awareness of the AFL when aiding in the diagnosis and intervention of at-risk students for SAFE.

Teaching formulaic language awareness and uses can aid in the development of successful writing (e.g., Biber et al., 2004; Hyland, 2012; Wood, 2015). This is especially evident as the notion of highly frequent formulaic language use is important for writing appropriately to the register and discipline, especially in academia (Cortes, 2004; Hyland, 2008). Furthermore, increasing the knowledge and awareness of the AFL can free up cognitive resources for other language tasks, such as recall of propositional information (Nekrasova, 2009). Hence by having better control of the AFL, at-risk writers can focus on other cognitively demanding writing areas (Ellis, 2002, 2012) assessed by SAFE.

Chapter 5: Conclusion

5.0 Introduction

The findings from this thesis have important pedagogical implications for remedial writing instruction provided to at-risk engineering students. The pedagogical implications will be discussed within the context of the Self-Assessment for Engineers (SAFE) diagnostic assessment as the main purpose of this thesis was to inform the SAFE assessment procedure for academic support offered to students to help improve their writing skills. SAFE aids in the early intervention of helping students who are academically at-risk of failing the engineering program (Fox & Artemeva, 2017). Part of the intervention includes recommending students to visit the learning support center specifically designed to accommodate first-year engineering students. There, there are Linguistic Scholars who provide recommendations for how to improve at-risk writing. The present findings can inform the learning support center for recommendations to benefit the students who are at-risk. In addition, the present findings can advise assessment procedures for both the SAFE raters and automatic assessment procedures using the results of the comparisons between at-risk and not-at-risk writing.

5.1 Pedagogical Implications

5.1.1 Vocabulary Use and Holistic Scores

The findings from the present study indicated that there were significant differences found between at-risk and not-at-risk writing holistic scores, as expected. The strong positive correlation between analytic vocabulary scores and writing holistic scores found in this study suggested that vocabulary was a contributing factor to holistic scores. This implies that vocabulary knowledge and use should be focused on as part of the intervention for students who are at-risk. This is considering there were differences found in the use of vocabulary in the successful not-at-risk writing counterpart which would be the goal for unsuccessful at-risk writers. Linguistics scholars who are helping students at-risk can first raise awareness about the value of using appropriate vocabulary because of the connection to holistic scores for writing. Furthermore, vocabulary may be one of the first and least challenging components of writing at-risk students could develop as a foundation in comparison to rhetoric and logic. The ability to use vocabulary skillfully aids in the improved readability of written texts and aids in higher ratings of writing quality (Laufer, 1994; Nation, 2013).

5.1.2 Four Lexical Profile Characteristics to Consider

There were many lexical differences found between at-risk and not-at-risk writing in the present study. The various relationships and differences found using lexical sophistication measures can be reduced to four lexical profile characteristics associated with not-at-risk writing which can help inform at-risk writing. These four characteristics will be discussed in the order presented: length of texts, diversity of vocabulary, academic single words and formulaic language.

The findings from the present study indicated that there is a relationship between length of text and holistic scores. At-risk students can be advised by the Linguistic Scholars to write more. It should be noted that when at-risk students are advised to write more, they should be cautious to still be concise and not “fluff” their writing with more words but rather ideas. Writing more is a simple recommendation that should not only be used for assessment practices but also in general in order to develop the skill of writing. The recommendation of practicing writing can help at-risk students develop better writing habits and become more comfortable with the use of vocabulary to help develop and convey ideas (Graham, MacArthur, & Fitzgerald, 2013).

Considering that lower proficiency writing has been found to be associated with shorter text lengths, as indicated in the previous literature (Banerjee et al. 2007; Douglas, 2015; Frase et al., 1998) and the findings of the present study, it may be suggested that at-risk writers do not have much experience writing which may result in the production of shorter texts. Therefore, the pedagogical recommendations for at-risk students to practice writing and write more during assessments could be made.

The next main finding of the present study was the relationship between the increased types of words (diversity) and increased writing proficiency. Additionally, the findings from the Lexical Stretch measures seem to be underlying variables that measured the diversity of words in writing as well. Based on these findings, it could be suggested that at-risk students be made more aware of the importance of adding a diversity of words in their writing. It could be suggested that Linguistic Scholars introduce the concepts of hypernyms, hyponyms, and synonyms to at-risk writers.

Additionally, it could be advised that at-risk students minimize repetition in their writing. This can mean recommending the thesaurus and having students practice using it appropriately by providing exercises. For example, an exercise could use samples of not-at-risk writing and omit verbs or nouns. Then, at-risk students can either think of potential words that could replace or are synonyms to the omitted word. Or, at-risk students could examine the not-at-risk samples in which certain verbs or nouns are underlined then at-risk students could have a list of words they could practice choosing synonyms from. Another way to promote awareness and use is to provide hyponym and hypernym exercises in which students are tasked with a word and asked to provide options or select options from vocabulary provided. This could be completed with frequently occurring vocabulary in the SAFE assessments, academia, or engineering texts.

Based on the findings that academic vocabulary and formulaic language have been identified as a characteristic of successful writing, the AWL and AFL could be recommended to students at-risk. This could help promote an increase in academic vocabulary knowledge and use. For at-risk native-speakers, they may already know the vocabulary in the respective lists. In this case, it is important for them to be made aware of the AWL and AFL important uses and functions in academia as these are commonly used in the genre. As for at-risk non-native speakers, they may first need to learn aspects of the words such as form, meaning, and use (Nation, 2001, 2013) to able to appropriately integrate the vocabulary in their own writing.

In addition to increasing knowledge and awareness of the AWL and AFL, students also need to be able to practice and utilize such vocabulary in their own writing. One way of teaching the uses of the respective word lists may be through Computer Assisted Language Learning (CALL) or Mobile Assisted Language Learning (MALL). Pearson Education conducted a survey of university students' use of technological devices in 2015. Eight in ten (or 86%) students regularly use a smartphone while 89% regularly use a laptop on a regular basis (Pearson Student Mobile Device Survey, 2015). Considering how it has been reported that a large portion of university students use technology, the AWL (Coxhead, 2000) and AFL (Simpson-Vlach & Ellis, 2010) knowledge and usage can be increased though CALL or MALL for at-risk students. Of course, academic word knowledge and practice do not have to be limited to the 570 AWL and 200 Written AFL vocabulary. Other academic words and formula lists can be included as well. For example, parts of the AVL (Gardner & Davies, 2014), may also be included. The findings of the AWL and AFL as discussed, point to a larger picture of academic vocabulary and formulas occurring in successful writing. The AWL and written AFL could be suggested as a way of operationalizing this as they occur frequently in academia and are methodologically informed lists.

The writing samples from the at-risk and not-at-risk students can be used to inform remedial instruction for at-risk students. In combination with Nation's (2001) three processes for successful vocabulary learning, exercises can be designed for the learning and practice of using ideal vocabulary in writing. Nation (2001) outlines three processes of successful vocabulary learning that can be applied to teach academic words and formulas. The first process is noticing which involves drawing attention to the vocabulary of interest. The second, is retrieving which emphasizes repeat encounters of the word. The third process is generative output in which the previously encountered word is met or used in a slightly different manner or context. Throughout the three processes, there in an emphasis on input and output in alternative contexts. Integrating the processes outlined above, Figure 1 is an example of the type of exercise that can be recommended using the not-at-risk writing samples.

> "The graph demonstrates that there is no trend from sunrise to sunset, and that the height/ elevation of the plane needs to be adjusted to avoid turbulence." (Not-at-risk, Case #2001_SI)
>
> 1. Select the words that are considered academic
> 2. Think of synonyms for words in the sentence
> 3. Fill in the blank with options you considered
>
> "The graph___________that there is no____________from sunrise to sunset, and that the height/ elevation of the plane needs to be _________avoid turbulence."

Figure 1. Example of not-a-risk writing exercise that at-risk students can complete

First, at-risk students could be shown an excerpt from not-at-risk writing. Then, in the following order (1-3) at-risk students could be asked to notice the academic words and retrieve words that are synonyms. Finally, at-risk students could practice the retrieval of words in context. Additionally, this idea can be replicated using at-risk writing as seen in Figure 2. This exercise idea can also be used for formulaic language practice and use as well.

"The leaf **gives you** at least 95 kilometer on a full battery in the winter **but** that is **still better** than paying for gas and electricity on a hybrid car." (At-risk, Case #1042_VL)

1. Select the words that are considered not-academic or informal
2. Think of synonyms for the bolded words in the sentence
3. Fill in the blank with options

"The leaf__________at least 95 kilometer on a full battery in the winter _______that is_________________than paying for gas and electricity on a hybrid car."

Figure 2. Example of at-risk writing exercise that at-risk students can complete

These exercises can be completed using a traditional on-paper approach or through CAAL or MALL. Providing lexical academic support through CALL and MALL allows for vocabulary learning and use to be easily accessible from anywhere at any time encouraging students who do not have time to go to the academic support center to still have resources. González-Lloret and Ortega (2014) suggested that technology mediated tasks appear to reduce learners' anxiety in comparison to face-to-face communication. Furthermore, students may have difficulties seeking academic support because of the stigma or embarrassment associated with asking for help.

Combining CALL/MALL with Nation's (2001) three processes for learning vocabulary allows for an emphasis on receptive and productive learning which may aid in writers use of such vocabulary. This is important because learners need to be given sufficient opportunities for focused output (Nation, 2001; Byrd & Coxhead, 2010). Lastly, encouraging appropriate and accurate academic vocabulary and formulaic language use in writing indicates the writer's membership of a particular group (Corson, 1985; Wray, 2002) in this case, academic writing for engineers.

5.1.3 Assessment

The findings from this study have assessment implications for both human raters and computer assisted automatic assessment. The lexical characteristics from both groups of at-risk and not-at-risk writing may inform the characteristics of successful and unsuccessful writing to observe when assessing a text.

For human raters, when assessing a SAFE text for the analytic vocabulary score and holistic score, they can consider if the writing piece has included the correct use of academic words and academic formulaic language. Raters can be made aware of these features by being introduced to the AWL and AFL. Of course, they should also be aware that writing can have multiple combinations of lexical features that contribute to successful writing (Crossley et al., 2014; McNamara et al., 2010)

At the level of computer assisted automatic assessment (computers provide a score for a piece of writing using lexical characteristics), the present study has suggested that at-risk and not-at-risk writing can be differentiated using lexical features. For the future, SAFE may consider moving to computer assisted assessment in which students could type their answers. Perhaps computer assisted automatic feedback using the lexical characteristic results would help contribute to the diagnostic procedure. This could help with the 'divide and conquer' method used for SAFE, as mentioned previously (Chapter 3: Methodology). When first receiving the completed SAFE tests, the raters are asked to divide or categorize the tests as at-risk, somewhat at-risk, or not-at-risk. Human raters use intuition to accomplish this and then the at-risk tests are the focus of marking. A computer software such as AntWord Profiler or TAALES could process the texts using the lexical characteristic findings from this thesis, as supported by previous findings. Therefore, accelerating and strengthening the diagnostic process (Goodfellow et al., 2002). It could be suggested that the key lexical features to consider in writing are the length of the compositions, diversity of words (types), and the percentage of tokens that are academic and formulaic. To apply these features in the diagnostic procedure, an average based on the test cohort or a database of previous tests could be taken for the aforementioned measures. Then, each test could be compared to the average. Students who are above the mean would be considered as not-a-risk while students below are considered at-risk. It should be noted that automatic assessment is still an underdeveloped concept in the literature (e.g., Crossley, Cai & McNamara, 2012), so it could be suggested that automatic assessment co-occur with human raters to help strengthen and accelerate the diagnosis categorization process, especially for at-risk tests because such compositions may likely be deficient in the mentioned lexical features.

5.2 Limitations and Future Research

Some limitations and future research directions were discussed briefly throughout Chapter 4 and will now be discussed further. The first limitation is the approach to lexical frequency profiling using the categorization of HFV, MFV, and LFV. Depending on the cut-off range for each of the categories, the definition of sophisticated vocabulary also changes. For example, the present study used the 11K list and beyond as a definition of sophisticated LFV. It is unclear whether more LFV may have been found if the cut-off was at a higher frequency list such as the 10K or 9K list.

The second limitation is related to the two test prompts pertaining to different topics (Car and Energy) included in the SAFE corpus. It is unclear whether different prompt topics would elicit different vocabulary profiles that could impact the results of the present thesis. A future study may investigate the differences in at-risk and not-at-risk writing according to the prompt topic to better understand such lexical profiles. Another future study may also investigate the vocabulary in the prompts and examine the recycling of vocabulary from such prompts in student writing.

Another limitation is the word lists used to operationalize specialized and academic formulaic language. While the AWL, EEWL, and AFL are representative of specialized vocabulary, there may have been a greater amount of specialized vocabulary that occurred in the SAFE corpus that were not included in these respective word lists. However, these word lists were used because they are pedagogically useful and are representative of vocabulary that was considered sophisticated. There are other lists that could be used in future research to operationalize sophisticated vocabulary such as an opaque formulaic language list (Hsu, 2014b).

An additional limitation is that the lexical sophistication characteristics identified in the present study were derived from the engineering discipline. It is unknown whether the lexical sophistication characteristics found presently may differ according to disciplines. A future study may examine texts from other disciplines to determine if such differences exist.

Considering that there are a variety of ways to define lexical sophistication, other indices may be considered for future research when examining the characteristics of student writing. TAALES 2.0 (Kyle et al., 2017) includes over 400 lexical sophistication indices pertaining to *n*-grams, word range, psycholinguistic word information, age of exposure, and word recognition norms. A multiple regression analysis of the respective indices contribution to holistic scores using TAALES may provide new insights into other lexical characteristics that may contribute to holistic writing scores. Similarly, a logistic regression could be conducted to determine if such hindices could accurately classify at-risk and not-at-risk writers.

The present study differentiated the at-risk and not-at-risk group according to holistic scores, a future study may include demographic information such as the test takers' first language(s). It is unclear whether native and non-native speakers of English may have produced different lexical profiles and therefore such information should be considered when providing remedial instruction.

While the present thesis has contributed to a better understanding of the use of sophisticated vocabulary as a lexical characteristic of not-at-risk and at-risk writing, there are still many other lexical characteristics to consider that may have an impact on writing quality and remedial instruction for a diagnostic assessment for engineers.

5.3 Conclusion

The purpose of this thesis has been to investigate the relationship between vocabulary use and holistic scores for an English for Specific Purposes diagnostic test for engineering students. Furthermore, there was a comparison of the lexical characteristics of at-risk and not-at-risk student writing. The corpus analysis utilized and expanded upon earlier approaches to lexical sophistication using written texts. The findings provide insight into the lexical sophistication characteristics that describe at-risk and not-at-risk writing. More specifically, the frequency of vocabulary and use of discipline-specific vocabulary may not be important for characterizing writing proficiency on a diagnostic assessment. Conversely, increased use of the diversity of words, academic words, and formulaic language may contribute to writing proficiency.

The present study contributes to the literature as it found that vocabulary is an important factor for a disciplinary, ESP-based writing task meant to identify the needs of entering undergraduate engineering students. This thesis has also contributed to the lack of studies in the literature that focus on discipline-specific diagnostic assessments, which are increasing in popularity. Furthermore, this study has examined a discipline-specific task for engineering while providing the lexical characteristics of not-at-risk and at-risk writers. Thus, informing the needs and potential effectiveness of the individualized academic support for at-risk students for a successful academic career.

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for specific purposes*, *31*(2), 81-92.

Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from: http://www.laurenceanthony.net/

Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from: http://www.laurenceanthony.net/

Appel, R., & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: Differences between high-and low-proficiency levels. *Language Assessment Quarterly*, *13*(1), 55-71.

Arnaud, J. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate component tests. In P.J.L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 133–145). London: Macmillan.

Astika, G. G. (1993). Analytical assessments of foreign students' writing. *RELC journal*, *24*(1), 61-70.

Banerjee, J., Franceschina, F., & Smith, A. M. (2007). Documenting features of written language production typical at different IELTS band score levels. *IELTS Research Reports*, *7*.British Council/IELTS Australia.

Beer, D. F., & McMurrey, D. A. (2014). *A guide to writing as an engineer* (4th ed.). Hoboken, NJ: Wiley Publishing.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525-531.

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

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

Bonanno, H., & Jones, J. (2007). *The MASUS procedure: Measuring the academic skills of university students: A diagnostic assessment.* Retrieved from https://sydney.edu.au/stuserv/documents/learning_centre/MASUS.pdf

Browne, S., & Doyle, H. (2010). *Discovering the benefits of the first-year experience program for under-represented students: A preliminary assessment of Lakehead universities gateway program.* Toronto, ON: Higher Education Quality Council of Ontario.

Brynildssen, S. (2000). *Vocabulary's influence on successful writing*: ERIC Digest D157 (ERIC Document Service No. ED446339). Bloomington, IN: ERIC Clearinghouse on Reading, English, and Communication. Retrieved from http://www.eric.ed.gov/PDFS/ED446339.pdf

Bruce, I. (2008*). Academic writing and genre: A systematic analysis.* London; New York: Bloomsbury

Budinski, K. G. (2001). *Engineers' guide to technical writing.* Materials Park, OH: ASM International.

Byrd, P. & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL, 5,* 31–64.

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

Cheng, A. (2007). Transferring generic features and recontextualizing genre awareness: Understanding writing performance in the ESP genre-based literacy framework. *English for Specific Purposes, 26*(3), 287-307.

Cobb, T. (n.d.a). *Compleat Web VP! (an adaptation of Heatley & Nation's 1994 Range).* Montréal, PQ: University of Québec à Montréal. Retrieved from http://www.lextutor.ca/vp/comp/

Corson, D. (1997). The learning and use of academic English words. *Language learning*, *47*(4), 671-718.

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

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-238.

Coxhead, A. (2006). *Essentials of teaching academic vocabulary*. Boston, MA: Heinle Coxhead, A. (2012). Academic vocabulary, writing and English for academic purposes: Perspectives from second language learners. *RELC Journal*, *43*(1), 137-145.

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 specialized vocabulary of English for academic purposes. In J. Flowerdew, & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252-267). Cambridge: Cambridge University Press.

Craswell, G., & Poore, M. (2012). *Writing for academic success* (2nd ed.). London: Sage Publishing.

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*, 965–981.

Crossley, S. A., Cai, Z., & McNamara, D. S. (2012). Syntagmatic, paradigmatic, and automatic n-gram approaches to assessing essay quality. In P. M. McCarthy & G. M. Youngblood (Eds.), *Proceedings of the 25th International Florida Artificial Intelli- gence Research Society (FLAIRS) Conference* (pp. 214–219) Menlo Park, CA: AAAI Press.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, *32*, 1-16.

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication, 31*(2), 184-214.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing, 28*(4), 561-580.

Csomay, E., & Prades, A. (2018). Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes*, *33,*100-118.

Cumming, A., Kantor, R., Baba, K., Erdoosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in writing-only and reading-to-write prototype tasks for next generation TOEFL. *Assessing Writing, 10*, 5-43.

Cumming, A., Kantor, R., Baba, K., Erdoosy, U., Eouanzoui, K., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated tasks for the new TOEFL.* Princeton, NJ: Educational Testing Service.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor Corpus of English. *Literary and Linguistic Computing, 25*(4), 447–464. https://doi.org/10.1093/llc/fqq018

Douglas, S. R. (2010). Non-native English speaking students at university: Lexical richness and academic success. (Unpublished doctoral dissertation). University of Calgary, Calgary, Canada. Retrieved from: https://doi.org/10.1017/CBO9781107415324.004

Douglas, S. R. (2013). The Lexical Breadth of Undergraduate Novice Level Writing Competency. *Canadian Journal of Applied Linguistics, 16*(1), 152-170.

Douglas, S. R. (2015). The relationship between lexical frequency profiling measures and rater judgements of spoken and written general English language proficiency on the CELPIP-General Test. *TESL Canada Journal*, *32*(9), 43–64.

Eckstein, G., & Ferris, D. (2018). Comparing L1 and L2 texts and writers in first-year composition. *TESOL Quarterly, 52*(1), 137-162.

Elder, C., Bright, C., & Bennett. (2007). The role of language proficiency in academic success: perspectives from a New Zealand university. *Melbourne Papers in Language Testing*, *12*(1), 24-28.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4*(2), 139–155.

Engineers Canada (2017). *The Engineering Enrolment and Degrees Awarded Report: Canadian Engineers for Tomorrow, Trends in Engineering Enrolment and Degrees Awarded 2013-2017.* Retrieved from: https://engineerscanada.ca/publications/canadian-engineers-for-tomorrow-2017#acknowledgements

Espinosa, S.M. (2005). Can P_Lex accurately measure lexical richness in the written production of young learners of EFL? *Porta Linguarum 4,* 7-21.

Evans, S., & Morrison, B. (2010). The first term at university: Implications for EAP. *ELT journal*, *65*(4), 387-397.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: Sage Publishing.

Flowerdew, L. (2012). English for academic purposes. In C. Chapelle (ed.) *The encyclopedia of applied linguistics.* Oxford, UK: Wiley-Blackwell.

Fox, J. (2005). Rethinking second language acquisition requirements: Problems with language-residency criteria and the need for language assessment and support. *Language Assessment Quarterly 2*(2), 85-115.

Fox, J., Haggerty, J., & Artemeva, N. (2016). Mitigating risk: The impact of a diagnostic assessment procedure on the first-year experience in engineering. In J. Read (Ed.), *Post-admission language assessment of university students* (pp. 43-66). Cham, Switzerland: Springer International Press.

Fox, J., von Randow, J., & Volkov, A. (2016). Identifying students at-risk through post-entry diagnostic assessment: An Australasian approach takes root in a Canadian university. In V. Aryadoust, & J. Fox (Eds), *Trends in language assessment research and practice: Theview from the Middle East and Pacific Rim* (pp. 266-285). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Fox, J., & Artemeva, N. (2017). From diagnosis toward academic support: developing a disciplinary, ESP-based writing task and rubric to identify the needs of entering undergraduate engineering students. *ESP Today, 5*(2), 148–171. https://doi.org/10.18485/esptoday.2017.5.2.2

Frase, L. T., Faletti, J., Ginther, A., & Grant, L. (1998). Computer analysis of the TOEFL Test of

Written English. *ETS Research Report Series*, *1998*(2), i-26.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*, 305-327.

Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, *8*(1), 1- 26.

González, M. C. (2017). The Contribution of Lexical Diversity to College Level Writing. *TESOL Journal, 8*(4), 899-919.

González-Lloret, M., & Ortega, L. (Eds.). (2014). *Technology-mediated TBLT: Researching technology and tasks* (Vol. 6). John Benjamins Publishing Company.

Goodfellow, R., Lamy, M. N., & Jones, G. (2002). Assessing learners' writing using lexical frequency. *ReCALL*, *14*(1), 133-145.

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size ofr. *The Journal of Experimental Education*, *74*(3), 249-266.

Graham, S., MacArthur, C. A., & Fitzgerald, J. (Eds.). (2013). *Best practices in writing instruction*. New York: Guilford Press.

Harrington, M., & Roche, T. (2014). Identifying academically at-risk students in an English-as-a-Lingua-Franca university setting. *Journal of English for Academic Purposes*, *15*, 37-47.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*(2), 237–258.

Heatley, A., & Nation, P. (1994). *Range [computer software]*. Victoria: University of Wellington, NZ. Retrieved from http://www.vuw.ac.nz/lals/.

Higginbotham, G., & Reid, J. (2019). The lexical sophistication of second language learners' academic essays. *Journal of English for Academic Purposes*, *37*, 127-140.

Hsu, W. (2014a). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes, 33,* 53-64.

Hsu, W. (2014b). The most frequent opaque formulaic sequences in English-medium college textbooks. *System*, *47*, 146-161.

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. Abingdon: Routledge.

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

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, *32,* 150–169.

Hyland, K., & Tse, P. (2007). Is there an 'academic vocabulary'? *TESOL Quarterly, 41*(2), 235-253.

Jacobs, H.L., Zingraf, S.A., Wormuth, D.R., Hartfiel, V.F., and Hughey, J.B. (1981*). Testing ESL Composition, Rowley*. MA: Newbury House Publishers.

JASP Team (2018). JASP (Version 0.9.2.0) [Computer software]. Amsterdam, The Netherlands: University of Amsterdam. Retrieved from: https://jasp-stats.org/ s

Kaplan, R. M., Wolff, S., Burstein, J. C., Lu, C., Rock, D., & Kaplan, B. (1998). Scoring essays automatically using surface features. *ETS Research Report Series*, *1998*(2), i-12.

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical Sophistication as a Multidimensional Phenomenon: Relations to Second Language Lexical Proficiency, Development, and Writing Quality. *Modern Language Journal, 102*(1), 120–141. https://doi.org/10.1111/modl.12447

Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly, 49*(4), 757–786.

Kyle, K., Crossley, S. A., & Berger, C. (2017). The tool for the analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods, 50*(3), 1030-1046.

Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal, 25*(2), 21-33.

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. Applied Linguistics, *16*(3), 307–322.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15-30.

Lee, Y. W., Gentile, C., & Kantor, R. (2009). Toward automated multi-trait scoring of essays: Investigating links among holistic, analytic, and text feature scores. *Applied Linguistics*, *31*(3), 391-417.

Leech, G., Rayson, P., & Wilson, A. (2002). Companion website for word frequencies in written and spoken English: based on the British National Corpus. Retrieved from http://ucrel.lancs.ac.uk/bncfreq

Lemmouh, Z. (2008). The relationship between grades and the lexical richness of student essays. *Nordic Journal of English Studies*, *7*(3), 163-180.

Lessard-Clouston, M. (2012). Technical vocabulary use in English-medium disciplinary writing: A native/non-native case study. *Linguistics Journal*, *6*(1), 127-150.

Li, C. (1997). *Is lexical richness an essential criterion in judging a piece of writing?* (Unpublished Master's dissertation). The University of Hong Kong, Hong Kong, PR China.

Li, Z., & Volkov, A. (2018). "To whom it may concern": A study on the use of lexical bundles in email writing tasks in an English proficiency test. *TESL Canada Journal*, *34*(3), 54-75.

Lowie, W., & Seton, B. (2012). *Essential statistics for applied linguistics*. Basingstoke, UK: Palgrave Macmillan.

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.

McCully, G. (1985). Writing quality, coherence, and cohesion. *Research in the Teaching of English, 19,* 269–282.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication, 27*(1), 57-86.

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods, 45*(2), 499-515.

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing, 23*, 35-59.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malm- kjaer & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge, England: Cambridge University Press.

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26*, 32–47. doi:10.1093/applin/amh037

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review, 4*(1), 151-172.

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System, 32*(1), 75–87. https://doi.org/10.1016/j.system.2003.05.001

Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System, 30*, 225–235.

Nadarajan, S. (2011). The challenges of getting L2 learners to use academic words in their writings. *Electronic Journal of Foreign Language Teaching, 8*(2), 184-200.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening*? The Canadian Modern Language Review, 63,* 59-82

Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press

Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nekrasova, T. M. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language learning*, *59*(3), 647-686.

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (6th ed.). Maidenhead, Berkshire, England: McGraw-Hill Education.

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. London, UK: Bloomsbury.

Paquot, M. (2018). Phraseological Competence: A Missing Component in University Entrance Language Tests? Insights From a Study of EFL Learners' Use of Statistical Collocations. *Language Assessment Quarterly*, *15*(1), 29-43.

Pearson Student Mobile Device Survey. (2015). National report: College students. Retrieved from http://www.pearsoned.com/wp-content/uploads/2015-Pearson-Student-MobileDevice-Survey-College.pdf

Pinchbeck, G. G. (2017). *Vocabulary Use in Academic-Track High-School English Literature Diploma Exam Essay Writing and its Relationship to Academic Achievement* (Unpublished Ph.D. Thesis). University of Calgary.

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878-912.

Powell, P. R. (2009). Retention and writing instruction: Implications for access and pedagogy. *College Composition and Communication*, 664-682.

Prior, P. (1998). *Writing/disciplinarity: A sociohistoric account of literate activity in the academy.* London/New York: Routledge.

Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL quarterly*, *19*(2), 229-258.

Raimes, A. (2002). The steps in planning a writing course and training teachers of writing. In J.C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 306–314). Cambridge, UK: Cambridge University Press.

Read, J. (1998). *Validating a test to measure depth of vocabulary knowledge*. In A. Kunnan (Ed.), Validation in language assessment (pp. 41–60). Mahwah, NJ: Lawrence Erlbaum.

Read, J. (2000). *Assessing vocabulary.* Cambridge University Press.

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: Cambridge University Press.

Rhee, S. C., & Jung, C. K. (2014). Compilation of the Yonsei English Learner Corpus (YELC) 2011 and its use for understanding current usage of English by Korean pre-university students. *The Journal of the Korea Contents Association, 14*(11), 1019-1029.

Rodgers, M. P., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly, 45*(4), 689-717.

Roessingh, H. (2008). Variability in ESL outcomes: The influence of age on arrival and length of residence on achievement in high school. *TESL Canada Journal*, *26*(1), 87-107.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484–503.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40–59.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied linguistics*, *31*(4), 487-512.

Smith, C. (2003). *Vocabulary's influence on successful writing: ERIC topical bibliography and commentary* (ERIC Document Reproduction Service No. ED480633). Bloomington, IN: ERIC Clearninghouse on Reading, English, and Communication. Retrieved from http://www.eric. ed.gov/PDFS/ED480633.pdf

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12*(3), 214–225. https://doi.org/10.1016/j.jeap.2013.05.002

Stockwell, G. R., & Harrington, M. W. (2003). The incidental development of L2 proficiency in NS-NNS email interactions. *CALICO Journal, 20*(2), 337-359.

Swales, J. M. (1998). Textography: Toward a contextualization of written academic discourse. *Research on Language and Social Interaction*, *31*(1), 109-121.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using Multivariate Statistics* (5th ed.). Boston, MA: Pearson.

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical word list. *English for Specific Purposes, 27*, 442-458.

Webb, S., & Paribakht, T. S. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes*, *38*, 34-43.

Webb, S., & Nation, I.S.P. (2017). *How Vocabulary is Learned.* Oxford: Oxford University Press.

West, M. (1953). *A General Service List of English Words.* London, UK: Longman. Winsor, D. A. (2013). *Writing like an engineer: A rhetorical education*. New York, NY: Routledge.

Wood, D. (2015). *Fundamentals of formulaic language*. London; New York: Bloomsbury Academic.

Wray, A. (2002). *Formulaic language and the lexicon.* New York, NY: Cambridge University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication, 20*(1), 1–28. https://doi.org/10.1016/S0271

Wu, W., & Hammond, M. (2011). Challenges of university adjustment in the UK: a study of East Asian Master's degree students. *Journal of Further and Higher Education*, *35*(3), 423-438.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215-229.

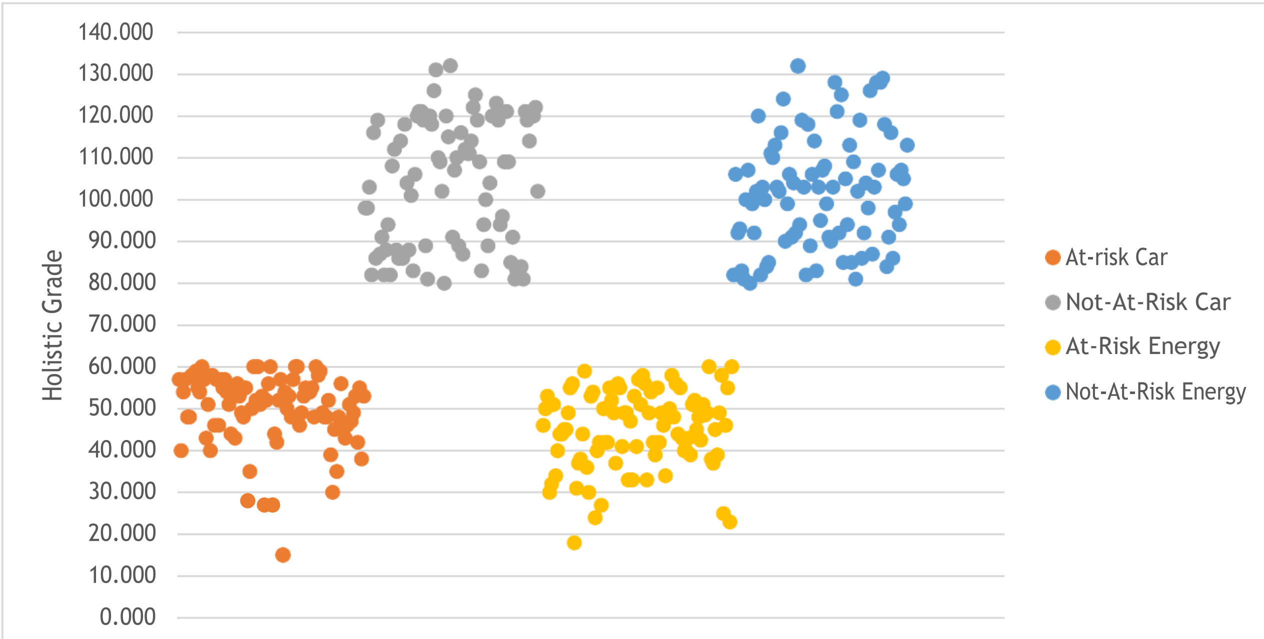
Yang, M. (2015). A nursing academic word list. E*nglish for Specific Purposes, 37*, 27-38

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*(2), 236-259.

Appendices

Appendix A

Distribution of Holistic Test Scores According to Test Versions



Appendix B

Occurrences of 3-word Academic Formulas for the at-risk group

Total number of formula types: 26 Total No. of Occurrences: 61

Rank	Freq	Range	Cluster
1	6	2	this means that
2	6	6	is affected by
3	4	4	are able to
4	4	4	depend (s/ing) on the
5	7	7	to the fact
6	3	3	which can be
7	2	2	can be seen
8	2	2	does not have
9	2	1	should not be
10	1	1	an attempt to
11	1	1	appears to be
12	1	1	are likely to
13	1	1	be explained by
14	1	1	is determined by
15	1	1	it is possible
16	3	3	they do/did not
17	1	1	to do so
18	1	1	two types of
19	3	3	as shown in
20	4	3	factors such as
21	2	2	the most important
22	1	1	could be used
23	1	1	has been used
24	1	1	it appears that
25	1	1	it is difficult
26	1	1	there are several

Appendix C

Occurrences of 4-word Academic Formulas Lists for the at-risk group

Total number of cluster types: 11 Total No. of Cluster Tokens: 31

Rank	Freq	Range	Cluster
1	9	9	the difference (s) between the
2	7	7	on the other hand
3	5	5	the other hand the
4	2	2	is that it is
5	2	2	in this case the
6	1	1	is based on the
7	1	1	it is obvious that
8	1	1	that there is no
9	1	1	can be seen in
10	1	1	as can be seen
11	1	1	it is clear that

Appendix D

Occurrences of 5-word Academic Formulas List for the at-risk group

Total number of cluster types: 2 Total No. of Cluster Tokens: 13

Rank	Freq	Range	Cluster
1	8	8	on the other hand the
2	5	3	due to the fact that

Appendix E

Occurrences of 3-word Academic Formulas Lists for the not-at-risk group

Total number of formula types: 42 Total No. of Occurrences: 117

Rank	Freq	Range	Cluster
1	10	10	depend(s,ing,ence) on the
2	9	9	this means that
3	8	8	needs to be
4	7	6	does not have
5	7	7	as shown in
6	7	7	factors such as
7	5	5	be explained by
8	4	4	are able to
9	4	4	is affected by
10	4	4	even though the
11	4	3	can be seen
12	4	4	appear (s) to be
13	3	3	which can be
14	3	3	they did/do not
15	3	3	two types of
16	2	1	at the time
17	2	2	it appears that
18	2	2	to the fact
19	2	2	are based on
20	2	2	could be used
21	2	2	if they are
22	2	2	should not be
23	2	2	there are no
24	1	1	be seen as
25	1	1	in the form
26	1	1	most likely to
27	1	1	on the part
28	1	1	take into account
29	1	1	there are several
30	1	1	this does not
31	1	1	was based on
32	1	1	we do not
33	1	1	wide range of
34	1	1	as a whole
35	1	1	been shown to
36	1	1	can also be

37	1	1	can be found
38	1	1	degree to which
39	1	1	does not appear
40	1	1	it has been
41	1	1	it is difficult
42	1	1	such as those

Appendix F

Occurrences of 4-word Academic Formulas Lists for the not-at-risk group

Total number of formula types: 15 Total No. of Occurrences: 43

Rank	Freq	Range	Cluster
1	7	7	on the other hand
2	5	4	it is important to
3	13	13	the difference (s) between the
4	2	1	be related to the
5	2	2	can be used to
6	1	1	is that it is
7	2	2	it is obvious that
8	1	1	it should be noted
9	3	3	it is clear that
10	2	2	it is possible that
11	1	1	as a result of
12	1	1	in this case the
13	1	1	it is impossible to
14	1	1	that there is no
15	1	1	the nature of the

Appendix G

Occurrences of 5-word Academic Formulas Lists for the not-at-risk group

Total number of formula types: 2 Total No. of Occurrences: 31

Rank	Freq	Range	Cluster
1	18	18	due to the fact that
2	13	13	on the other hand the