
Electronic Theses and Dissertations, 2004-2019

2013

The Intricate Relationship Between Measures Of Vocabulary Size And Lexical Diversity As Evidenced In Non-native And Native Speaker Academic Compositions

Melanie Gonzalez
University of Central Florida



Part of the [Education Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Gonzalez, Melanie, "The Intricate Relationship Between Measures Of Vocabulary Size And Lexical Diversity As Evidenced In Non-native And Native Speaker Academic Compositions" (2013). *Electronic Theses and Dissertations, 2004-2019*. 2633.

<https://stars.library.ucf.edu/etd/2633>



THE INTRICATE RELATIONSHIP BETWEEN MEASURES OF VOCABULARY
SIZE AND LEXICAL DIVERSITY AS EVIDENCED IN NON-NATIVE AND
NATIVE SPEAKER ACADEMIC COMPOSITIONS

by

MELANIE C. GONZÁLEZ

B.A. University of Central Florida, 2003

M.A. University of Central Florida, 2007

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the College of Education and Human Performance
at the University of Central Florida
Orlando, Florida

Summer Term
2013

Major Advisor: Keith S. Folse

ABSTRACT

The present study, a quantitative lexical analysis, examines the extent to which vocabulary size and lexical diversity contribute to writing scores on advanced non-native speakers' and native speakers' academic compositions. The data consists of essays composed by 104 adult non-native English learners enrolled in advanced second language writing courses and 68 native speaking university students in a first-year composition course. The lexical diversity of the sample essays is quantified by both the Measure of Textual Lexical Diversity (MTLD) and the *voc-D* while vocabulary size is measured by CELEX word frequency means, three instruments that are available in the computational linguistics program Coh-Metrix 3.0. Writing scores are provided by three independent raters' evaluations according to the TOEFL iBT Independent Writing Rubric.

Results from a binary logistic regression reveal that lexical diversity has a significantly greater impact on writing score than vocabulary size ($p < .01$). Nevertheless, a series of MANOVAs indicate that vocabulary size initially facilitates writing scores at the lower proficiency levels, but it is an essay's lexical diversity that promotes it to the higher score levels. Additional findings from the MANOVAs demonstrate that native speakers' profiles of lexical diversity and size are significantly different from their non-native peers ($p < .001$). The lexical profiles also differed significantly among the individual score levels of the TOEFL iBT rubric ($p < .05$). A final outcome from a Pearson's product moment correlation analysis shows that

vocabulary size has only a moderate relationship to lexical diversity, suggesting that variation of mid-range vocabulary may be more important to writing proficiency than the use of more sophisticated terms that occur less frequently in natural language.

Implications for practice suggest that it is not enough to simply teach vocabulary words in the L2 composition classroom, but also to guide learners in how to employ these words in a varied manner within their writing. Furthermore, the results of this study indicate that teachers should spend more time on helping students use medium frequency words along with synonyms of a similar frequency rather than teaching students infrequent vocabulary, which may appear to sound more advanced.

ACKNOWLEDGMENTS

As this rite of Ph.D passage comes to a close, there are many people who readily and kindly contributed to not just this study, but also my doctoral program and overall well being through these past few years.

First and foremost, my sincerest and heartfelt gratitude goes to Dr. Keith Folse for his unwavering support, guidance, and encouragement that truly helped me see this study through to its end. His absolute faith in my capabilities as a burgeoning researcher and professor truly warmed my heart. Dr. Folse's resolute dedication to the field and to his students is a source of never-ending inspiration for me. I could not have dreamed of a better mentor.

I am immensely grateful to Dr. Stuart Webb who challenged me to think of my research questions in multiple ways. Many aspects of this study were inspired by his research, and I feel truly honored for his service on this committee. I am fortunate to have had his direction in the formulation of this study.

Additionally, I must also send a huge thanks to Dr. Haiyan Bai and Dr. Florin Mihai for their integral membership in my dissertation committee. Dr. Bai's patient guidance as I maneuvered through the multiple statistical analyses aided me to problem-solve in order to find the best possible research design for my data. Likewise, I am very much grateful for Dr. Mihai's practical advice and assessment expertise without which this study could have been far more complicated than it needed to be.

I also would like to thank the many instructors and administrators who allowed for me to collect student essays from their institutions. The 183 learners themselves also deserve recognition and thanks for freely allowing me to use their written work for this study. Without their eager participation, this study would not have been possible.

Next, I must express my sincerest gratitude to the three raters who graciously volunteered their time to grade essays that were not even composed by their own students! It is no easy feat to evaluate and score so many essays, and they did not hesitate to agree to the task. I am truly humbled by their service and contribution of a most vital part of this study.

A special thanks goes to Alison Youngblood, my partner in doctoral student “crime” and biggest champion. It all started with an email inquiring, “are you going to do the Ph.D program?” Her vivacious nature, friendship, and sense of humor helped to carry me through this program. Most of all, I must thank her for always having my back when challenges arose.

My most earnest thanks are for my family and friends’ loving support, patience, and delivery of the occasional humorous Internet meme that kept my spirits bright. They forgave and stuck by me for all the times I had to sacrifice precious time with them in order to do something related to my studies, and for that, I am eternally grateful. Of this important group of people in my life, I first thank my husband, Matt Kantrowitz, for being my greatest supporter and source of comfort. He never wavered in expressing his pride and conviction that I would succeed in my doctoral study efforts. Additionally, I

must thank him for the all-important reminders to take breaks from my marathon study and writing sessions!

To my parents, Vicente Gonzalez and Jeany Todd Gonzalez, and my sister, Janelle, no words ever seem perfect enough to express my gratitude for their unconditional love and support. Their cheers have been the loudest whenever I experienced both achievements and doubts. They are truly anchors in my life.

A final thanks is reserved for my two dachshunds, Charlie and Sammie. They warmed my lap for many of the long hours in front a computer or a book. They were the most loyal friends and never failed to offer their companionship during this program and my life. They are forever in my heart.

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF TABLES	xi
LIST OF ACRONYMS	xiii
CHAPTER 1: INTRODUCTION	1
Background	4
Rationale	6
Research Questions	11
Hypotheses	12
Significance of the Study	14
Study Limitations	16
Study Assumptions	17
Definitions of Terms for the Study	18
Organization of the Study	20
CHAPTER 2: RESEARCH AND LITERATURE REVIEW	22
Defining Academic Language	23
BICS and CALP	23
Linguistic Dimension of Academic Language	24
Cognitive Dimension of Academic Language	25
Language Discovery Dimension of Academic Language	26
Academic Language and NNS learners	26
Defining and Measuring Lexical Knowledge	27
The Role of Language Proficiency in L2 Writing Ability	32
The Role of Lexical Knowledge in the Scoring and Evaluation of L2 Texts	34
Composition Rubrics	35
Empirical Studies	38
Lexical Features of Writing Quality	40
Empirical Studies	41
Vocabulary Size and Writing Quality	43
Lexical Richness Features and Writing Quality	44
Lexical Diversity	44
Lexical Sophistication	46
Lexical Control	48
Operationalizing Vocabulary Size and Lexical Diversity	50
Measuring Productive Vocabulary Size	52
Words to Know	53
Unit of Measurement	56
Testing Vocabulary Size	57
Testing Productive Vocabulary Size in L2 Writing	61
How Many Words Does a NNS Writer Need to Know?	69
Measuring Lexical Diversity	69
Type-Token Ratio	71
Voc-D	72

The MTLTLD	73
Vocabulary Size versus Lexical Diversity as Predictors of Writing Quality.....	75
Conclusion	76
CHAPTER 3: RESEARCH METHODOLOGY	78
Orientation to Research Design	79
Research Setting and Population	80
The Intensive English Programs	80
First-Year Composition	81
Research Population Determinations	82
Sampling Procedures	84
Sample Size Determinations	84
Recruitment of Participants.....	85
Non-Native Speaking Learners.....	85
Native Speaking Learners.....	85
Raters	85
Data Collection Procedures.....	86
Instrumentation	87
Coh-Metrix 3.0.....	87
TOEFL iBT Independent Writing Rubric.....	93
Research Questions.....	94
Hypotheses	95
Hypothesis for Research Question One	95
Hypothesis for Research Question Two	96
Hypothesis for Research Question Three	97
Data Analysis Procedures	99
Ethical Considerations	100
Conclusion	101
CHAPTER 4: RESULTS.....	102
Research Questions.....	102
Sampling Procedures	104
Descriptive Data Results.....	107
Description of the Sample.....	107
Initial Data Screening of the Independent Variables	110
Initial Data Screening of the Dependent Variable	111
Data Analyses for Research Questions and Hypotheses.....	112
Final Data Screening.....	113
Research Question One.....	114
Research Question Two	117
Research Question Three	117
Additional Analyses.....	121
Conclusion	123
CHAPTER 5: DISCUSSION.....	125
Purpose of the Study	125
Summary of the Findings.....	127

Results.....	129
Significance of the Findings	134
Limitations of the Study.....	136
Implications for Practice	136
Recommendations for Future Research	139
Conclusion	140
APPENDIX: IRB LETTER OF APPROVAL.....	142
LIST OF REFERENCES	144

LIST OF FIGURES

Figure 1. Bloom's taxonomy of cognitive domains	25
Figure 2. Orientation to research design	80
Figure 3. Lexical diversity by writing score	132
Figure 4. Means for vocabulary size by score level.....	133

LIST OF TABLES

Table 1 Nation’s Aspects Involved in Knowing a Word	29
Table 2 Vocabulary Knowledge Scale	30
Table 3 Writing Rubrics	37
Table 4 Number of Words for Language Skills and Coverage	44
Table 5 Lexical Impact on Writing Quality	49
Table 6 Word Lists	56
Table 7 Examples of Nation’s Vocabulary Levels Test	58
Table 8 Examples from Meara and Buxton’s Eurocentres Vocabulary Size Test	59
Table 9 Examples from Laufer & Nation’s Productive Vocabulary Levels Test	60
Table 10 Review of Measures of Productive Vocabulary Size	68
Table 11 Measures of Lexical Diversity	75
Table 12 Number of Essays by Language Designation	105
Table 13 Number of Essays by First Language	106
Table 14 Number of Essays by Genre	107
Table 15 Inter-Rater Reliability Analysis	107
Table 16 Text Length	108
Table 17 Lexical Diversity and Vocabulary Size	108
Table 18 Descriptive Statistics for Writing Score	109
Table 19 Number of Essays by Writing Score	110
Table 20 Data Screening for the MTLN, voc-D, and CELEX Frequency Scores	110
Table 21 K-S Tests of Normality for the MTLN, Voc-D, and CELEX	111
Table 22 Data Screening for Writing Score	112
Table 23 K-S Tests of Normality for Writing Score	112
Table 24 Correlations between the MTLN, Voc-D, and CELEX Frequency Scores	114
Table 25 Multivariate Analysis of Variance between of Lexical Diversity and Vocabulary Size by Language Designation	115
Table 26 Tests of Between-Subjects Effects for Lexical Diversity and Vocabulary Size by Language Designation	115
Table 27 Means for Language Designation	116
Table 28 Split-File Analysis of Between-Subjects Effects for Lexical Diversity and Vocabulary Size by Language Designation	116
Table 29 Correlations between the MTLN, Voc-D, and CELEX Frequency Scores	117
Table 30 Logistic Regression Predicting Likelihood of Scoring a 5	118
Table 31 Logistic Regression Predicting NNS Likelihood of Scoring a 5	119
Table 32 Multivariate Analysis of Variance of Lexical Diversity and Vocabulary Size by Writing Score	120
Table 33 Tests of Between-Subjects Effects for Lexical Diversity and Vocabulary Size by Writing Score	120
Table 34 Means for Writing Score Levels	121
Table 35 Correlations between Lexical Diversity and Vocabulary Size with Each Writing Score Level	122

Table 36 Correlations between Lexical Diversity and Vocabulary Size with Each Writing Score Level by Language Designation123

LIST OF ACRONYMS

- *CELEX* – Title of word frequency mean scores
- *EAP* – English for Academic Purposes
- *EFL* – English as a Foreign Language
- *ELT* – English Language Teaching
- *ESL* – English as a Second Language
- *IELTS* – International English Language Testing System
- *IEP* – Intensive English Program
- *IRB* – Internal Review Board
- *L1* – First Language
- *L2* – Second Language
- *NNS* – Non-Native Speaking
- *NS* – Native Speaking
- *MTLD* – Measure of Textual Lexical Diversity
- *Voc-D* – Vocabulary Diversity (computational linguistics program)
- *SPSS* – Statistical Package for the Social Sciences
- *TOEFL* – Test of English as a Foreign Language
- *WAC* – Writing Across the Curriculum

CHAPTER 1: INTRODUCTION

The United States college classroom continues to undergo a transformation from a relatively homogeneous and monolingual environment to one that serves students from a variety of ethnic, cultural, and linguistic backgrounds. For the past century, an increasingly diverse population of students have sought out and matriculated into American institutions of higher education as post-secondary study became less exclusive (Institute of International Education, 2011; Matsuda, Cox, Jordan, & Ortmeier-Hooper, 2006). Included in this demographic change are non-native English-speaking (NNS) learners, students for whom English is not their primary or first language (L1) but a second language (L2). More than just foreign students learning English, NNS learners include international visa students, naturalized and native-born permanent U.S. residents, and refugees from a variety of linguistic backgrounds. According to the 2010 U.S. Census Bureau's American Community Survey of the college-age demographic (i.e., 18 and older), 55.4 million, or 14 percent of the total U.S. population, spoke a language other than English in the home (U.S. Census Bureau, 2011). Of this large number, nearly 1.7 million were enrolled in an institution of higher education, comprising almost 12 percent of the total student population (2011).

As a result of this demographic shift, many higher education institutions have added initiatives to increase diversity-related awareness within the faculty, staff, and student populations (Matsuda et al., 2006). The student population has changed so dramatically that education institutions are thinking about how they might change to

better serve today's student population. Faculty members have also begun to recognize a need to consider the cultural and linguistic needs of NNS learners in their instructional practices, particularly in regard to the academic language demands of college-level coursework (Ferris, 2009; Harklau, Siegal, & Losey, 2009). Despite the growing awareness of L2 issues in the college classroom, many college instructors feel ill equipped to adequately address the needs of NNS learners (Ferris, Brown, Liu, & Arnaudo Stine, 2011). This inadequacy perhaps stems from the ethical and political dilemmas regarding academic expectations raised due to NNS learners' presence in the college classroom (Harklau et al., 2009). For example, instructors and specialists have questioned the equity and appropriateness of holding native English-speaking (NS) and NNS learners to the same academic standards given that proficiency in English is a key criterion for entry into higher education (Harklau et al., 2009; Silva, 1993). Despite this language requirement, NNS learners with varying linguistic skills find their way into the mainstream classroom either via graduation from K-12 schools, the obtainment of international student visas, intensive English program (IEPs) completion, or by satisfying college Test of English as a Foreign Language (TOEFL) and/or International English Language Testing System (IELTS) score requirements (Ferris, 2009).

However NNS learners satisfy the benchmarks to enroll in U.S. colleges and universities, even those with high levels of English proficiency find that undergraduate and graduate coursework expectations still present major challenges for them, particularly in their written compositions (Casanave & Hubbard, 1992). Common college-level writing tasks are challenging for any college student, requiring learners to

produce sizable compositions such as reflections, critiques, literature reviews, and research papers (Ferris, 2009; Knoblauch & Matsuda, 2008). In view of factors relating to cultural, rhetorical, and linguistic differences, NNS learners find the experience of striving to meet the academic expectations of these language-intensive writing tasks to be much more daunting than their NS peers.

This problem is intensified when raters differentially weight these factors when assessing NNS learners' compositions (Huang, 2008; Huang, 2009; Huang & Foote, 2010; Roberts & Cimasko, 2008). Studies comparing raters' scoring of NNS and NS compositions show that raters are more likely to penalize texts produced by NNS writers based on linguistic issues such as vocabulary and grammar whereas NS texts tend to receive lower scores based on content and rhetorical style (Sweedler-Brown, 1993; Weigle, Boldt, & Valsecchi, 2003). Consequently, linguistic proficiency is clearly a crucial factor leading to successful NNS composition in the college classroom.

Given the impact NNS writers' linguistic proficiency has on their ultimate writing score, there is a need to better understand the various linguistic issues that contribute to writing quality and/or deficiency. Research into what types of language errors affect raters' perceptions of L2 compositions has revealed a hierarchy of error gravity that elucidates which linguistic forms have the greatest influence on writing quality. Of all the possible components impacting writing scores, lexical richness issues in L2 writing such as poor word choice and lack of lexical diversity claim the top hierarchical level as the errors most bothersome to raters (Agustín Llach, 2011; Chastain, 1980; 1981; Dordick, 1996; Ferris, 1994; Santos, 1988). Lexical concerns such as these frequently

occur in L2 compositions as a result of gaps in learners' NNS lexicon and prevent learners from being able to produce the diverse range of vocabulary that enriches writing quality. To better address the needs of NNS college learners, it is therefore advantageous to examine the impact of learners' vocabulary size and lexical diversity on writing proficiency.

Keeping this research purpose in mind, this chapter first elucidates the factors influencing the present study and the general state of knowledge surrounding the intersection of vocabulary and composition. Together, these sections form the background to the study. Next, the rationale and purpose for the study are presented followed by the research questions and corresponding hypotheses. The limitations and assumptions of the research are discussed in the next section. Lastly, the chapter concludes with the provision of definitions of key terms and the organization of the remaining dissertation.

Background

For decades, writing in the L2 classroom was the means to practice linguistic form, namely grammar and vocabulary but especially grammar. This practice was congruent with the longstanding grammar-translation method of language instruction (Ferris, 2009). However, with the advent of contrastive analysis approaches in the late 1970s, a pattern-approach to L2 writing took root where NNS writers were instructed to emulate English paragraphs. Not long afterwards, L2 writing instruction adopted the process-oriented technique, which focused on drafting, revising, and editing and remains

a popular approach to writing instruction (Silva, 1992). Current writing practices advocate a genre-analysis approach that incorporates the emulation of the structure, rhetorical devices, and linguistic features of specific writing genres (Johns, 2008). This shift back to a more form-focused approach to L2 composition practices stresses the importance of using the language characteristic of individual genres with a particular emphasis on its lexical items (Silva, 1992).

This resurgence of interest in linguistic form in composition theory is a result of recognizing that NNS writers, even those with advanced levels of proficiency, continue to struggle with grammatical and lexical items in production (Ferris, 1995; Silva, 1993; 1993). Although writing quality is judged from a variety of different perspectives, the central criterion remains the writer's ability to successfully convey meaning as evidenced by holistic scoring methods (Cumming, Kantor, Baba, Erdosy, Eouanzoui, & James, 2005; Hawkey & Barker, 2004; Huang & Foote, 2010; Janopoulus, 1993). However, problems arise when grammatical and lexical issues obscure meaning, thereby impeding communicative effectiveness. Accordingly, most rubrics utilized to evaluate and score learner writing contain scoring criteria relating to grammar and lexis (Agustín-Llach, 2011). Studies describing L2 composition have formed a consensus that the presence of syntactic and lexical errors in NNS essays is common and such errors often overshadow other textual features to impact score (2011; Engber, 1995; Silva, 1993). In terms of the gravity of linguistic errors impeding communicability, lexical errors have proven most unforgivable (Santos, 1988). Consequently, researchers and instructors have renewed

interest in explaining and analyzing vocabulary in writing with hopes of effectively preventing or appropriately treating lexical errors.

Rationale

With the writing-across-the-curriculum (WAC) movement of the past three decades, the majority of college-level tasks and assignments are dedicated to writing tasks (Kinneavy, 1983; Leki & Carson, 1994). These tasks usually intend for the learners to demonstrate the state of their knowledge or viewpoints on a certain topic. Surveys of NNS learners in the college classroom cite writing as being their biggest challenge in the college classroom (Burke & Wyatt-Smith, 1996). These nonnative learners state that writing is a slow process, requiring more time and effort than their NS peers due to translation from their L1 to English and the heavy use of a bilingual dictionary to find appropriate words (1996; Silva, 1992; Wang & Wen, 2002). Moreover, NNS learners feel that their limited L2 lexicon hinders their ability to adequately express their ideas, opinions, and points thus leading to shorter, more error-filled texts than their NS counterparts (Silva, 1992). Therefore, NNS writers are very much aware of the marked differences between their compositions and those by NS writers.

For many faculty members working with NNS learners, appropriately addressing and adjusting their instructional practices to handle these differences between NS and NNS compositions is a daunting and overwhelming notion (Ferris, 2007; 2011). Many college instructors have limited experience with and awareness of L2 writing evaluation and fallback to methods of grading monolingual L1 writing (Ferris, 2007). Analyses

comparing L1 and L2 writing reveal that these two groups of writers commit different types of errors thereby requiring different treatments (Ferris, 2011; Silva, 1993). Ferris (2007) describes the problem of non-ESL specialists face when evaluating L2 writing alongside L1 essays. She states “ESL language features must be put into their proper perspective when weighed against other strengths and weaknesses in student texts” (2007, p. 65). While L1 essays may have stylistic or content issues, L2 writing contains linguistic features related to NNS learners’ developing L2 acquisition, which can be a lifelong process for even those with advanced L2 proficiency (Silva, 1993).

As a result, college instructors’ reactions to L2 writing are highly variable. Research reveals that instructor feedback to L2 writing runs the gamut between being overly strict and overly lenient (Ferris, 2007; 2011). In other words, grading and treatment practices may penalize NNS learners for every linguistic mistake or award credit for L2 writing that is quite unsound. Since studies show that L2 writing differs from L1 writing in terms of grammar and lexis (Silva, 1993), teacher feedback must address these linguistic issues in a manner more helpful for NNS learners’ L2 development.

Lexical errors that impede comprehensibility the most are the most troubling for instructors to deal with because it is difficult to correct these types of errors (Ferris, 2007; Folse, 2008). Research indicates that vocabulary issues are often untreatable and are a result of partial L2 vocabulary knowledge and a small vocabulary size (Ferris, 2011; Schmitt & Zimmerman, 2002). Compared with NS equivalents, NNS learners lack sufficient vocabulary knowledge to meet the academic demands of the college classroom.

Unlike learning L2 vocabulary, L1 vocabulary acquisition progresses longitudinally. Estimates report that the average NS acquires roughly 1,000 word families every year with a plateau around the age of twenty (Nation, 2006). Thus, the lexicon of an average adult native speaker contains approximately 20,000 word families, or 70,000 words (Goulden, Nation, & Read, 1990). In comparison, NNS learners grow their lexicon at the same rate as NS, but they begin L2 vocabulary acquisition at a much later time and therefore their depth of vocabulary knowledge varies greatly (Schmitt, 2000). Clearly, many NNS learners are at a lexical disadvantage in college classrooms at the beginning of their learning, a lexical gap that widens even more by the end of their studies (Laufer, 1997).

Although there has been much more attention given to importance of vocabulary in the learning of an L2 within the body of second language acquisition (SLA) and teaching research, explicit focus on vocabulary has yet to translate to the classroom (Folse, 2004). Perusal of courses, programs, and materials used to prepare NNS learners for the English-speaking college classroom reveals that grammar continues to dominate curricula (Folse, 2010). When vocabulary is addressed, it is likely to depend on the teacher's affinity for and willingness to teach vocabulary (2010) or occur in the reading course due to an abundance of research linking receptive vocabulary knowledge to reading skills (Chung & Nation, 2003; Cobb, 2008; Grabe & Stoller, 2002; Hirsh & Nation, 1992; Horst, Cobb, & Meara, 1998; Hu & Nation, 2000; Laufer & Sim, 1985; Nation, 2006; Nation & Wang, 1999).

Consequently, regulating vocabulary instruction to the reading classroom implies that new words are only encountered receptively in texts. However, the ability to accurately and purposefully produce a lexical item in written discourse is one of the most linguistically challenging skills for NNS learners, requiring a complex network of lexical knowledge and leads many learners in the classroom to consult dictionaries or ask vocabulary questions during the composition process (Henriksen, 1999; Nation, 2001; Read, 1993). Therefore, given that the facility to vary lexis allows NNS learners to more precisely and richly express ideas in writing, explicit attention to productive vocabulary needs to be front and center in the composition classroom and not regulated only to the receptive context of reading (Folse, 2010).

Presently, the state of the research indicates that lexical proficiency correlates well with writing quality (see Alderson, 2005; Astika, 1993; Engber, 1995; Hawkey & Barker, 2004; Laufer, 1998). NNS students in the college classroom are faced with substantive course assignments and are expected to meet the same standards as their NS peers who have the lexical advantage of their L1. Academic literacy studies have embraced the benefits of a large lexicon for L2 reading proficiency in terms of text coverage (Nation, 2006; Laufer & Ravenhorst-Kalovski, 2010). Hence, there is an underlying assumption that the same notion applies to writing proficiency: L2 texts that demonstrate a larger vocabulary size are more deemed proficient (Laufer & Nation, 1995).

However, the body of literature has yet to clearly and empirically validate the claim that vocabulary size is the key lexical predictor of L2 academic writing success.

Unlike L2 reading studies, research into productive lexical thresholds has not yet been able to pinpoint a critical number of words NNS writers need to possess to write well for academic purposes (Nation & Webb, 2011). Instead, research has indicated that lexical diversity (i.e., using varied vocabulary choices) significantly impacts academic writing quality (Engber, 1995; Laufer, 1994; Linnarud, 1986). Determining whether it is the sophisticated breadth of learners' lexicon or their ability to purposefully utilize the words they already possess within their bank of words or a combination of both answers the pedagogical question of how best to design a lexical syllabus for advanced L2 writing courses. Hence, investigating learners' vocabulary size and variation in relationship to academic writing tasks is of value to language research and teaching.

Though there is a clear need for comparing the individual effects of specific measures of lexical size versus variation on academic writing ability, relatively few quantitative studies have explicitly and solely investigated this intersection. Past and present studies have tended to be more exploratory or global in nature with aims to profile the productive lexical features of L2 writings. Moreover, the comparison of lexical richness measures between authentic essay assignments composed by advanced NNS and NS writers has received little attention despite the fact that NNS learners are often held to NS standards within the college classroom. Furthermore, there are indications that it is not always a writer's ability to use a large number of low frequency, difficult, and rare words within texts that leads to writing proficiency (Laufer, 1994). Instead, clear and accessible writing may rely more on the writer's ability to employ a wide range of high frequency words in a more varied and stylistic manner. Since NS

writers tend to possess a larger lexicon than NNS writers at the onset of post-secondary study, comparing advanced NNS writers' lexical size and diversity profiles to NS equivalents may shed light on this lexical dilemma. As a result of this gap in the literature, the present study aspires to compare the effect of productive vocabulary size and lexical diversity on the writing ability of both advanced NNS learners and NS learners.

Research Questions

Given that a more detailed profile of the individual characteristics of lexical richness in NNS and NS writing has yet to be determined in the body of literature, this study aims to answer the following main research questions:

1. Is there a significant difference between advanced NNS learners' and NS learners' measures of vocabulary size and lexical diversity (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser, McNamara, Louwerse, M., & Cai, 2004; the *voc-D* and MTLD in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) as evidenced in academic writing?
2. Is there a relationship between vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLD in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) as evidenced in NNS and NS essays?

3. Is vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser, et al., 2004) or lexical diversity (as measured by the *voc-D* and MTLD in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) a greater predictor of writing score achievement (as measured by the TOEFL iBT Independent Writing Rubric; ETS, 2005) in non-native and native speaking college writing?

Due to the entwined nature of lexical knowledge, the ensuing hypotheses were formulated.

Hypotheses

The null and directional hypotheses for the first research question are as follows:

- H_0 : There is no significant difference between the vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLD in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) of advanced NNS learner and NS writing.
- H_1 : The vocabulary size (as measured by the CELEX raw frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLD in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) of NS compositions will be greater than the same measures in advanced NNS writers' compositions.

For the second question, the null and directional hypotheses are listed below:

- H_0 : There is no significant relationship between vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLT in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) as evidenced in NNS and NS essays.
- H_1 : Vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) significantly contributes to lexical diversity (as measured by the *voc-D* and MTLT in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) as evidenced in NNS and NS essays.

The null and directional hypotheses for the third and principle research question are the following:

- H_0 : Neither vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) nor lexical diversity (as measured by the *voc-D* and MTLT in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) is a significant predictor of writing score (as measured by the TOEFL iBT Independent Writing rubric; ETS, 2005) in non-native and native speaker college writing.
- H_1 : Learners who demonstrate a larger vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) will receive higher writing scores (as measured by the TOEFL iBT

Independent Writing rubric; ETS, 2005) in non-native and native-English speaking college writing.

- H₂: Learners who demonstrate greater lexical variation (as measured by the *voc-D* and MTL_D in Coh-Metrix; Malvern & Richards, 1997; McCarthy & Jarvis, 2010) will receive higher writing scores (as measured by the TOEFL iBT Independent Writing rubric; ETS, 2005) in non-native and native-English speaking college writing.

Further explication and support for each hypothesis is provided in Chapters Two and Three.

Significance of the Study

The main principle that formed the basis of this study is that particular lexical richness features must be present in learner language in order to successfully perform academic writing tasks proficiently and fluently (Laufer & Nation, 1999). Since the skill of producing a word's form, meaning, and use accurately in writing requires a high level of depth of word knowledge, it is useful to examine and compare how learners at the advanced and proficient levels of English use words appropriately and inappropriately in a given context. Therefore, the present study is meaningful for the three reasons delineated below.

First, it is rare that a NNS learner acquires the equivalent lexicon size to that of a NS peer. The research into vocabulary thresholds for language skills reveals the wide gap between NNS learners' and NS' vocabulary sizes. Looking at the productive lexical

size of NS college academic writing can serve as a baseline for future research and vocabulary development in IEP course syllabi.

Second, given the concerns researchers, college-level instructors, and NNS learners themselves have over the quality of L2 writings, studying the intersection of vocabulary measures and writing quality can provide insight in to the linguistic demands of college composition tasks and what professors expect to find lexically. Profiling learners' vocabulary size and variation could therefore serve as a diagnostic test for appropriate levels placement and writing interventions, potentially providing a more accurate judgment learners' L2 writing ability. Furthermore, IEPs and other ESL programs preparing NNS learners for U.S. postsecondary study need to be accountable for their L2 graduates' transition into the mainstream college classroom. Therefore, ESL instructors need to be aware of the various vocabulary features and expectations of the academic writing tasks learners will encounter outside of the ESL program. Thus far, little research to date has investigated how prepared NNS writers will be able to meet the productive lexical demands of post-secondary composition tasks.

Lastly, it can be reasoned that there is a solid justification for including vocabulary instruction within the ESL writing classroom. A variety of vocabulary measures are available for instructors to see if particular aspects of vocabulary knowledge are being overlooked in learner compositions. For example, learners who demonstrate a large vocabulary size, but do not use the full richness of their vocabulary in writing, may need some lexical interventions and well-designed tasks to help draw out and apply their vocabulary knowledge in diverse ways (Laufer & Nation, 1999).

Study Limitations

There are several limitations that apply to the present study. First, productive vocabulary size is difficult to measure because writers only use a fraction of their lexicon in writing (Nation & Webb, 2011). It is constrained by task, purpose, and topic. However, word list research has shown that vocabulary used within academic expository texts tend to be more consistent than that in narrative or creative writing (Coxhead, 2000). A recommendation to examine the vocabulary size of a text is to utilize corpus-based word frequency counts to calculate the average use of high, medium, and low-occurring words in natural language (Schmitt, 2010).

Second, lexical diversity is tied to the length of composition (Laufer & Nation, 1995; Durán, Malvern, Richards, & Chipere, 2004; McCarthy & Jarvis, 2010; Tweedie & Baayen, 1998). Given that the corpus of learner essays contained within this study is from a large number of different participants, assignments, and genres, the text length of these samples varies considerably. Recognizing this potential weakness, the measures of lexical diversity selected for this study take into consideration the text length by sampling various portions of learner texts to increase reliability (McCarthy & Jarvis, 2010); however, the limitation that shorter texts tend to display greater lexical diversity must still be noted (2010).

Lastly, the use of convenience and purposive sampling techniques are subject to systemic bias and can thus affect the generalizability of results (Fraenkel & Wallen, 2009). Despite this possible bias, convenience and purposive sampling procedures in this case were appropriate to ensure: (a) that participation was voluntary given the sometimes

personal content in academic essays per IRB stipulations, and (b) that the participants recruited meet the criteria for NNS and NS writers preparing for college-level composition.

Study Assumptions

The present study is based on the following two main assumptions relating to vocabulary knowledge in production.

The first assumption relates to NNS learners' depth of word knowledge. If a learner can accurately use a word in written production, then this usage implies that the learner has a proficient depth of lexical knowledge for that particular word based on research demonstrating that productive mastery lies on the complex end of the word knowledge continuum (Henriksen, 1999; Laufer & Goldstein, 2004; Melka, 1997). In other words, the ability to recall a word from memory and produce it in discourse is more difficult than being able to recognize it receptively (Laufer & Goldstein, 2004).

NNS learners' acquisition of vocabulary at various frequency levels comprises the second assumption. If a learner's text demonstrates productive knowledge of words in the lower-frequency levels (i.e., having knowledge of academic and technical vocabulary that appears less frequently in discourse), then it can be assumed that the learner knows the words in the higher frequency bands preceding their level results (such as the first 2,000 frequent words listed in the General Service List; West, 1955). This assumption is grounded in research stating that both NNS and NS learners acquire higher-frequency words first before words in the lower-frequency bands (Meara & Bell, 2001; Read, 1988;

Schmitt, Schmitt, & Clapham, 2001). When learners first encounter a new word, a “memory trace” is left behind, and the more interactions learners have with a word, the more likely they are to remember it (Tremblay, Baayen, Derwing, & Libben, 2011). Furthermore, these high-frequency words are less likely to be constrained by register and connotation, allowing them to appear often in informal discourse (Schmitt, 2010). In addition, corpus research reveals that frequency interrelates with complexity in word form, which can also affect acquisition. In other words, the more syllables in a word, the less likely it is to occur in language use and the less likely a NNS learner will acquire it (Edwards & Collins, 2011).

Definitions of Terms for the Study

In order to provide clarity and understanding regarding the constructs within the study, the following terms and acronyms are defined below.

- *Productive vocabulary* – vocabulary recalled and produced into order to express meaning in written discourse.
- *Receptive vocabulary* – vocabulary recognized and connected to meaning when reading the word in written texts.
- *Lexical diversity* – when a writer uses a variety of different words in composition (also known as *lexical variation*)
- *Lexical richness* – refers to the large, sophisticated, and diverse lexicon found within a learner’s composition
- *First language (L1)* – the native language of the learner acquired from birth

- *Second language (L2)* – an additional language learned some time after the learner’s first language
- *ESL* – an acronym for English as a second language. This refers to the study of English in a country where English is the dominant, native language
- *EFL* – an acronym for English as a foreign language. This refers to the study of English in a country where English is not the native language.
- *IEP* – and acronym for intensive English program. IEPs are language learning centers that prepare NNS learners for postsecondary study in English in a university where English is the native language.
- *Native-Speaking (NS)* – adjective describing a learner who is either monolingual and/or possesses English as his or her first language
- *Non-Native Speaking (NNS)* – adjective describing a learner for whom English is a second or additional language
- *Frequency* – the words most likely to be encountered in discourse. Learners generally acquire more frequent lexis before less frequent words
- *Token* – a group of letters with space around it, i.e. running words
- *Type* – number of different words in a text, i.e. term for a word that is counted once within a text, but possibly used more than one time
- *Word family* - a group of words that share the same base to which derived and inflected affixes are attached

Organization of the Study

Chapter One first presented the background to the study, introducing the factors that form the foundation of the research questions. It also detailed the problem underpinning the rationale for conducting the dissertation study. Next, the chapter presented the research questions and hypotheses investigated. Finally, the chapter included the professional significance of the study, the assumptions and limitations, and defined key terms.

Next, Chapter Two reviews the related literature and research related to the problem being investigated. Given the role of vocabulary in writing quality, the chapter first discusses research and studies providing evidence that lexical knowledge correlates positively to writing score. Due to this relationship, the chapter proceeds to analyze studies examining the lexical features of quality writing and the instruments of measurement used therein. The chapter concludes with a review of the few empirical studies conducted that have investigated the intersection between lexical size and variation measures and writing.

The methods and procedures used to collect data for the study are presented in Chapter Three. It includes details regarding the research design and sample to be used. The chapter also presents the instruments employed to collect data and the statistical analysis procedures that will be utilized to analyze resulting data. Lastly, it concludes with the ethical considerations for research.

The results of statistical procedures and findings from the study will be provided in Chapter Four. Chapter Five will discuss conclusions drawn from the findings, a

summary of the study, its findings, and limitations. The dissertation closes with pedagogical implications and recommendations for future research.

CHAPTER 2: RESEARCH AND LITERATURE REVIEW

The field of English language teaching (ELT) and research holds in general accord that vocabulary is indispensable for language acquisition and communication (Folse, 2004; Laufer, 1991; Schmitt, 2010). The commonly cited observation that learners often carry and use dictionaries in the language classroom underlines the importance of vocabulary in language learning and use. Words function as the gatekeepers to language and without them, communication does not succeed.

In response, there has been a notable increase in the number of vocabulary-related studies within the body of second language acquisition (SLA) and English language teaching (ELT) research. Several explanations expound the upturn of vocabulary-centered research. First, learners themselves view vocabulary as critical to academic language success (Leki & Carson, 1994). Second, the lines between lexis and other linguistic systems are viewed as more blurred than demarcated. Word knowledge has been observed to inform syntax, morphology, phonology, pragmatics, and rhetoric (Biber & Conrad, 2001; Schmitt, 2010). Lastly, receptive and productive word knowledge is a key aspect in language assessments. Indicators of lexical knowledge such as word choice and deployment impact communicability and thus affect learner fluency. These markers of vocabulary proficiency are of particular importance in the case of academic writing quality. Nowhere does vocabulary play a more active role in assessment than in evaluations of student writing with high correlations of .70-.79 to writing ability

(Schmitt, 2010). Therefore, lexical proficiency is essential for developing the academic language necessary for college composition.

Thus, this chapter reviews the scholarship relating to the impact of learners' vocabulary size and variation on academic writing ability and achievement. It first discusses the theoretical constructs that form the conceptual framework for the present study. Next, the chapter presents the extant empirical research into the role lexical knowledge plays in second language (L2) writing. The chapter concludes with a review of research and instruments specifically related to vocabulary size and lexical diversity in composition, which are the targeted variables in the present study.

Defining Academic Language

Academic language is the heart of academic composition and refers to the discourse encountered within educational settings. It is the language associated with textbooks, lectures, assessments, and academic disciplines. However, academic language is best defined in terms of its distinction from its counterpart social language. Cummins' (1979) BICS/CALP dichotomy between social and academic language is cited so frequently in the field because of its simple yet powerful message.

BICS and CALP

Basic Interpersonal Communication Skills (BICS) are the everyday linguistic features that all children acquire naturally within the first language (L1) (Cummins, 1979; Cummins & Swain, 1986). BICS is considered to be informal in register and is the language of daily conversation. Additionally, BICS is characterized by the use of slang,

informal grammatical structure, and the most frequently occurring words in English (Coxhead, 2000; Folse, 2004; Nation, 2001; Scarcella, 1996). For non-native English speaking (NNS) learners, there is evidence that they acquire BICS first, mirroring native speakers' L1 acquisition processes (Ellis, 2008).

In contrast, Cognitive Academic Language Proficiency (CALP) is acquired through academic study and develops over a longer period of time for both native and nonnative learners. Unlike BICS, CALP utilizes formal discourse features such as sophisticated vocabulary words and complex grammatical structures (Coxhead, 2000; Folse, 2004; Nation, 2001; Scarcella, 1996). Deviations from the formal aspects of CALP in discourse are viewed as errors and often indicate a learner's limited proficiency in academic language.

However, Singhal (2004) claimed the BICS/CALP distinction does not provide the level of precision necessary to explain the language demands of academic writing. To explain this lack of precision, she detailed three distinct dimensions of academic language: the linguistic, cognitive, and language discovery dimensions. Together, they provide a more comprehensive framework for defining academic language (2004).

Linguistic Dimension of Academic Language

The linguistic dimension of academic English is similar to Cummins' (1979) depiction of CALP. Academic writing in the linguistic dimension of Singhal's (2004) model goes beyond the lexical and syntactic descriptors of academic language to include the morphological, semantic, sociolinguistic, and discourse aspects of academic English. Mastery of academic writing requires intimate knowledge of each of these linguistic

features as they appear in specific rhetorical modes. For example, certain transition words signal particular writing genres, provide cohesion, and allow the audience to access the appropriate schema to achieve reading comprehension.

Cognitive Dimension of Academic Language

The second dimension involves the cognitive components of academic language such as higher-order thinking skills, accumulated knowledge, and background knowledge. Singhal (2004) grounds the cognitive dimension in the progressively complex skills delineated in Bloom's (1956) *Taxonomy of Cognitive Domains* (Figure 1). Higher-order thinking in academic writing concerns the author's ability to execute rhetorical devices such as support a thesis, develop logical arguments, understand fact and opinion, and synthesize multiple and countering perspectives (Ferris, 2009).

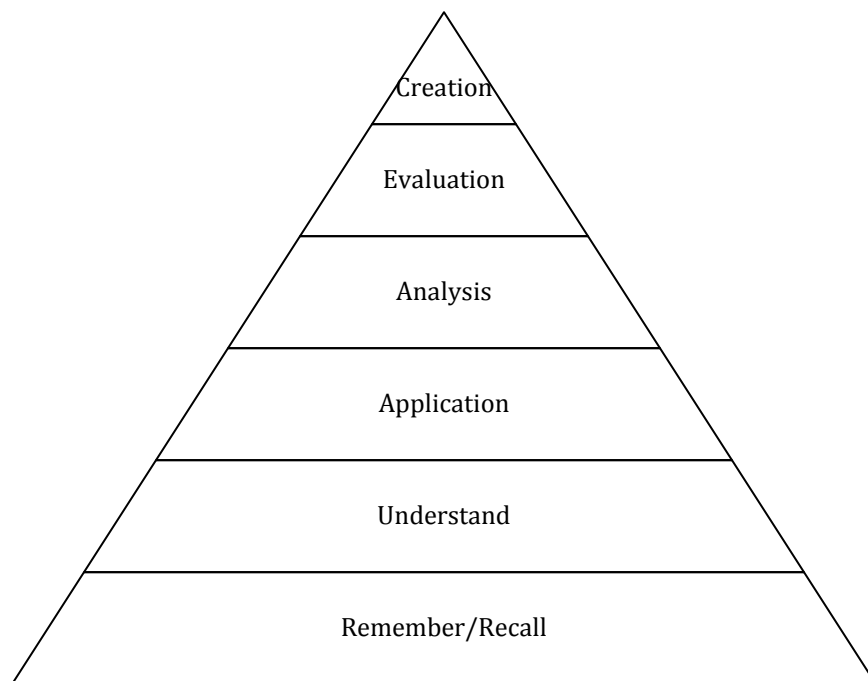


Figure 1. *Bloom's taxonomy of cognitive domains (1956)*

Language Discovery Dimension of Academic Language

The language discovery dimension makes up the final component of Singhal's (2004) academic language model. This dimension involves the metacognitive and tactical skills associated with academic study. It requires learners to recognize their own intellectual aptitudes and employ specific strategies that enhance knowledge or compensate for lack of understanding. This includes skills such as avoiding plagiarism, highlighting, note taking, and extracting information to name a few (Ferris, 2009). In terms of academic writing, the language discovery dimension is shaped by learners' ability to carry out the process of writing by pre-writing, drafting, editing, and revising (2009). It also involves aspects of analyzing texts closely for particular conventions in order to enhance metalinguistic awareness (Singhal, 2004).

Academic Language and NNS learners

These models of academic language reveal that academic language is a complex network of various linguistic, cognitive, and metacognitive skills. Unlike social language, academic discourse may not be fully mastered by all learners regardless of language learning designation. Estimates of the length of time it takes to attain proficiency in academic language range from seven to ten years at a minimum (Cummins, 1979; Cummins & Swain, 1986). This is a timespan that many NNS learners preparing for postsecondary study do not have. Recommendations for aiding NNS learners to overcome this disadvantage center upon the explicit instruction of the structures, conventions, and vocabulary of academic discourse (Biber & Conrad, 2001; Coxhead, 2000; Ferris, 2011; Scarcella, 2003; Singhal, 2004).

However, recent studies in corpus linguistics have begun to articulate that vocabulary plays a much more significant role in academic language development than any other factor (Coxhead & Byrd, 2007; Crossley, Salsbury, McNamara, & Jarvis, 2010; Conrad, 2000; Biber & Conrad, 2001; Coxhead, 2000; Simpson-Vlach & Ellis, 2010). The study of large corpora has revealed that academic prose is highly formulaic in nature and teaching learners to recognize and incorporate these lexical patterns in their writing results in higher levels of fluency (Biber & Conrad, 2001, Simpson-Vlach & Ellis, 2010). The assertion that academic language chiefly consists of such lexical chunks requires an in-depth analysis of what constitutes word knowledge.

Defining and Measuring Lexical Knowledge

The current position of lexical knowledge research states that knowing a word goes far beyond the ability to reproduce or approximate a dictionary definition of the word (Folse, 2004; Henriksen, 1999; Nation, 1990; 2001; Nation & Webb, 2011; Read, 2004). The link of form to meaning, though important, encompasses only one of the many dimensions of lexical knowledge needed for accurate and fluent use. Various researchers have proposed hierarchies framing the multiple aspects of word knowledge that lead to lexical competence.

Henriksen (1999) put forth three dimensions of word knowledge. The first dimension of partial to precise knowledge refers to the linking of form to meaning. The second dimension, depth of knowledge, relates to knowing a word's orthographic, phonological, syntactic, semantic, morphological, collocational, and pragmatic

characteristics. The third dimension of receptive to productive use refers to the ability to understand the word and use it accordingly in appropriate contexts.

Read (2004) also classified word knowledge into three constructs similar to Henriksen, but he grounds his model in the ways depth of word knowledge is usually assessed. Read labeled his classifications as: (1) precision of meaning, (2) comprehensive word knowledge, and (3) network knowledge. The first two categories closely relate to Henriksen's (1999) partial to precise and depth of knowledge dimensions. However, his third category, network knowledge, refers to how words are associated with each other within the lexicon (e.g., the statement of the word *cat* elicits *dog*) and is based on the Word Associates Test (Read, 1993; 1998).

Nation (1990; 2001) described nine different aspects of a word's form, meaning, and use (see Table 1). This comprehensive framework aimed to increase the sensitivity in the assessment of depth of vocabulary knowledge to explore and discover learners' partial knowledge of word items (Nation, 2001). In this framework, a learner must have receptive and productive mastery in all nine aspects in order to truly know a word. For NNS writers, this is a massive undertaking. Deficits in any one of these areas of word knowledge can cause a writer's message to be clouded, misinterpreted, or even appear awkward (Folse, 2008).

Table 1

Nation's Aspects Involved in Knowing a Word (Nation, 2001, p.27)

Form	Spoken	Receptive	What does the word sound like?
		Productive	How is the word pronounced?
	Written	Receptive	What does the word look like?
		Productive	How is the word spelled?
	Word parts	Receptive	What parts can we recognize in this word?
		Productive	What word parts are needed to express meaning?
Meaning	Form and meaning	Receptive	What meaning does this form signal?
		Productive	What word form can be used to express this meaning?
	Concept and referents	Receptive	What is included in this concept?
		Productive	What items does the concept refer to?
	Associations	Receptive	What other words does this make us think of?
		Productive	What other words are possible to use instead of this one?
Use	Grammatical functions	Receptive	In what patterns does this word occur?
		Productive	In what patterns is this word required to use?
	Collocations	Receptive	What other words or types of words occur with this one?
		Productive	What words or types of words must we use with this one?
	Constraints on use (register, frequency, etc.)	Receptive	Where, when, and how often would we expect to encounter this word?
		Productive	Where, when, and how often can we use this word?

In order to assess vocabulary knowledge completely, lexical tests must measure multiple aspects of word knowledge and to what extent each of these aspects is known. The majority of tests of word knowledge used in research either involve matching and/or elicitation tasks or ranking knowledge via ordinal scales. The two most influential

measures are perhaps the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993) and the Word Associates Test (Read, 1993; 1998a).

Paribakht and Wesche (1993) developed a five-point, self-report scale called the Vocabulary Knowledge Scale (VKS) to gauge a NNS learner's receptive and productive understanding of a word. The VKS contains five levels of word knowledge (Table 2). In this approach, learners self-describe the scope of their knowledge and understanding of a particular word's meaning and use ranging from receptive knowledge of a word (*I have never seen this word*) to their productive knowledge (*I can use this word in a sentence*) (1993).

Table 2

Vocabulary Knowledge Scale (Paribakht & Wesche, 1993)

Point value	Self-report categories
1 point	I don't remember having seen this word before.
2 points	I have seen this word before, but I don't know what it means.
3 points	I have seen this word before, and I think it means _____. ^a
4 points	I know this word. It means _____. ^b
5 points	I can use this word in a sentence. Write your sentence here _____. ^c

^aLearner needs to provide a synonym in English or an L1 translation.

^bLearner needs to provide a synonym in English or an L1 translation.

^cLearner needs to complete number 4 in addition to 5.

Although widely used, some critics of the VKS claim that it only measures a fraction of a learner's actual knowledge of a word (Nation & Webb, 2011; Schmitt, 2010). For example, in order to accurately produce a target word in a sentence in number five on the scale, a learner would need to possess knowledge of the all words that

surround it, such as collocations, syntax, etc. (Nation & Webb, 2011). An additional problem of ranking word knowledge using a scale is that it assumes that there is a scaled progression through the aspects of word knowledge. This advancement is not always the case. In some situations, learners may be able to produce a sentence using the target word based on receptive knowledge, but unable to provide a synonym or definition of the word (Nation & Webb, 2011; Schmitt, 2010). Another negative aspect of the VKS involves the use of an ordinal scale for vocabulary. While the five rankings on the scale appear to be a simple lexical measure, the use of five points on an ordinal scale here is problematic. For example, do learners who write four (*I know this word. It means ____*) for a word really know twice as much about that word as learners who write a two (*I have seen this word before, but I don't know what it means.*)? Although the VKS is lauded for its ease of use, these issues with its reliability must be factored in when analyzing learners' self-reports of word knowledge (Folse, 2006).

The second influential depth of word knowledge test is Read's (1993; 1998a) Word Associates Test. This test assesses receptive word knowledge based on the aspects of form and meaning, concept and referents, and collocation (Nation & Webb, 2011). A learner must select four words that are synonyms, collocates, or aspects of meaning of the target word from eight options. For example, the stimulus word is *beautiful* and the eight available answer options are: *enjoyable, expensive, free, loud, education, face, music, and weather* (Read, 1998b). The correct answers in this case are: *enjoyable, face, music, and weather*. *Face, music, and weather* are collocates of *beautiful*, and *enjoyable* represents an aspect of the word's meaning. Read asserts that the Word Associates Test is an

effective and rigorous depth of knowledge test because native speakers were not able to find all four correct responses in pilot studies (1993). As such, its design has been frequently used in many lexical studies.

In attempting to derive an accurate picture of the relationship of lexis and writing proficiency, the depth of a learner's word knowledge of particular words needs to be factored into the analysis. Studies profiling NNS learners' word knowledge have revealed that partial knowledge is the norm (Schmitt & Zimmerman, 2002). How well a learner knows a word affects his/her ability to accurately generate a word in context. The tests of depth of word knowledge discussed in this section all arrive at the same conclusion that the ability to accurately produce a word in writing lies at the most complex and concrete end of the word knowledge continuum. Consequently, insufficient word knowledge can result in a composition rife with lexical issues and errors thus affecting writing quality. Therefore, the examination of learners' written productive vocabulary capabilities can provide a reliable snapshot of the state of learners' proficiency in terms of words they actually know and use accurately in production (Laufer & Nation, 1995).

The Role of Language Proficiency in L2 Writing Ability

Studies examining L2 writing skills have concluded that the greater the level of learners' language proficiency, the higher the quality of their compositions (see de Haan & van Esch, 2005; Grant & Ginther, 2000; Hawkey & Barker, 2004; Wang, 2003, Zareva, Schwanenflugel, & Nikolova, 2005). In several studies of written production,

proficiency level correlated to higher composition scores. Results indicated that compositions by highly proficient learners tend to be longer in essay length, lexically and grammatically sophisticated, have better cohesion, and demonstrate an improved command of rhetorical conventions (Cumming, 2001; Cumming, 2006; Grant & Ginther, 2000; Crossley, et al., 2010). As NNS learners advance in their linguistic competence, they move away from utilizing discourse that mirrors spoken language toward written discourse constructions that demonstrate an increase in the use of nominalizations, passive voice, and lexical forms that signal cohesion and subordination (Agustín Llach, 2007; Grant & Ginther, 2000).

Not surprisingly, research has shown that learners' writing skills are transferred from the L1 to the L2 (Bernhardt & Kamil, 1995; Cummins, 1979). Learners' L2 proficiency level provides the avenue through which this transfer can occur (Cumming, 2006). Research into L1 and L2 composing processes have shown that learners must reach a particular threshold level of proficiency in the L2 before the transfer of writing skills from the L1 can occur (Cumming, 2006; Cummins, 1979). As a result of this linguistic threshold, NNS writers with a higher level of linguistic competence are better able to access writing skills developed in the L1 and their compositions are more likely to garner higher scores from raters (Agustín Llach, 2011).

Relatedly, studies examining the use of the L1 during the L2 composing process indicate that problems at the lexical level are the most frequent errors (Lennon, 1991; Wang, 2003). Therefore, this error pattern indicates that L2 vocabulary knowledge is a key contributor to L2 writing ability. Learners with a lower level of proficiency possess

fewer words in their lexicon and tend to make use of their L1 in order to reconcile lexical issues such as word choice, generation, and solicitation (2003). This limitation leads to an increased likelihood that lexical errors such as calques, borrowings, or misspellings will appear in their compositions (Agustín Llach, 2007; Grant & Ginther, 2000). Studies coding lexical errors in L2 composition reveal that the number of overall L1-influenced errors decreases as L2 language proficiency increases and as learners grow their L2 lexical knowledge base (Bardovi-Harlig & Bofman, 1989; Hawkey & Barker, 2004; Lennon, 1991). Therefore, lexical proficiency is an important indication of L2 writing proficiency.

The Role of Lexical Knowledge in the Scoring and Evaluation of L2 Texts

Research examining the relationship between learners' vocabulary use and writing quality have produced a general finding that lexical issues greatly impact the scoring of NNS writers' compositions (Engber, 1995; Laufer & Nation, 1995; Roberts & Cimasko, 2008; Ruegg, Fritz, & Holland, 2011; Santos, 1988). Writing in its basic form is essentially the formulation of nonverbal ideas into textual representations, a process that requires purposeful and accurate use of lexical items (Schoonen, van Gelderen, Stoel, Hulstijn, & de Glopper, 2010). NNS writers who lack sufficient lexical knowledge to transform their thoughts to paper may encounter barriers to the expression and communication of their ideas, thus leading to lower writing scores (Folse, 2008; Santos, 1988; Silva, 1993). Research supporting the important role vocabulary plays in the

scoring of writing appears in two strands: the rubrics used to score writing and studies exploring vocabulary's contribution to writing score variance.

Composition Rubrics

The first widely used rubric to judge English as a second language (ESL) writing was the ESL Composition Profile created by Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey (1981). This 100-point rubric contains five broad categories of content, organization, vocabulary, language use, and mechanics. Raters evaluate a composition based on the point values assigned to each category. Although vocabulary is only assigned 20 points, careful study of the rubric reveals that vocabulary noticeably affects other categories. For example, in order to gain points in content, writers must demonstrate “range” and provide “detail[s],” which require a substantial amount of vocabulary to achieve. Another example of vocabulary's influence on other categories would be in the area of organization, where the use of transition words, connectors, etc. help to structure an essay (Folse, 2008). In a test of the ESL Composition Profile's validity and reliability, Astika (1993) found that when raters scored writing with this rubric, vocabulary accounted for almost 84% of total score variance ($F = 1071.76, p < .001$). As a result, vocabulary has the ability to influence overall scores well beyond its assigned 20 points on this rubric.

A second notable rubric used to evaluate academic writing is the TOEFL iBT Integrated Writing Rubric (Educational Testing Service [ETS], 2005). The scoring standards on this rubric measure an examinee's ability to understand an academic lecture and reading passage and then apply the information to respond to a writing prompt

(Folse, 2008). Evaluators holistically score essays on a scale of 0 to 5, with 0 being the lowest score and 5 the highest. Each of the score values includes a language component involving vocabulary. For instance, scores two and three deal with errors in expression “that obscure connections or meaning” (ETS, 2005, p. 196). Scores of four and five mention minor language errors that “do not result in inaccurate or imprecise presentation of content” (p.196).

The third rubric worth mentioning here is the Independent Writing Rubric of the TOEFL iBT (ETS, 2005). The rubric measures learners’ ability to argue a point of view on a particular issue. This rubric is also scored on a scale of 0 to 5 as with the Integrated Writing Rubric. The criterion for evaluation focus on four areas: content, organization, coherence, and language use. Though there is no independent lexical category, each score once again contains lexical benchmarks, ranging from “variety and range of vocabulary” to “inappropriate word choice/forms” and “minor lexical errors” (2005, p. 195).

Finally, the International English Language Testing System’s writing rubric (IELTS) is the final rubric measuring L2 academic writing preparedness. The IELTS writing examination consists of two academic writing tasks. Task one asks examinees to interpret visual information such as a chart, graph, table, or diagram and compose a 150-word minimum description of the figure in English (IELTS, 2013). Task 2 presents test-takers with a point of view or problem about which they must argue a stance in at least 250 words. Each task is rated holistically on a scale of 0 to 9. Scores are assigned based on four criteria: overall task achievement, coherence and cohesion, lexical resource, and

grammatical range and accuracy (2013). In contrast from the TOEFL Independent Writing Rubric, the IELTS rubric explicitly qualifies the use of “a wide range of vocabulary” and the “[skillful] use of uncommon lexical items” as indicators of writing quality (2013, p. 1).

Hawkey and Barker (2004), in a large-scale qualitative analysis of writing scale criteria and benchmarks, aimed to tease out a collective set of writing standards for the development of a common writing assessment scale. In this comprehensive review of the various ESL rubrics and scales, they identified several linguistic characteristics that are widely-accepted as indicators of L2 writing proficiency. Results place lexical richness features such as lexical diversity and sophistication as essential gauge of fluent L2 writing. This finding is in line with the rubrics’ proficiency benchmarks examined earlier in this section.

Table 3

Writing Rubrics

Rubric	Point Value	Lexical Criteria
ESL Composition Profile (Jacobs et al., 1981)	18-20	Sophisticated range, effective word choice, word form mastery, appropriate register
TOEFL Integrated Writing (ETS, 2005)	4-5	Minor lexical errors
TOEFL Independent Writing (ETS, 2005)	4-5	Variety and range of vocabulary, occasional noticeable minor errors in word form and use of idiomatic language; Appropriate word choice and idiomaticity, minor lexical errors
IELTS Tasks 1 and 2	9	Uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as ‘slips’, use of uncommon lexical items

As the four widely used and influential rubrics presented in this section clearly show, measures of competency in writing consider vocabulary an integral component of

writing ability and award sizable values to lexical proficiency in L2 writing, often far greater than the assigned point value (Folse, 2008; Hawkey & Barker, 2004).

Empirical Studies

Since many writing rubrics include vocabulary as a critical scoring standard, research has investigated to what extent vocabulary predicts overall writing score. The majority of early studies on the contribution of vocabulary to writing quality followed a discourse/text analysis framework (Chastain, 1980; 1981; Dordick, 1996; Engber, 1995; Grobe, 1981). Consequently, these studies were largely exploratory to determine which types of errors were most bothersome to raters. In other words, these studies concentrated on how many points a writer would lose for vocabulary errors or weak lexical proficiency rather than how many points a writer could gain for good vocabulary usage.

Chastain (1980) conducted a study of university L2 Spanish learners to examine which language errors interfered the most with comprehension by native speakers of Spanish. He generated a list of 35 Spanish sentences containing at least one error. Native Spanish speakers in Spain then evaluated the sentences for comprehensibility. Results indicated that comprehension was most severely inhibited by lexical errors such as using the wrong word or the addition or omission of a word. In response to criticisms that using isolated sentences weakened data, Chastain (1981) replicated his previous study using short essays. Results were almost identical with incorrect use of vocabulary words rated as the most serious error.

Similarly, Dordick (1996) performed an exploratory study with 289 NS college students to see what types of errors in ten different versions of an essay written by an NNS student interfered most with comprehension of the text. The various versions of the essays focused on different types of writing errors such as subject-verb agreement, syntax, vocabulary, weak rhetorical style, etc. Results demonstrated that the versions containing lexical errors such as poor word choice or form interfered the most with reader comprehension (1996).

While Chastain (1980; 1981) and Dordick (1996) investigated error gravity from the perspective of native speaker judgments, Santos (1988) investigated which errors were perceived to be most serious in the eyes of teachers. He surveyed 178 university instructors on their perceptions of two compositions written by non-native speakers of English using six 10-point scales that included content as well as language-focused criteria. Santos found that the instructors rated errors relating to vocabulary as the gravest problem with the writings and stressed the importance of careful word choice and use (1988).

Other empirical studies have focused solely on the types of lexical issues impacting the holistic score of L2 writings. In a study with 66 international students enrolled in an Intensive English Program (IEP) in an American university, Engber (1995) examined the specific relationship between lexical proficiency and reader perception of the overall quality of timed essays written by the IEP students. The scores of these essays were compared to four areas of lexical proficiency: lexical variation, error-free

variation, percentage of word errors, and density of word errors. Engber found that error-free lexical variation had the highest significant correlation to writing score (1995).

In a later study, Ruegg et al. (2011) investigated the interplay between lexis and grammar in the scoring of timed essays. They administered an English writing proficiency test to 140 Japanese incoming university students. Next, 45 native or near-native English-speaking professors at the university graded the essays using a variety of lexical and grammatical measures. The researchers then analyzed the scores to determine what portion of the variance in score is accounted for by different lexicogrammatical qualities such as accuracy, variety, and frequency (2011). A regression analysis revealed that 82% of variance was accounted for by the grammatically accurate use of lexis ($\beta = 0.820$). Therefore, using words correctly and appropriately facilitated perceived writing quality.

In sum, the studies presented in this section provide ample evidence that knowing a variety of words and how to employ them appropriately and accurately in context will garner higher ratings by evaluators. Considering the many aspects of knowing a word's productive form, meaning, and use (see Table 1), L2 writing can sometimes be characterized by its deviations from L1 lexical use, namely in the area of lexical richness, i.e., purposeful, specific, and diverse use of vocabulary in compositions.

Lexical Features of Writing Quality

Because NNS learners are at different stages of acquiring and learning to use vocabulary items, their writing is often marked by certain lexical differences that reveal

their developing acquisition of the language. Given that a writing task in a second or foreign language draws heavily on the linguistic resources a writer possesses, research has revealed that NNS writers encounter different lexical issues in the composing process from their native-speaking counterparts (Schoonen et al., 2010). As discussed previously, these lexical differences unfortunately are likely to be marked as errors by evaluators of writing due to rater sensitivity towards comprehensibility, acceptability, and fluency when assessing L2 writing (Roberts & Cimasko, 2008). Furthermore, it is very important to note that lexical errors tend to be more irksome to readers, interfering with readers' comprehension of the text, than any other types of writing errors (Chastain, 1980; 1981; Dordick, 1996). Since vocabulary clearly affects evaluators' judgment of overall writing quality, it is then prudent to examine which lexical features contribute to quality writing.

Empirical Studies

In a study of first language (L1) writing, Grobe (1981) compared the writing scores of 5th, 8th, and 11th grade students with a variety of syntactic, mechanical, and lexical measures. Results indicated that the total number of different words was the strongest predictor of writing score. Thus, Grobe (1981) arrived at the conclusion that “what teachers perceived as ‘good’ writing is closely associated with vocabulary diversity” (p. 85).

Linnarud (1986) found a similar result in a study comparing the compositions of 17-year-old Swedish advanced learners of English ($n = 42$) and those by English native speakers of the same age ($n = 21$). Despite their high level of English proficiency, the

Swedish learners' essays lacked lexical variation and originality. Additionally, there was a difference in the frequency of individual words used by the native speakers of English and the NNS learners. The native speakers tended to use more adjectives and adverbs thereby expressing more originality and specificity in their compositions.

In an investigation of a target L2 language other than English, de Haan and van Esch's (2005) longitudinal study of 31 non-native Spanish and 29 native Spanish texts arrived at a similar outcome. Both native and non-native Spanish teachers scored the essays holistically and the ratings compared to a range of lexical richness features. Findings not only concluded that lexical diversity more significantly impacted writing score, but lexical diversity also grows over time as NNS grow their overall L2 proficiency.

Using a within-subjects design in their longitudinal study of 281 Dutch secondary students learning English as a foreign language (EFL), Schoonen et al. (2003) compared writing performance and fluency between the participants' compositions written in their L1 (Dutch), and L2 (English). Their results also determined that writing proficiency and fluency in the L1 and the L2 relied more substantially on the writers' lexical knowledge in both languages more than any other criteria. In addition, the learners' L1 compositions demonstrated a higher level of lexical diversity than their L2 essays suggesting that vocabulary expertise played a key role in the composition process.

Finally, in an analysis of a corpus of L1 undergraduate essays, McNamara, Crossley, and McCarthy (2009) found that more proficient writers demonstrated a greater range of lexical diversity in their essays. The essays earning the highest scores contained

words that occur less frequently in language and rarely repeated words. This result suggests that high-proficiency writers have a larger lexicon from which to draw rich, diverse vocabulary items to express ideas.

The empirical research therefore demonstrates that lexical diversity most closely relates to writing proficiency. Evaluators are more likely to award points to writer who are able to vary their lexis and utilize specific word choices in composition. The studies discussed within this section further indicate that there are clear differences in the lexical diversity profiles of L1 and L2 compositions. One explanation for NS writers' ability to vary lexis in text stems from their relatively large vocabulary size as compared to that of NNS learners'.

Vocabulary Size and Writing Quality

Several studies have demonstrated a positive correlation between a large lexicon and the ability to execute language skills (Alderson, 2005; Table 4). In the areas of listening and speaking, research indicates that the NNS learner needs to know approximately the most frequently occurring 2,000 to 3,000 word families in order to listen to and carry out a conversation (Nation, 2001; Schmitt, 2000). This figure allows learners to understand about 90% of spoken language (Nation, 2001). In order to read and comprehend text, Hirsch and Nation (1992) and Laufer (1992) found that a 5,000 word family threshold is required for basic reading. However, in order to achieve adequate reading comprehension for academic study, the optimal threshold increases 8,000 to 9,000 word families, translating to roughly 98% text coverage (Hu and Nation,

2000; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006). An adequate lexical threshold level for writing has yet to be identified. Some preliminary figures for writing (see Staehr, 2008) suggest that they mirror those of the reading figures; however, the aforementioned depth of vocabulary knowledge research shows that a higher level of word knowledge is required in order to accurately produce vocabulary items in writing and therefore the figures for writing could be much higher (Laufer, 1998; Staehr, 2008).

Table 4

Number of Words for Language Skills and Coverage

	Spoken English	Basic Reading	Academic Reading	Writing
Number of Word Families	2,000-3,000	5,000	8,000-9,000	--
% of Text Coverage	90%	95%	98%	--

Lexical Richness Features and Writing Quality

The studies in the previous sections looking into quality of lexis in writing share the conclusion that lexical richness tends to be the central component within L1 and high-proficiency texts that often needs development in L2 compositions (Laufer & Nation, 1995; Silva, 1993; Tiball & Treffers-Daller, 2008; Table 5). The term lexical richness consists of three facets: lexical diversity, lexical sophistication, and control of words in production (Read, 2000).

Lexical Diversity

The first component, lexical diversity, refers to the fact that proficient writers in their L1s tend to use a wider variety of words in academic compositions (Laufer & Nation, 1995; Read, 2000; Silva, 1993). NNS writers may have issues in varying their words within their writing because academic terms are less likely to be encountered in

ordinary conversations (Coxhead & Byrd, 2007). Therefore, learners receive less input and exposure to lower-frequency words found in academic discourse, making it more difficult to retrieve these lexical items for the purposes of varying their use of words (Laufer, 1994). Lexical variation is roughly calculated by counting the number of word *types*, or number of different words, in a text, multiplying it by 100, and then dividing the result by the number of *tokens*, or running word count (Laufer & Nation, 1995). The resulting figure indicates how varied a writer's vocabulary is; that is, the larger the type-token ratio is, the more varied the vocabulary (1995). Research measuring variation and diversity of vocabulary in L2 writing state that texts that contain more repetition and a smaller set of words are perceived as lower quality than texts that contain a greater proportion of different words (Engber, 1995; Schoonen et al, 2003; Silva, 1993).

In a corpus analysis study of 545 samples of spoken and written language, Biber (1986) probed each sample for differences in a broad number of 41 linguistic categories. In a factor analysis, Biber (1986) weighted each of the 41 variables to find five derived factors that underlie communicative competence across spoken and written registers. His findings indicated that the use of varied and different words led to the conveyance of specific meanings and indicated better use of productive vocabulary in communication.

Likewise, Ferris (1994) conducted a study to explore the lexical and syntactic features prevalent in L2 compositions. She used a corpus of 160 ESL student essays written for a university placement exam. From this corpus, 28 variables related to composition such as number of words, verb tense, and synonymy/antonymy were selected to be compared with holistic scores given by three independent raters. Analysis

revealed that the number of words, synonymy/antonymy, and word length all contributed significantly to higher holistic scores. Moreover, students with higher language proficiency tended to vary their word choices more often than lower proficiency students (Ferris, 1994).

In a more recent study, Yu (2009) compiled a corpus of 200 compositions and 25 interviews from an international language test battery to examine the impact of lexical diversity on speaking and writing quality. Yu (2009) utilized *D* as the instrument (see Malvern & Richards, 1997; Durán, Malvern, Richards, & Chipere, 2004) to quantify each text and interview's lexical diversity and then correlated the values to writing and speaking scores. Results demonstrated that lexical diversity has a significant, positive relationship to both writing and speaking scores.

Further in-depth analysis of the instruments relating solely to the measurement of lexical diversity is presented in a later section.

Lexical Sophistication

Lexical diversity also leads to a more sophisticated use of vocabulary, another component of lexical richness that affects L2 writing (Tidball & Treffers-Daller, 2008). Read (2000) defines lexical sophistication as “the use of technical terms and jargon as well as the kind of uncommon words that allow writers to express their meanings in a precise and sophisticated manner” (p. 200). This definition is based on the operationalization that a sophisticated lexical item is one that does not occur frequently in use. Laufer (1994) and Laufer and Nation's (1995) Lexical Frequency Profile (LFP) follows this same assumption, rating higher-frequency words as easier than lower-

frequency items. The LFP generates a rating of a text's use of sophisticated vocabulary items based on four frequency bands constructed from Xue and Nation's (1984) University Word List (UWL). The first band is made up of the 1,000 most frequent words in English; the second band includes the second 1,000 most common words; the third band then consists of the 836 academic words on the UWL. The final band contains words not found within the aforementioned three bands (Laufer & Nation, 1995). Corpus-based word frequency has been the longstanding method of quantifying a texts' average lexical sophistication and vocabulary size.

Studies measuring lexical sophistication have found that the level of sophistication increases in tandem with proficiency (Laufer, 1994; Laufer & Nation 1995; Silva, 1993). In a study of 65 EFL learners enrolled in an English for academic purposes (EAP) course in Israel and New Zealand, Laufer and Nation (1995) determined that learners with lower language proficiency used more high-frequency words than those with higher proficiency who tended to use less frequent, more specialized words.

Earlier, Laufer (1994) examined the lexical profiles of 48 undergraduate EFL learners in Israel. Results from independent *t*-tests examining lexical growth between an entrance and end-of-term exam reported that as learners progressed in language, the more words from lower-frequency bands of the UWL they used. In other words, as learners grew their L2 lexical knowledge through study, higher-frequency words were replaced with lower-frequency items indicating a larger, more sophisticated vocabulary.

Lexical Control

The third component of lexical richness is lexical control (Read, 2000). Proficient writers demonstrate larger control over their use of vocabulary during the composition process, meaning these writers are purposeful and precise when choosing words to express their ideas (Silva, 1993). In other words, lexical control refers to evidence within written compositions that the author make conscious choices to deploy particular words to convey specific meanings or stylize the text.

Crossley, Salsbury, McNamara, and Jarvis (2010) used a computational linguistics program called Coh-Metrix to run a comparison of 180 ESL essays to 60 essays written by native speakers of English to compare and contrast lexical proficiency between the two corpora. To validate computer results, three human raters also analyzed the 240 essays. Analysis from both the computer program and the raters revealed that L1 essays employed more selective word choice such as hypernymy, synonymy, and antonymy in their writing, indicating that native speaking authors choose words that are specific rather than general to convey ideas. For example, where a NNS writer might write *car*, a NS writer would employ a more specific item such as *coupe*. These findings are similar to those of Ferris's aforementioned (1994) study.

In another study, Roberts and Cimasko (2008) asked 71 social science and engineering faculty to correct the top five errors on an essay written by an ESL student and assign a holistic score on a scale of 1 to 10. The researchers found that the professors significantly chose to correct errors related to vague and poor word choice over grammatical errors. Furthermore, some faculty members chose to rewrite or edit parts of

the essay despite the technical correctness of the information within the paper. The professors' inclination to reword particular sentences indicates the importance of lexical control in academic writing to convey content and meaning (2008).

Grant and Ginther (2000) operationalized lexical control in terms of specific word items that serve a particular discourse function within the text. Using a computerized tagging program, they analyzed 90 TOEFL Test of Written English essays for differences in a variety of lexical and grammatical features among three levels of L2 proficiency. Results indicated that highly rated essays employed unique, sophisticated words such as conjuncts and demonstratives that allowed the writers to more logically and connect ideas and confidently support claims and viewpoints. Analysis of lexical features only also suggested that control over the production of words led to an increase in lexical diversity. This finding suggests that individual components of lexical richness are interdependent upon each other.

Table 5

Lexical Impact on Writing Quality

Findings	Study
Lexical errors significantly impede comprehensibility	Chastain (1980; 1981), Dordick (1996); Engber (1995), Roberts & Cimasko (2008), Ruegg et al. (2011), Santos (1988)
Lexical diversity correlates highly with writing quality	Biber (1986), Crossley et al. (2010), de Haan & van Esch (2005); Engber (1995), Ferris (1994), Grant & Ginther, (2000), Grobe (1981), Hawkey & Barker (2004), Laufer (1994), Linnarud (1986), McNamara et al. (2009), Schoonen et al. (2003); Yu (2009)
L1 essays demonstrate higher lexical diversity, sophistication, and control than L2 equivalents	Crossley et al., (2010), de Haan & van Esch (2005), Laufer (1994), Laufer & Nation (1995), McNamara et al. (2009), Schoonen et al. (2003)
Vocabulary size affects writing proficiency	Laufer (1994), Laufer & Nation (1995), Staehr (2008)

While Read's (2000) delineations of lexical richness and the various empirical studies dedicated to its study have provided three relatively clear indicators of a lexically-rich text, the lines between each indicator seem to blend. Many of the studies in this section appear to operationalize each indicator in similar terms thus making it difficult to discern what aspect of lexical richness is being investigated. For example, Crossley et al.'s (2009) finding that higher-proficiency texts executed lexical control when specific words in a manner to create hypernymy, synonymy, and antonymy could be argued as indications of both lexical diversity and sophistication. This observation provides strong proof of the multifaceted and interrelated relationship between lexical richness features.

Operationalizing Vocabulary Size and Lexical Diversity

As a result of such an interdependent nature between the three facets of lexical richness, the present study must clearly operationalize the variables being investigated. For this reason, the study adopted Laufer and Nation's (1995) more dichotomized view of lexical richness. What makes a word more "sophisticated" or "specific" than another and thereby indicative of a large lexicon is largely subjective and would require an aspect of lexical control on the part of the writer. As such, Laufer and Nation (1995) recommend defining a text's advanced vocabulary size by word frequency. Therefore, the average word frequency, as measured by CELEX frequency scores (Baayen, Piepenbrock, & Gulikers, 1995), served to measure the productive vocabulary sizes of the collected data and allowed for a more holistic point of view. Other instruments based on frequency bands (such as Lexical Frequency Profile; Laufer & Nation, 1995), calculate the

proportion of word frequency at different bands, which relates more to lexical sophistication than the average size of a text's lexicon.

In order to ensure that lexical variation is separate from vocabulary size, Laufer and Nation (1995) report that its calculation should “not distinguish [between] what kinds of words are used” (p. 310). As a result, they state that learners can actually achieve lexical diversity using smaller lexicons. While Laufer and Nation (1995) argue this is a negative outcome, investigating this possibility is a fundamental aim of the present investigation. Consequently, lexical diversity in this study is defined as the number of different words produced with a text.

In sum, all of the discussed lexical richness components that contribute to proficiency in writing share one thing in common: the need for a large, well-rounded vocabulary repository containing low-frequency words from which the NNS writer can draw in order to accurately, clearly, and succinctly express ideas in writing (Table 5). However, there are indications that the key lexical predictors of writing quality is not the presence of many rare, low-frequency words in text, but a more varied approach to word choice and use (Laufer, 1994). In fact, Engber's (1995) study revealed that having too many low-frequency words in a passage could have a negative effect on writing scores. Texts that contain too many low-frequency vocabulary items can affect readability and ease of comprehension (1995). Therefore, this question of whether using low-occurring, sophisticated words or stylistically varying more frequently used words leads to writing proficiency is of interest to both teachers and researchers. Thus, the remaining literature review sections examine the popular methods of measuring productive vocabulary size

and lexical diversity and the dearth of research comparing their effects on writing proficiency.

Measuring Productive Vocabulary Size

When studying vocabulary size, there are generally three approaches research can take to measure how large a vocabulary NNS writers need to know (Nation & Waring, 1997). The first approach would be to consider how many words exist in the English language. A second could consider how many words native speakers, especially parallel speakers (e.g., native university students and non-native university students), know. Third, vocabulary size could be approached from the framework of how many words NNS writers need to know to successfully fulfill writing tasks (1997; Nation & Webb, 2011).

The first approach is not a valid measure for vocabulary size because not even native speakers of English know every word in existence. The second approach, to use the number of words native speakers know, is a bit more reachable. Research has estimated that native speakers' receptive vocabulary knowledge is around 20,000 word families based on a rule of thumb that a person's vocabulary increases by roughly 1,000 word families per year up to the age of 20 (Goulden, Nation, & Read, 1990). Along the same lines, the vocabulary size of NNS learners immersed in the target language environment has been discovered to grow at roughly this same rate (Nation & Webb, 2011). However, the variable with NNS learners is that they start acquiring vocabulary at a later age than native speakers. Therefore, there tends to be a large gap between the size

of a NNS learner's lexicon and that of a native speaker (Laufer, 1994; Nation & Waring, 1997). Consequently, a more feasible tactic to investigating how many words a NNS learner needs to know is the third approach of determining how many words are needed to accomplish language tasks, in this case, academic writing.

Words to Know

Since research has shown that a writing sample demonstrating a rich, varied, and sophisticated vocabulary tends to receive favorable perceptions from readers and raters, a logical question that follows involves which words should be used in order to facilitate a favorable perception. Recent advances in the field of corpus linguistics have shed insight into which words are most useful for NNS writers, especially in the area of English for academic purposes. As mentioned before, NNS writers have lexical issues because terms used in academic writing occur less frequently than those found in everyday conversation (Coxhead & Byrd, 2007). To increase instructors' and researchers' understanding of English academic vocabulary, several word lists have been created (Table 6).

Word lists. The first notable word list designed for non-native speakers of English was West's (1953; 1955) General Service List (GSL). West selected the 2,000 word families on the GSL from a corpus of five million words. His selection criteria for the chosen word families revolved around frequency of occurrence, learnability, and usefulness of the concepts represented by the selected terms. The resulting word list covers 90% of words in works of fiction and 75% of nonfiction texts. While this list covers a great deal of language, academic texts have been found to include a number of

important, specialized terms that are not found on the GSL (Xue & Nation, 1984; Coxhead, 2000).

A second important word list for academic study came in the form of Xue and Nation's (1984) previously discussed University Word List (UWL). The UWL contains 836 headwords that are not found in the GSL, but frequently appear in a wider range of academic texts. Xue and Nation took care to use a corpus based on a wide range of 19 different academic disciplines to include words that are found across the board in academic study (1984). The UWL makes up approximately 8% of the terms found in academic texts.

However, there was a need for a more categorized list. As a result, Coxhead (2000) developed the Academic Word List (AWL) to provide NNS learners with a word list based on a larger, more structured corpus than its precedents. She created an academic corpus from 414 academic texts from the disciplines of arts, commerce, law, and science. From this large bank of words, Coxhead generated the AWL, a list containing 570 word families that frequently occurred across all four areas of academic study. Combined with West's GSL, the AWL covers 86% of academic vocabulary (Coxhead, 2000).

Lexical bundles. Corpus linguistics has confirmed what many teachers have thought for years that much of the English language may actually be comprised of pre-fabricated multiword phrases that are grouped based on meaning and form (Biber & Conrad, 2001; Biber, Conrad, & Cortes, 2004; Byrd & Coxhead, 2010; Hyland, 2008). Research identifies various terms such as multiword units, lexical bundles, lexical

phrases, formulaic phrases, routines, and formulas to refer to these phrases made up of multiple words (Biber et al., 2004). Lexical bundles has routinely been the term of choice for reference in creating words lists containing multiword units in order to adopt a frequency approach to studying these phrases (2004; Byrd & Coxhead, 2010). In terms of vocabulary size and richness, these lexical bundles can aid writers to increase fluency and naturalness within their compositions.

In an analysis of the T2K-SWAL, a corpus of spoken and written language encountered in US universities in a variety of academic contexts such as classroom lectures, study groups, campus service encounters, and institutional writing, Biber et al. (2004) aimed to identify which lexical bundles were prevalent within academic study. Results provided evidence that academic prose makes heavy use of lexical bundles that incorporated noun phrases and prepositional phrases such as *the way in which* or *as well as the*. Referential bundles such as *on the basis of* are also prevalent in academic texts. As a result, Biber et al. suggest that even though lexical bundles are not traditionally regarded as single lexical or grammatical items, they function in a similar manner and need to be pointed out to NNS learners in order to increase their fluency (2004).

In the tradition of the AWL, Simpson-Vlach and Ellis (2010) looked into identifying recurrent formulaic phrases within the MICASE and British National corpora. The result of their study was the Academic Formulas List (AFL), a list made up of the AFL Core list of formulaic phrases and the top 200 formulas from the Spoken and Written AFL. The researchers designed this list with the pedagogical intention of aiding EAP students to employ rhetorical techniques such as hedging, asserting arguments,

organizing discourse in a more fluent manner. Their recommendations follow Biber et al.'s (2004) assertion that such formulaic phrases do have lexical properties that allow them to function as single lexical units (Simpson-Vlach & Ellis, 2010).

Word list research points to the importance of a large and targeted vocabulary for NNS learners in academic settings. As it is nonsensical to expect a NNS learner to know all the words in the English language, word lists can aid NNS writers to focus on growing their lexicon in a directed, purposeful manner to meet their communicative goals, especially as their language needs become more specialized as with academic writing. In order for instructors of NNS writers to tailor their pedagogical practices to facilitate this vocabulary growth, it is prudent to investigate how a learner's vocabulary size is measured (Laufer, 1994).

Table 6
Word Lists

Word List	Number of Words	Text Coverage
General Service List (GSL; West, 1955)	2,000 of the most frequent word in English	90% works of fiction 75% of nonfiction texts
University Word List (UWL; Xue & Nation, 1984)	836 headwords from 19 academic disciplines	8% of academic texts
Academic Word Lists (AWL; Coxhead, 2000)	570 word families from four academic discipline categories	86% of academic texts when combined with the GSL
Academic Formulas List (AFL; Simpson-Vlach & Ellis, 2010)	108 core lexical formulas and 200 most frequent lexical formulas	n/a

Unit of Measurement

One of the central issues of debate in measuring productive vocabulary size is the unit of counting. Research in vocabulary size tends to use types, word families, or lemmas to as the counting unit (Nation & Webb, 2011; Schmitt, 2010). A type is simply a technical term for a word that is counted once within a text. A word family is a group

of words that share the same base to which derived and inflected affixes are attached such as *family, familiar, unfamiliar*, etc. A lemma is a base-word plus all its inflected forms as in *favor, favors, and favored*. When attempting to measure vocabulary size, choosing the appropriate unit of counting is essential because depending on the level of vocabulary knowledge measured, the counting unit may vary. Research has found that counting word families is appropriate for measuring receptive vocabulary size, whereas counting lemmas is better suited for measuring productive size (Nation & Webb, 2011; Schmitt, 2010).

Testing Vocabulary Size

In the previous discussion of the words NNS learners need to know, the answer to how many words an L2 academic writer needs to know may at first appear deceptively simple: add the GSL and the UWL or AWL together and the resulting number assures academic writing success. While these words are of great importance in L2 writing, the ability to measure vocabulary size in writing first is fundamental for investigating how the growth of a learner's lexicon is related to his or her vocabulary use and depth of knowledge (Laufer & Nation, 1995). As a result, the majority of indexes of vocabulary size are based on word frequency operating under the assumption that the presence of lower-frequency words signifies a larger lexicon (Xue & Nation, 1984). This section presents several tests that researchers have used to measure vocabulary size (Table 10).

Vocabulary levels test. First, Nation's Vocabulary Levels Test (VLT) (1983) measures a learner's vocabulary size using word family frequency bands ranging from the highest frequency words to the lowest (Table 7). The test presents a target vocabulary

word, a sentence containing a derived or inflected version of the word, and multiple-choice items defining the target word. Learners then select the appropriate definition to the word's meaning within the context of the sentence (see Table 3). As the learners progress in the test, the frequency of the word's occurrence lowers culminating in vocabulary items that are academic or specialized in nature (Nation, n.d.).

Table 7

Examples of Nation's Vocabulary Levels Test (Nation, n.d)

Frequency band	Sample items
First 1000	see: They saw it. a. cut b. waited for c. looked at d. started
Second 1000	maintain: Can they maintain it? a. keep it as it is b. make it larger c. get a better one than it d. get it
Third 1000	soldier: He is a soldier . a. person in a business b. student c. person who uses metal d. person in the army
Eighth 1000	erratic: He was erratic . a. without fault b. very bad c. very polite d. unsteady
Thirteenth 1000	plankton: We saw a lot of plankton . a. poisonous weeds that spread very quickly b. very small plants or animals found in water c. trees producing hard wood d. gray clay that often causes land to slip

The goal of the VLT is to aid instructors to decide what vocabulary items to focus on for different levels of student proficiency (Laufer & Nation, 1999).

Eurocentres vocabulary size test. Another vocabulary size test is Meara and Buxton’s Eurocentres Vocabulary Size Test (1987), which is a computerized placement test in which learners are given a series of words, one at a time, and learners self-report whether they know the word or not by pressing one of two keys on the computer keyboard. The reliability check in the program is that the words are one of two types of items: real English vocabulary words and invented non-words created to mimic English morphology (Meara, 1990). The test operates within two blocks where the first block focuses on high-frequency words and the second increases in difficulty, presenting lower frequency words. Each block mixes genuine and non-words (1990; see Table 8). This allows for a large number of vocabulary items to be tested in a short amount of time, making it an ideal diagnostic test (Laufer & Nation, 1999).

Table 8

Examples from Meara and Buxton’s Eurocentres Vocabulary Size Test (Meara, 1990)

Block a	adviser morlorn weekend	ghastly patiful discard	contord profess disdain	implore stourge gleanse
Block b	mascule penepplain leat	palangane rangue prunella	bezel aliver kellick	orduad quoddity windlestraw

However, critics of the Eurocentres test have argued that the test does little more than assess passive receptive vocabulary size and question the effect the non-words have on the test-takers (Meara, 1990). Therefore, researchers who assert that NNS learners’ “productive vocabulary ability is not a yes/no phenomenon” cite the importance of

assessing vocabulary size within the context in which the words would be produced (Laufer & Nation, 1999, p. 36).

Productive vocabulary levels test. One such measure is Laufer and Nation's (1999) Productive Vocabulary Levels Test (PVLТ). In this assessment, the test solicits vocabulary knowledge within the context of a meaningful sentence and the first few letters of the target word is given. Test-takers then fill in blanks with the remaining part of the word (see Table 9).

Table 9

Examples from Laufer & Nation's Productive Vocabulary Levels Test (1999)

Frequency band	Sample items
2000 word level	<ol style="list-style-type: none"> 1. I'm glad we had this opp___ to talk. 2. There are a doz___ eggs in the basket. 3. Every working person must pay income t___.
5000 word level	<ol style="list-style-type: none"> 1. Soldiers usually swear an oa___ of loyalty to their country. 2. The voter placed the ball___ in the box. 3. They keep their valuables in a vau___ at the bank.
UWL word level	<ol style="list-style-type: none"> 1. There has been a recent tr___ among prosperous families towards a smaller number of children. 2. The ar___ of his office is 25 square meters. 3. Phil___ examines the meaning of life.
10,000 word level	<ol style="list-style-type: none"> 1. The prisoner was released on par___. 2. Her favorite flowers were or___. 3. The dead bodies were mu___ beyond recognition.

The rationale behind providing the first few letters on the test is to control for the learner filling in the blank with a synonym or semantically related term. Again, this test is organized according the frequency bands of the target words like the VLT (1999). In a study with 79 EFL students in Israel ranging from 10th grade to freshmen at a university,

Laufer and Nation (1999) tested the reliability and validity of the PVLТ and found that it is a practical measure of learners' vocabulary size growth.

Lex30. Fitzpatrick and Clenton (2010) tested the ability of a computational linguistics program called the Lex30 to estimate lexical size. This program attempts to minimize any influence of receptive vocabulary knowledge to focus solely on productive ability. To achieve this, the program uses word association tasks to elicit target words. Results are then scaled and scored according to the frequency of the produced words. The total possible score is 120 points; however, the researchers report that native speakers rarely score above 70 and non-native speakers' scores range from 10 to 40. In Fitzpatrick and Clenton's (2010) study, they used the Lex30 to estimate the vocabulary size of 40 Japanese EFL medical students. The participants' results from the Lex30 were then compared with the Productive Vocabulary Levels Test to check validity (Laufer & Nation, 1999). Findings suggested that the Lex30 produced accurate scores with no significant differences between the computer tool and the Productive Vocabulary Levels Test. These results suggest that the Lex30 succeeds in its ability to use word associations to measure the vocabulary size of learners.

Testing Productive Vocabulary Size in L2 Writing

However useful the above tests are in measuring vocabulary size, they do not measure vocabulary size as authentically used by NNS writers (Crossley et al., 2010). Although the PVLТ (Laufer & Nation, 1999) provides some contextualization of the target words within meaningful sentences, it does not give an accurate picture on how NNS writers deal with actually producing these target words during the composition

process (Crossley et al., 2010). Therefore, researchers looked for measures to test vocabulary size in learner-produced texts.

Subsequently, with the increasing sophistication of computer technology, researchers began to see the benefit of measuring productive vocabulary size in natural-occurring contexts such as free-writes, summaries, and compositions (Crossley et al. 2010). Computer programs allow for investigators to examine longer texts rather than multiple-choice or fill-in-the-blank items. Computers also allowed vocabulary size studies to increase efficiency and remove the variables of human subjectivity and inter-rater reliability in measuring vocabulary size in production by developing computer-profiling programs to count word units (2010).

Lexical frequency profile. Laufer (1994) proposed the aforementioned LFP as a reliable measure to estimate the size of a NNS writer's vocabulary size. The LFP creates a profile of a learner's vocabulary based on their use of high- versus low-frequency words (1994; Laufer & Nation, 1995). The LFP program can be set to calculate tokens, types, or word families. In a longitudinal study to test the effectiveness of the LFP to gauge vocabulary size, Laufer (1994) analyzed free-write compositions by 48 EFL university students in Israel using the program. Different compositions made by the same author were processed through the LFP three times during the study. Analysis demonstrated that learner's vocabulary size increased over time as indicated by their writing samples, signifying that productive vocabulary size could be measured rather well using the LFP.

In the Laufer and Nation (1995) study, a more optimized version of the LFP was used to compare results of the LFP to the VLT. The researchers administered Nation's (1983) Vocabulary Levels Test (VLT) to the 65 participants, which elicited the use of target words at varying frequency levels in English. The results of the VLT were then compared to the LFPs of the participants' compositions. First, results demonstrated that higher proficiency students possessed larger vocabulary sizes as demonstrated by their writing samples due to the inclusion of lower-frequency vocabulary items. Secondly, the comparison of the VLT to the LFP indicated that it is possible that learners' vocabulary size will be reflected within their compositions.

P_Lex. However, Meara and Bell (2001) criticized Laufer and Nation's LFP, stating it was a problematic measure of vocabulary use in writing largely due to first, the program's preference for lower-frequency vocabulary and second, that it requires long texts to function. In its place, Meara and Bell propose a different profiler called P_Lex that is also based on frequency like the LFP, but is a more reliable predictor of productive vocabulary size in shorter compositions (Schmidt, 2010). They conducted a study testing P_Lex's capabilities with 49 EFL students in a summer course at the University of Wales Swansea. The researchers replicated the Laufer and Nation study, administering a VLT to the participants and even having them write on the same topics (Meara & Bell, 2001). Results demonstrated that P_Lex profiler produced similar results as the LFP but worked just as well with shorter texts and with all levels of word frequency.

Meara again challenged the LFP (2005) by running a series of Monte Carlo simulations to test the LFP's sensitivity to detect small changes in vocabulary size.

Meara's Monte Carlo simulations relied on a series of recalculations of results, each time using different random samplings from Laufer and Nation's (1995) data to retest their conclusions. Analysis of the outcomes and distributions produced suggested that the LFP "did not reliably produce strong correlations between scores on different texts produced by the same author" nor within large sample sizes (Meara, 2005, p. 46). Meara claimed then that the LFP was only reliable enough to detect large changes in vocabulary size.

Zipf's law. Continuing the debate over the reliability of the LFP, Edwards and Collins (2011) evaluated the validity of the LFP's construct using mathematical equations based on Meara's (2005) Monte Carlo simulations. Edwards and Collins (2011) combined a mathematical equation with Zipf's law to reexamine the reliability of the LFP's results. Zipf's law states that if words are ranked by their frequency of use in natural language, then "the most frequent word is twice as frequent as the second most frequent word, three times as frequent as the third most frequent word, and so on" (Edwards & Collins, 2011, p. 4). The researchers thus created a mathematical formula using the same frequency bands from the LFP and corroborated Laufer and Nation's findings that as learners' vocabulary sizes increase, the number of words found in the high-frequency bands goes down. The equation also revealed that while the LFP is not sensitive enough for large samples, confirming assertions by Meara (2005), the instrument is reliable enough for around 20 participants, which is a group size often available for research studies (Edwards & Collins, 2011).

Coh-Metrix. Advances in computational linguistics have led to powerful software that can use a variety of multi-leveled measures to analyze lexis in writing. The

Coh-Metrix program analyzes texts for cohesion based on the rationale that purposeful vocabulary leads to better connectedness in composition and thereby calculates nuances in lexical richness and diversity (Grasser, McNamara, Louwerse, & Cai, 2004). One of the advantages of using Coh-Metrix is that it utilizes the CELEX corpus frequency lists that uses the lemma as its unit of counting, which researchers argue more accurately depicts learners' vocabulary size (Nation & Webb, 2011; Schmitt, 2010). Crossley and McNamara (2009) compared the differences between NS and NNS writers using Coh-Metrix. They found that NS writers used significantly more words than NNS writers. This suggests that in order to truly succeed in academic writing, NNS writers need a larger vocabulary bank of words in order to make their compositions longer.

Receptive vocabulary size as a predictor of productive vocabulary size.

Computer programs such as the LFP, P_Lex, Lex_30, and the Coh Metrix are a few of the valid methods for measuring productive vocabulary size. However, studies have indicated that learners only produce a fraction of the words they actually know (Laufer, 1998; Waring, 1997). Therefore, research has looked into using receptive knowledge to measure a learner's productive vocabulary size and how the two lexicon types differ.

In a study of 73 female Japanese EFL learners of English enrolled in a university, Waring (1997) used the VLT and the PVLTL in order to investigate the difference between receptive and productive vocabulary sizes in terms of percentage. The participants were recruited from three different proficiency levels in order to examine the lexicon sizes of varying levels of language competence. The tests were scored twice, once by Waring himself and then cross-checked by another native speaker. Results revealed that

receptive vocabulary size is larger than the productive lexicon and is hypothesized that productive capability is acquired at a later stage in language acquisition.

Likewise, Laufer (1998) investigated EFL learners' growth in vocabulary to examine how passive (receptive) and active (productive) changed over the course of a year. The three instruments used to measure receptive and productive vocabulary were the VLT, the Productive VLT, and the LFP. Results demonstrated that while the learners' receptive lexicon demonstrated growth over time, their productive vocabulary size remained relatively unchanged (1998). This finding provided evidence that even though NNS learners' receptive vocabulary size can increase significantly over time, they did not transfer their new word knowledge into their writing when assigned a composition task.

Webb (2008) conducted a study with 83 Japanese EFL students in a Japanese university to see if receptive vocabulary size could predict the size of a learner's productive vocabulary. Webb selected 180 target words, categorized by frequency, from the COBUILD dictionary. The participants were given a receptive and productive translation test. On the receptive test, the participants translated L2 target words to Japanese and responses were scored twice at two levels of sensitivity, sensitive and strict, by two bilingual native speakers of Japanese. On the productive test, participants were provided with the L1 meanings of the target words in Japanese and asked to write the English translation. Results were twofold. First, findings demonstrated that receptive vocabulary size is larger than productive vocabulary size, upholding Laufer's (1998) and Waring's (1997) findings. Second, analysis indicated that learners with a larger receptive

vocabulary are likely to be able to use those words in production that learners with a smaller receptive lexicon (Webb, 2008).

Also looking into how receptive vocabulary size affects its productive counterpart, Staehr (2008) performed a study of 88 EFL learners in secondary schools in Denmark to examine the impact of reading, listening, and writing on vocabulary size and to investigate if a vocabulary threshold exists in order to read, listen, and write in the L2. Participants were given the VLT to measure their vocabulary size, then a multiple choice reading comprehension test, a multiple choice listening comprehension test, and a writing prompt to write a job cover letter. Results provided further evidence that a large vocabulary size is critical for reading and writing and slightly less so for listening. Comparisons between the reading, listening, and writing scores revealed that knowing the most frequent 2000 word families aided participants the most on the writing test (2008). Participants whose VLT revealed they lacked knowledge in the first 2000 frequent words still were able to perform well on the reading and listening tests but failed the writing portion, suggesting that the productive vocabulary threshold to succeed in writing is higher than receptive vocabulary thresholds (2008). More than half the variance in writing score was accounted for by vocabulary size, lending more insight into how many words NNS writers would need to know to perform academic writing tasks.

Table 10

Review of Measures of Productive Vocabulary Size

Instrument	Strengths	Limitations
VLT (Nation, 1983)	Word frequency-based; multiple-choice test	Measures receptive knowledge only; limited context
Eurocentres (Meara & Buxton, 1987)	Computerized; use of invented non-words	Decontextualized; does not account for partial word knowledge; measures receptive knowledge only
PVLT (Laufer & Nation, 1999)	Word frequency-based; contains primes to elicit target words	Limited range of word-frequency bands
Lex30 (Fitzpatrick & Clenton, 2010)	Uses word associations; word frequency-based	NS rarely score perfectly
LFP (Laufer & Nation, 1995)	Measures vocabulary size in authentic compositions; profiles a text's vocabulary use based on word frequency	Tends to require long texts to accurately profile vocabulary
P_Lex (Meara & Bell, 2001)	Word frequency-based; measures vocabulary size in short texts	Reliable for short texts
Zipf's Law	Simple mathematical formula to determine word frequency	Highly sensitive to text length
Coh-Metrix (Grasser et al., 2004)	Contains over 60 lexical indices; computer-based; uses CELEX log lemmas to gauge word frequency	Does not use Nation's more widely used frequency bands
Receptive vocabulary size (Laufer, 1998; Waring, 1997; Webb, 2008)	Provides insight into learners' lexicon and what words could potentially use in production	Receptive word knowledge tends to be greater than productive knowledge

The research in this section provides insight into the various instruments used to measure vocabulary size and how they can be employed to determine how the size of a learners' lexicon in written productive tasks. While a conclusive figure has yet to be narrowed down, the studies do suggest that because receptive vocabulary sizes tend to be larger than productive estimates, NNS writers may need explicit guidance to develop their productive vocabulary repertoire (Laufer & Nation, 1999; Staehr, 2008). This confirms other research that instructors cannot rely on vocabulary to be learned and acquired incidentally (Nation, 2006).

How Many Words Does a NNS Writer Need to Know?

Despite the numerous empirical studies investigating vocabulary size and writing, a clear-cut number of how many word families is needed to write academic compositions has yet to be produced in the literature. Staehr's (2008) study indicated that knowledge of the first 2000 frequent word families was a central threshold for his participants. Learners who passed the 2000 word family threshold tended to perform above average on writing tasks. However, this figure lacks the specialized, academic terms that are expected in academic compositions where less frequent words account for 8% of the language used (Xue & Nation, 1984; Coxhead, 2000).

Determining a concrete figure for how much vocabulary is needed to write in academic contexts is a tricky task (Staehr, 2008). Nation and Webb (2011) go as far as to say that it may not even be possible given that writing only reveals a portion of a NNS writer's lexicon because word choice depends heavily on the task and writers are not likely to use every word they know within a composition (Nation & Webb, 2011; Staehr, 2008). However, some preliminary figures of 8000 to 9000 word families based on the 98% receptive vocabulary threshold proposed by Nation offer a place to start (2006). As Laufer (1998) and Staehr's (2008) studies indicate, the threshold for written production may be an even higher figure than 9000.

Measuring Lexical Diversity

Since lexical diversity is typically considered an "end-product" of language, it is used as a measure of the productive capabilities of speaking and writing (Yu, 2009). The

literature has utilized a number of terms interchangeably with lexical diversity. Lexical variation, individuality, range, and balance all describe the use of a wide range of different words within discourse (Read, 2000; Yu, 2009). The theoretical principle underlying the foundation for the correlation between lexical diversity and language proficiency is situated in that a greater number of different words produced in spoken or written discourses leads to a higher quality of language produced (Daller, Van Hout, & Treffers-Daller, 2003; Durán, et al., 2004; Engber, 1995; Jarvis, 2002; Laufer & Nation, 1995; Yu, 2009).

In terms of writing proficiency, empirical L1 and L2 writing studies have corroborated and reinforced this relationship with results showing that raters perceive compositions demonstrating a greater proportion of different words as higher quality writing (Arnaud, 1984; Engber, 1995; Linnarud, 1986; Lovejoy, 1991). Furthermore, research has shown that NNS learners' written lexical diversity is significantly much lower compared to NS learners (Linnarud, 1986).

Accurately and reliably measuring lexical diversity has been a challenge for vocabulary researchers. Nor has it garnered the amount of attention that vocabulary size has enjoyed in the literature. The overarching source of computing a text's use of varied vocabulary lies in text length: the longer the text, the likelihood of new words appearing is reduced (Durán, et al., 2004; Graesser et al., 2004; Jarvis, 2002; McCarthy & Jarvis, 2010). In other words, the first paragraph of any given text exhibits a wider variety of unique words whereas paragraphs number 50, 100, and 150 likely contain words that have been used previously. As a result, longer texts appear less diverse. Therefore, it is

difficult to ascertain if a particular measure of lexical diversity is reliably measuring vocabulary variation or text length. A number of techniques and instruments have been devised to reliably and sensitively assess vocabulary diversity in text (see Table 11).

Type-Token Ratio

The traditional method of quantifying lexical diversity is the type-token ratio (TTR; Arnauad, 1984; Jarvis, 2002; Laufer & Nation, 1995; McCarthy & Jarvis, 2007; Tweedie & Baayen, 1998). Types are the unique words that are counted the first time they appear within a text, whereas tokens count each word, each time it appears. For example, the sentence, “the white car is larger than the red car” contains the types, *the*, *white*, *car*, *larger*, *than*, and *red* for a total of six types. However, this sentence has nine tokens because the total number of words equals nine. Lexical diversity is then measured by calculating the ratio of types to tokens (measuring in a range of 0 to 1) by dividing the total number of types by the total number of tokens. For the example sentence above, the TTR is .78.

While the advantage of using TTR to examine lexical diversity is its ease of use, its fatal flaw is that it is highly impacted by text length (Laufer & Nation, 1995; Tweedie & Baayen, 1998). This makes using authentic writing samples in lexical diversity studies very difficult because of a number of reasons. First, learners do not produce texts containing exactly the same number of words. This could be controlled in two ways: (a) assigning a writing task where learners are restricted to a specific number of words (McCarthy & Jarvis, 2010), or (b) sampling only the first set number of words in essays

(Biber, 1988). Manipulation of the number of words, however, affects natural vocabulary use and thus becomes a limitation.

Ratner and Silverman (2000) proposed the correction to TTR by only reporting the total number of different words (NDW), i.e., the types contained within the text. Thus, the greater the number of types indicates more diversity. However, text length continues to affect diversity because the NDW depends on how many tokens are available from which to extrapolate the types (McCarthy & Jarvis, 2010).

Voc-*D*

The *D* measure of lexical diversity is one of the first diversity formulas proposed as able to reliably control for length of text (Durán, et al., 2004; Malvern & Richards, 1997). Using a computer program called *voc-D*, the *D* statistic figures diversity using a formula that randomly samples 35 to 50 tokens from a text 100 times and fitted to a theoretical curve. Then, a TTR score is calculated for each of the samples and a mean is produced resulting in *D*-scores for each sample. Next, each of the *D*-scores is averaged together. This procedure repeats three times. Finally, an average *D*-score from the entire process forms a more reliable lexical diversity rating (Durán et al., 2004).

Durán et al. (2004) conducted extensive tests of the lexical diversity of transcripts spoken English produced by 32 pre-school children over three years in Bristol. They used the *voc-D* program to examine if lexical diversity in children's speech increased with maturation by comparing *D*-scores to participants' age and mean length of structured utterances (MLUS). Historically, when TTR has been applied to developmental trends in children's speech, lexical diversity decreases clearly

exemplifying the unreliability of TTR as a measure (Durán et al., 2004). Voc-*D* results revealed that children's lexical diversity increased in tandem with age and MLUS, validating voc-*D* as able to control for text length.

However, *D* is not without limitations. McCarthy and Jarvis (2007) assert the *D* has two critical limitations: (a) its random sampling procedures produce a different *D* coefficient when computed multiple times, and (b) text length still affects the coefficient for longer texts. To test these assertions, McCarthy and Jarvis (2007) ran a database of 266 texts, each of which was roughly 2,000 tokens in length, through a series of analyses using the voc-*D*. The researchers found that *D*, despite being a valid measure, is only reliable for texts likely containing low diversity, such as children's or lower proficiency NNS learner language (McCarthy & Jarvis, 2007).

The MTL D

The Measure of Textual and Lexical Diversity (MTLD) is the latest approach to quantifying lexical diversity. The MTL D operates within the principle that, given the length of any text, the TTR score at the beginning of any text will be closer to 1 and progressively falls towards 0 in tandem with text length (McCarthy & Jarvis, 2010). The MTL D attempts to calculate the TTR at the precise point of decline. Pearson correlations (2008) found a general TTR of 0.72 as a reliable point of analysis. The MTL D then counts how many times a TTR of 0.72 occurs within the text and divides the resulting figure into the total number of tokens. Average scores lie between 70 and 120, with the latter score indicating the greatest lexical diversity (McCarthy & Jarvis, 2010). The final calculation step runs the entire process backwards through the text primarily as a

reliability check for the .72 cutoff score. Therefore, the MTLT solves some of the accuracy and reliability issues that *D* possessed.

Currently, the largest limitation of using the MTLT is that it is still new and in need of studies further validating its effectiveness, especially applying it to a particular corpus. McCarthy and Jarvis (2010) admit that it remains untested if reliable under 100 tokens or beyond 2000 tokens, their critical limits for text dependency (McCarthy & Jarvis, 2007). The developers also suggest further validation using a variety of writing genres, as topic and task may affect lexical diversity (Yu, 2009). However, when the MTLT scores were subjected to Pearson correlations against TTR, NDW, and *D* within the two corpora used in the validation study, the MTLT demonstrated that it better controlled for text length than the more established measures (McCarthy & Jarvis, 2010).

In sum, accurate measures of lexical diversity have been challenging to devise, but the MTLT offers encouragement that vocabulary variation can be calculated without suffering the effects of text length. Comparing the performance MTLT score and CELEX word frequency figures to writing score may help to enlighten the contribution markers of lexical proficiency on the quality of composition. The previously discussed computer-based textual analyzer, the Coh-Metrix, includes the CELEX word frequency measures and the MTLT. In addition, the Coh-Metrix allows for validation of the MTLT score using the more established *voc-D*.

Table 11

Measures of Lexical Diversity

Instrument	Strengths	Limitations
Type-Token Ratio	Easy to use measure of lexical diversity	Highly affected by text length, i.e., the longer the text, the less reliable it gets
<i>D</i> (Malvern & Richards, 1997)	Randomly samples 35-50 tokens to compensate for text length	Prefers short texts; moderately affected by text length
MTLD (Jarvis & McCarthy, 2010)	Randomly samples 50 tokens forwards and backwards to compensate for text length; calculates lexical diversity at precise point of decline; reliable for essays of 100+ tokens	It is new; needs further validation with long texts to see if reliable under 100 tokens and beyond 2,000 tokens

Vocabulary Size versus Lexical Diversity as Predictors of Writing Quality

Because of the complex nature of lexical knowledge and use, there is an assumption within the field that the ability to vary words in discourse requires a large vocabulary size (Laufer, 1994). Daller et al. (2003) claim that lexical diversity cannot be measured without including a measure of lexical sophistication, in which they operationalize as higher instances of rare and less frequent vocabulary items. In other words, a text that contains a greater percentage of rare words would garner a higher score of lexical diversity than a similar text that still scores a high TTR, but uses more general terms. Aforementioned research into linguistic thresholds certainly indicates that a larger vocabulary increases learners' ability to vary vocabulary items in use (see Zareva et al., 2005). However, very few studies have further explored the accuracy of this claim. In fact, there are indications that mid-range vocabulary items might explain some of the variance in the lexical profiles between NS and NNS.

In fact, there is preliminary evidence that vocabulary size does not correlate well to vocabulary variation. Laufer (1994) in a study of learner lexical profiles found no significant correlation between learners increase in vocabulary size and their progress in terms of lexical diversity. This apparent incongruity suggests “learners who are able to vary their lexis well are not necessarily those who have the richest vocabulary in terms of size” (1994, p. 30). Therefore, quality may be more closely related to learners’ ability to express themselves with the words they already possess within their own lexicon and not the recurrent use of low-frequency words (1994). Yet, research relating to the vocabulary size for linguistic ability indicates that L2 lexicons must pass a certain lexical threshold in order to successfully perform language tasks (Cummins, 1986; Nation & Hwang, 1995). Few studies to date (with the exception of Laufer, 1994) has expounded on this seemingly contradictory relationship thus leading to the present investigation of two lexical richness measures in composition.

Conclusion

The body of literature reviewed indicates that the size of NNS learners’ lexicon clearly impacts their ability to write well in the L2; however, attributing writing quality to the number of sophisticated words learners produce does not take into account that quality writing sometimes relies on the use of accessible, layman terminology. Evidence from the studies correlating learners’ lexical breadth to writing achievement demonstrate that greater word knowledge leads to better essays that are more likely to earn higher scores by raters. Likewise, research investigating what lexical features contribute to

higher writing quality show that NNS learners who possess a large, sophisticated, and varied lexicon are better able to perform writing tasks than learners who do not possess as rich of a vocabulary. Lastly, lexical knowledge and lexical richness both require NNS learners to know a lot of words in order to successfully compose an academic essay. Studies examining vocabulary size indicate that NNS learners' productive lexicons tend to be smaller than their receptive equivalent. However, there is anecdotal evidence that too many sophisticated words can lower writing quality, but little empirical research has investigated these seeming opposing tenets of proficient writing. Considerably more studies have focused on what Cobb has termed the "outer edges of vocabulary" rather than looking at learners' use of the mid-range vocabulary items (2013). Consequently, the state of the research could benefit from attempts to examine authentic academic uses of lexical size and variation for proficient L2 writing purposes.

CHAPTER 3: RESEARCH METHODOLOGY

For instructors of non-native speaking (NNS) learners preparing and intending to study in a U.S. college classroom, examining how native speaking (NS) learners utilize the full richness of their lexicon in writing in comparison to NS learners can inform curricula and classroom practices to justify and increase teacher and learner attention to lexis in English as a second language academic preparation programs. As detailed in Chapter Two, research has revealed that lexical diversity is an important predictor of writing quality; however, the relationship between vocabulary size and the ability to achieve lexical diversity remains unclear (Laufer, 1994).

Therefore, the primary objective of the present study was to profile the specific lexical richness measures of vocabulary size and diversity within advanced NNS and NS learner compositions in order to investigate: (a) which measure results in more proficient writing, (b) if there is a correlation between these two measures, and (c) if there is a difference between advanced NNS and NS learners' lexical richness profiles. The present chapter first provides a general overview of research design followed in this study to meet the research aims followed by a detailed description the methods undertaken including the research setting, population and sampling procedures, data collection procedures, and instrumentation. The chapter then concludes with the detailed research questions, hypotheses, and statistical analysis procedures carried out to garner results.

Orientation to Research Design

Based on the above research objectives, this study undertook a quantitative, lexical analysis approach for data collection and analysis. A corpus of NNS and NS learner essays gathered from post-secondary composition courses was created and analyzed using an online textual profiler tool, the Coh-Metrix (Graesser, McNamara, Louwrese, & Cai, 2004), freely available for studying the productive features of compositions. For the quantitative analysis, a correlational design was implemented to answer the main research question of the ability of the independent variables of vocabulary size and lexical diversity measures to predict the dependent variable of writing achievement. Since the dependent variable measuring writing score is a categorical variable, a binary logistic regression was selected as the statistical test.

For the first research sub-question, a test of statistical differences compared NNS and NS writers' scores. Figure 2 provides a visual overview of the research design that will be further elaborated in subsequent sections. The second sub-question to determine if vocabulary size predicts lexical variation also used a correlational test. Because both the independent variables are continuous data, a Pearson product moment correlation was an appropriate statistical test.

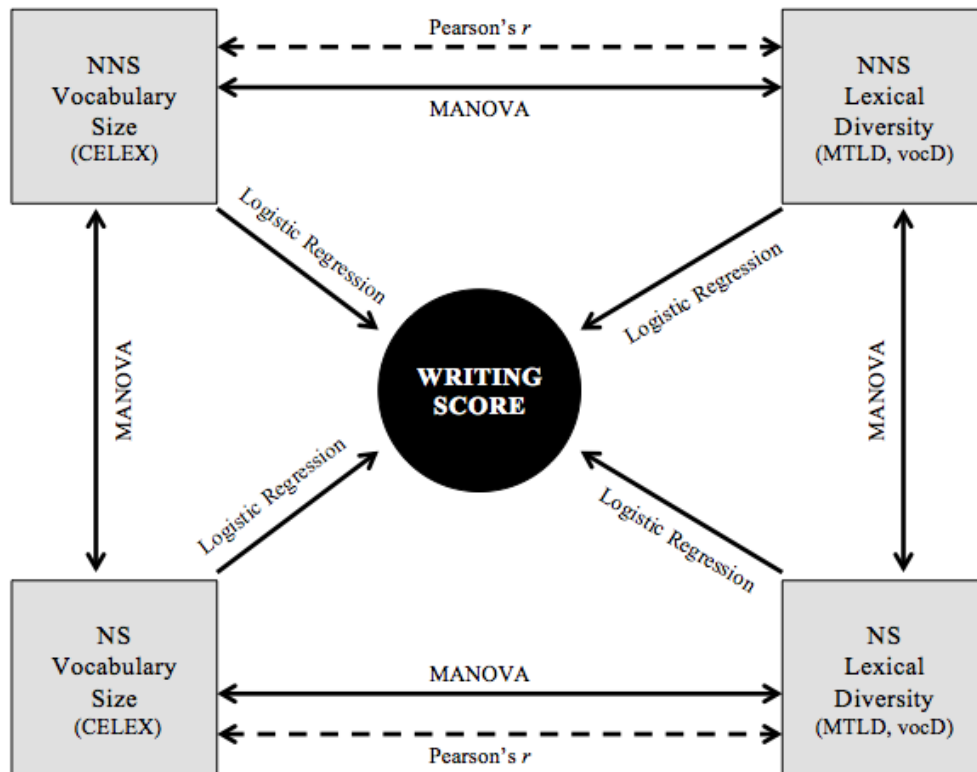


Figure 2. Orientation to research design

Research Setting and Population

The study involved data collected from five intensive English programs affiliated with post-secondary institutions across the United States, and one large, public university located in the southeastern United States. To protect the anonymity of the institutions and participants involved, the participating institutions are coded using alphabetical designations to which only the researcher is privy.

The Intensive English Programs

The advanced NNS learner participants are students enrolled in six intensive English programs (IEPs) affiliated with post-secondary colleges and universities in the

United States. In order to facilitate the language transition and prepare NNS learners for the mainstream college classroom, these institutions offer rigorous English courses designed to prepare learners for the academic language used within English-only college courses. Typically, NNS learners matriculate within IEPs prior to admittance into an American college or university in order to improve their English academic skills. The participants in this study were from a variety of language backgrounds and were in their second to last or last semester of IEP study with the intention to apply for admittance into a U.S. college or university at the time of data collection. The advanced-level writing courses at each institution is a required course for all IEP students in the higher levels of study and generally follow a process-oriented approach to writing instruction. The composition curriculum at each IEP focused on the teaching and learning of typical college writing genres and tasks.

First-Year Composition

The NS learners who participated in this study are enrolled in four sections of a first-year composition class at a large, public university in the Southeastern United States. The first-year composition course (ENC 1101) is housed within the Department of Writing and Rhetoric at this university. The ENC 1101 course is an undergraduate requirement for every admitted student and is designed to aid learners in the mastery of a variety of college-level writing tasks. Instructors of ENC 1101 usually are graduate students in the Writing and Rhetoric Master's or Doctoral degree programs fulfilling a graduate teaching assistant requirement. The first-year composition curriculum also follows the process-writing approach to develop writing. This population of students was

sampled in order to provide a basic profile of the typical language evaluated in college compositions.

Research Population Determinations

There were several decisions that guided the selection of the population and sample used in this study. First, the choice to limit the L2 population to those learners at the highest levels of English proficiency assured that the study participants were at the brink of taking the Test of English as a Foreign Language (TOEFL) or the International English Language Testing System (IELTS) assessments required for admission into an American university. Minimum score requirements on English language proficiency tests such as the TOEFL and IELTS are used by post-secondary institutions to ascertain if L2 students meet satisfactory levels of academic language proficiency to participate in college-level coursework. Given this admission requirement, NNS learners in the advanced courses are at the end of their English language study and profiling the lexical richness of their compositions gave insight into how well prepared these learners are for the language demands of college writing tasks. The NNS learners within this sample all have passed their respective institution's minimum requirements to enroll in the advanced level writing course.

Secondly, the decision to include participants from a variety of language backgrounds ensured that: (a) essays from various linguistic backgrounds were represented within the corpus, (b) the sample mirrored typical IEP classroom demographics, and (c) essays displayed more accurate lexical profiles found in diverse ESL classrooms.

The third decision to include NS learner compositions from a first-year writing course within the corpus served the dual purposes of providing a typical lexical profile of novice NS writers and a comparison group to predict lexical impact on writing score. Moreover, the inclusion of NS peers was able to provide a baseline of the typical lexical resource of novice college writers. The primary objective of the ENC 1101 is to “provide an entry point to university-level writing practices and instruction” (First-Year Writing Program, n.d.). Therefore, the course intends to equip freshmen writers with knowledge about writing that is transferable to later writing situations encountered in post-secondary level study. As such, the inclusion of NS essay data from this course allowed for a comparison group within the corpus to examine if NNS learners preparing to exit the IEP are ready for the even the most novice lexical demands of college writing of such a course.

Lastly, the decision to only draw NS essays from one university whereas L2 essays arrived from five different IEPs could be viewed as a potential limitation. However, first-year composition courses contain students who matriculated from schools throughout the county bringing with them a variety of writing experiences. Furthermore, graduates from IEPs do not typically enroll into the university to which their IEP is connected. Lastly, the goal behind including L1 writing samples was to provide an example assignment that NNS learners would potentially encounter once enrolled in a college composition course. Consequently, the population determinations used in this research mirrored normal enrollment practices.

Sampling Procedures

The study followed the convenience sampling method using intact classes of NNS learners in advanced English writing courses and NS freshmen in first-year composition courses at the participating institutions as detailed in the research setting above. The convenience sampling method was selected at the so as to target a population that suited the research objectives and guaranteed that the use of essay data was voluntary (Fraenkel & Wallen, 2009). As a result, 112 essays were collected from the advanced L2 writing courses, and 71 essays were gathered from four sections of the first-year composition course to yield a corpus that comprised 183 academic compositions. Procedures on how the study arrived at these figures are detailed in the following sections.

Sample Size Determinations

For correlational research, Fraenkel and Wallen (2009) suggest a minimum of 30 participants to yield data to run Pearson product moment and logistic regression analyses. However, a binary logistic regression analysis requires a sample size that is appropriate for the number of independent variables (2009). Shavelson (1996) indicated that ten cases per independent variable is a good rule of thumb to satisfactorily achieve the power necessary to generalize the findings to a population.

To ensure a more precise and appropriate sample size for correlational research, Cohen's (1992) statistical power analysis was consulted. The study adhered to the conventional 80% power at a .05 level of significance in order to achieve a medium effect size of .30 thus yielding a desired sample size of 76 essays needed to conduct statistical analyses. Recruitment of the study participants thus proceeded with this number in mind.

Recruitment of Participants

Non-Native Speaking Learners

Participation in the study was on a voluntary basis. The six IEPs that contributed learner essays were recruited through a call for research participants on the professional teacher's listserv associated with directors, coordinators, and instructors of NNS learners within IEPs. An explanation of research and the approved protocol was distributed to instructors for distribution to the NNS learners in their classes (see Appendix for a copy of the *IRB Approval and Explanation of Research*). Instructors were given the option of emailing scanned or electronic essays to the researcher's email address or mailing hard copies to the researcher with each learner's L1 and language level identified.

Native Speaking Learners

The director of the first-year composition program aided the researcher to contact instructors of ENC 1101 course sections. The instructor of four sections volunteered to submit essays from their existing students' coursework with learners' L1s identified. All learners received the explanation of research and those who wished not to participate in the study had the opportunity to exclude their essays from the data set.

Raters

Two volunteer raters were recruited to evaluate the quality of the essays. The first rater was a non-native speaker with a Master's degree in Teaching English Speakers Of Other Languages from a U.S. public university and IEP composition instructor. The second rater was a native speaker instructor of ENC 1101 in the Writing and Rhetoric department with a Master's degree in Rhetoric and Composition. A third rater, an

English for Academic Purposes (EAP) instructor and Master's degree in TESOL, was consulted in the few cases of conflicting scores to provide a third score to reconcile rating differences.

Data Collection Procedures

Upon approval from the University of Central Florida's Institutional Review Board (IRB), data collection procedures occurred in the following steps:

1. Participating instructors of advanced L2 composition and ENC 1101 courses emailed or mailed ungraded final drafts of student essays with L1s, gender, and course level indicated to the researcher.
2. Identifying information relating to names, institutions, or instructor names was removed from the essays.
3. Raters were trained to rate essays according to ETS' *TOEFL iBT Independent Writing Task Scoring Guide* (2005) that contains benchmark and example essays for each rating.
4. Raters were given a training set of ten essays to rate per ETS' TOEFL iBT Independent Writing Task rubric.
5. Once the raters were satisfactorily trained, they were given to the essays for rating. Nineteen essays of the 183 were returned with conflicting scores; these essays were delivered to the third rater to evaluate and assign a third score.

6. Inter-rater reliability check using Pearson's product moment correlations was performed to validate scores (Fraenkel & Wallen, 2009).
7. Essays were prepared for entry into the Coh-Metrix program according to Graesser et al.'s (2004) and Crossley, Salsbury, McNamara, and Jarvis' (2010) recommendations to remove special characters, correct misspellings for word recognition, delete extra spacing or consecutively repeated words, and correct any spatial formatting issues that could affect the reliable performance of the tool. Lexical errors remained in the essays as the present study is concerned with vocabulary usage not accuracy.
8. The Coh-Metrix ran textual analyses and the results from the *voc-D*, *MTLD*, and *CELEX* log frequency of all words were extracted from the 108 indices and entered into the computer statistical analysis program Statistical Package for the Social Sciences (SPSS) 21.

Instrumentation

Coh-Metrix 3.0

The Coh-Metrix 3.0 is a computational linguistics program developed by Graesser et al. (2004) that produces 108 validated linguistic indices that analyze the discourse representations of a text to examine its cohesive properties. It is freely available online and allows researchers to copy and paste electronic texts into the program for analysis (<http://cohmetrix.com/>). The Coh-Metrix 3.0 identifies both the syntactic and semantic aspects of various texts leading to textual cohesion and computes a score for each index.

For lexical analysis, it contains several useful tools such as word counts, frequency counts, latent semantic analyses, diversity measures, and readability values.

The Coh-Metrix has been found to be a reliable program. Crossley et al. (2010) conducted a study in which 300 essays written by NNS and NS were rated by graders trained to use a rubric of lexical proficiency. The essays were then inputted into the Coh-Metrix tool to analyze the same lexical indices in the rubric to validate the computational tool's ability to perform similarly to human raters' judgments (2010). Results demonstrated a significant correlation between human raters' scores and the Coh-Metrix indices' values ($r = .66$). Therefore, Coh-Metrix is a valid tool for analyzing the lexical quality of a text (2010).

The present study utilized three of the Coh-Metrix 3.0 indices for the analysis of the essays collected: log frequency mean for all words, the *voc-D*, and the MTLD. The psychometrics for each individual index used in this study is explicated below.

CELEX. The measure of word frequency to indicate learner vocabulary size in Coh-Metrix is derived from CELEX (Baayen, Piepenbrock, & Gulikers, 1995), a lemmatized database that contains corpus-based frequencies taken from the 17.9 million-word COBUILD corpus. One million of these words are tokens from spoken English sources such as transcribed TV programs and taped telephone conversations. The remaining tokens within the corpus were sourced from written texts such as newspapers, books, and academic texts (Baayen et al., 1995; Graesser et al., 2004).

The Coh-Metrix provides three CELEX indices available for indicating lexical size and sophistication: the average word frequency for content words (WRDFRQc),

average word frequency for all words (WRDFRQa), and the average minimum word frequency in sentences (WRDFRQmc) (Coh-Metrix, n.d.). For the purposes of this study, the average word frequency for all words for all words in the text (WRDFRQa) was selected because limiting analysis to content words could potentially be affected by topic and task; since the corpus gathered represents a variety of topics and genres, examining the mean word frequency score for all words allowed for better comparison. CELEX word frequency scores range from 0 to 6, with 0 indicating the rarest, most sophisticated words in the text and 6 referring to the most common words used in the English language. The CELEX scores are considered continuous, interval data.

Using frequency counts removes the subjectivity of defining sophisticated vocabulary. Laufer and Nation (1995) argue that word frequency information is more suitable for research that intends to compare groups from different learning environments. Meara and Bell (2001) corroborate the use of word frequency to measure vocabulary size as “people with big vocabularies are more likely to use infrequent words than people with smaller vocabularies are” (p. 4). This is further supported by research demonstrating that higher frequency words are easier to learn because they are shorter, occur more often in production, (de Groot, 1992; Nation, 1984; Meara and Bell, 2001).

There were a couple of reasons that guided the decision to utilize CELEX word frequency scores to measure the productive vocabulary size of the sample essays. First, the CELEX scores are calculated by counting lemmas as opposed to word families. For examining learners’ productive vocabulary size, the lemma offers a better judgment of productive word knowledge because it is restricted to the stem word and its inflected

forms only (Nation & Webb, 2011). This counting stipulation allows for better estimates of word frequency and vocabulary size over using word families that include derived and inflected word forms. As a result, the word family as a counting unit tends to be biased towards higher-frequency word items that have more forms than low-frequency words (2011).

A final reason the present study utilized the CELEX scores was that they are measured on an interval scale. In popular other measures of productive vocabulary size such as Laufer and Nation's (1995) Lexical Frequency Profile, vocabulary size is calculated in terms of the proportion of words used according to frequency bands within a single composition. For the purposes of this study, utilizing a log mean word frequency, such as CELEX that quantifies the average total frequency of all the words within an essay, permitted a more holistic account of each essay's productive lexicon. In other words, the research questions were more concerned with the average vocabulary size of each text in the sample rather than comparing the proportions of words employed from varying frequency bands.

MTLD. The Measure of Textual Lexical Diversity (MTLD) is a new tool available to quantify lexical diversity in written texts (McCarthy & Jarvis, 2010). Scores have an indefinite range; however, validation studies identified a range of interval data scores of 70 to 120 where the higher scores indicate greater diversity (2010). Of all instruments of lexical diversity, the MTLD appears to be the most resistant to text length effects (Crossley et al., 2010; Koizumi, 2012; McCarthy & Jarvis, 2010). Since the MTLD is relatively new, there is a need for studies to utilize this measure of lexical

diversity to provide further validation of its insensitivity to text length effects (McCarthy & Jarvis, 2010). Because the sample corpus included essays that varied greatly in text length, the study needed an instrument that would produce reliable estimations of lexical diversity regardless of token count. Consequently, the MTL D was selected from within the Coh-Metrix tool as an independent variable for the present study.

The MTL D solves the text length problem by sampling strings of words multiple times to determine how many times a text a segment reaches a token-type ratio (TTR) of .72 or below (McCarthy & Jarvis, 2010). At the point where the text fails to score at below .72, the remaining segment then receives a TTR score and that number is added to the number of times the text scored at the cutoff point (known as a “factor” count). For example, if a fifty-word text reaches the factor count twice and then the remaining segment’s TTR is .32, the MTL D performs the following calculation: $2 + .32 = 2.32$. This figure is then divided into the number of tokens, i.e. $50 \div 2.32$. This procedure is then reversed beginning with the last word of the text and ending with the first. The two results from the forwards and backwards procedures are averaged together to result in the MTL D value (Koizumi, 2012; McCarthy & Jarvis, 2010).

In the validation study of the MTL D, McCarthy and Jarvis (2010) compared the MTL D values to those resulting from four other established measures of lexical diversity: the *voc-D*, *HD-D*, *K*, and the *Maas*. The texts spanned 16 registers and varied between 100 and 2,000 words in length. Result indicated that the MTL D performs better than the other measures in terms of controlling for text length (2010). Furthermore, the MTL D performed equally as well or better than the other four more established measures. The

MTLD had the highest correlation with *voc-D* ($r = .84$). An ANOVA revealed that only the MTLD and the *voc-D* accurately predicted the model ($\eta^2 = 36.9$). However, *voc-D* also demonstrated a contribution to the predicted model for shorter texts (2010). As a result, McCarthy and Jarvis (2010) validated the MTLD as a powerful tool for examining lexical diversity on large texts, but also advised that analyses of lexical diversity run the *voc-D* alongside and together with the MTLD to derive a more solid profile of textual lexical diversity (2010).

Voc-D. Per McCarthy and Jarvis' (2010) suggestion to include other measures of lexical diversity, the *voc-D* was also chosen as an independent variable. Similar to the MTLD, the *voc-D* is also interval data and has a variable score range. Previous studies have found that *voc-D* scores can typically range between 10 and 100, with the higher number indicating greater lexical diversity (McCarthy & Jarvis, 2007). As discussed in Chapter Two, the *voc-D* also calculates a lexical diversity score based on randomized sampling of a given text (Malvern & Richards, 1997). It samples 35 tokens from a text 100 times and the resulting TTRs are averaged to form a diversity score to create an empirical curve. This procedure is conducted three times to arrive at the *D*-score for a text (Malvern et al., 2004). Final scores often range from 10 to 100 with the higher scores equating to greater diversity (McCarthy & Jarvis, 2010). The *voc-D* was validated in Durán, Malvern, Richards, and Chipere' (2004) study of 32 children at 10 different ages participating in the Bristol Study of Language Development. Results indicated that *D* accurately correlated lexical diversity to spoken language proficiency ($r = .87$).

However, Duran et al. (2004) and McCarthy and Jarvis (2010) have indicated that *voc-D* is best used to analyze the lexical diversity of short texts.

TOEFL iBT Independent Writing Rubric

To score the essays, the TOEFL iBT Independent Writing Rubric (2005) was selected to reflect an authentic measure of L2 writing proficiency. It is a holistic rubric that produces categorical scores ranging from 0 (the lowest possible score) to 5 (the highest possible score). This computer-based assessment is widely used to judge the preparedness of NNS learners for the academic writing tasks to be encountered in the mainstream college classroom. Numerous colleges and universities use this score as a criterion for admittance. According to the Educational Testing Service (n.d.), more than 8,500 colleges and universities around the world require NNS students to submit TOEFL scores as part of the application for admission procedures. Thus, the TOEFL iBT Rubric was an ideal choice to holistically rate NNS learner essays.

The TOEFL iBT Writing Rubric (2005) is based on Cumming, Kantor, Powers, Santos, and Taylor's (2000) monograph detailing the writing framework for the TOEFL 2000 next generation test. This framework details two overarching characteristics of successful L2 writing. Thus, the objectives of the rubric relate first to learners' production of "discourse and ideas," addressing the organization and presentation of ideas, and second to "language use," the syntactical, morphological, and lexical aspects of the text (p. 14). Cumming et al. (2000) thus deconstructed these two macro-level features of quality writing using empirical studies of the many variables that function

systemically together to form proficiency L2 writing. Over 64 empirical studies formed the evaluative criteria utilized in the TOEFL iBT Independent Writing Rubric.

Despite the TOEFL iBT Independent Writing Rubric being validated for NNS writing proficiency, it was decided to use the same rubric to rate the NS essays as well for the following reasons. First, if the TOEFL iBT Independent Writing Rubric is to accurately assess NNS writers' ability to produce proficient academic texts in English, then NS writers were predicted to score at the highest level of 5. This provides a point of reference for which comparisons of L2 writing and lexical proficiency can be performed. Second, solid research design necessitates the use of the same instruments across groups in order to make direct comparisons (Fraenkel & Wallen, 2009). Lastly, clear and reliable training materials for raters must accompany an assessment rubric in order to ensure the accuracy of its use. ETS has a training set and scoring guide freely available from which to advise raters on its proper use (www.ets.org/toefl). As such, the TOEFL iBT Independent Writing Rubric was selected over other assessment instruments.

Research Questions

The research was guided by the following research questions that have emerged from the review of research and literature:

1. Is there a significant difference between advanced NNS learners' and NS learners' measures of vocabulary size and lexical diversity (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser

et al., 2004; the *voc-D* and MTLD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) as evidenced in academic writing?

2. Is there a relationship between vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) as evidenced in NNS and NS essays?
3. Is vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser, et al., 2004) or lexical diversity (as measured by the *voc-D* and MTLD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) a greater predictor of writing score achievement (as measured by the TOEFL iBT Independent Writing Rubric; ETS, 2005) in non-native and native speaking college writing?

Hypotheses

The null and directional hypotheses for the three research questions were as follows.

Hypothesis for Research Question One

- H_0 : There is no significant difference between the vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-*

D and MTLD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) of advanced NNS learner and NS writing.

- H₁: The vocabulary size (as measured by the CELEX raw frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) of NS compositions will be greater than the same measures in advanced NNS writers' compositions.

The estimation that NS texts would exhibit significantly higher levels of lexical diversity and utilize lower-frequency words was based on previous research findings that compared L1 and L2 lexical profiles and arrived at this very conclusion (Crossley et al., 2010; de Haan & van Esch, 2005; Laufer, 1994; Laufer & Nation, 1995; McNamara, Crossley, and McCarthy, 2009; Schoonen, van Gelderen, Stoel, Hulstijn, & de Glopper, 2010; Silva, 1993).

Hypothesis for Research Question Two

- H₀: There is no significant relationship between vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) as evidenced in NNS and NS essays.
- H₁: Vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) significantly contributes to

lexical diversity (as measured by the *voc-D* and MTLTD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) as evidenced in NNS and NS essays.

Laufer's (1994) finding that the use of lower-frequency words did not correlate significantly to lexical diversity in her study of English as a foreign language (EFL) university students formed the hypothesis behind this question with the express purpose of testing if this result holds true for NS and advanced NNS in an English as a second language (ESL) context. Given the many studies demonstrate that higher-proficiency NNS and NS employ low-frequency words and lexical variation in writing (Crossley et al., 2010; Laufer, 1994; Laufer & Nation, 1995; McNamara et al., 2009; Schoonen et al., 2010), it is hypothesized that there will be a significant relationship between vocabulary size and variation; however, it may not be a large correlation as evidenced in Laufer's (1994) study.

Hypothesis for Research Question Three

- H_0 : Neither vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) nor lexical diversity (as measured by the *voc-D* and MTLTD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) is a significant predictor of writing score (as measured by the TOEFL iBT Independent Writing rubric; ETS, 2005) in non-native and native speaker college writing.
- H_1 : Learners who demonstrate a larger vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) will receive higher writing scores (as measured by the TOEFL iBT

Independent Writing rubric; ETS, 2005) in non-native and native-English speaking college writing.

- H₂: Learners who demonstrate greater lexical variation (as measured by the *voc-D* and MTLN in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) will receive higher writing scores (as measured by the TOEFL iBT Independent Writing rubric; ETS, 2005) in non-native and native-English speaking college writing.

Two directional hypotheses were formulated due to the lack of literature investigating this specific research question on whether vocabulary size or lexical diversity has a greater affect on writing score. Crossley et al. (2010) in a large-scale lexical analysis of over 240 texts within a corpus of NS and NNS essays found that lexical diversity best predicted writing score, closely followed by lexical sophistication and word hypernymy ($p < .001$). Together, these three variables accounted for 44% of the variance in writing score ($r^2 = .44$). However, the study was conducted with a population at varying levels of English proficiency and only randomized chunks of texts were selected for analysis. Therefore, this study proceeded under the hypothesis that lexical diversity would be a greater predictor of writing score for both the more tightly controlled sample populations of NS and NNS learner texts to provide further evidence for Crossley et al.'s (2010) findings.

Data Analysis Procedures

In order to answer the first research question examining the differences between advanced NNS and NS learners' vocabulary size and diversity, a multivariate analysis of variance (MANOVA) was conducted. According to Tabachnik and Fidell (2007), a MANOVA is the ideal statistical analysis when exploring the differences in two or more groups of interval data sorted by a categorical variable. In the case of the present study, there were three variables comprising the analysis: the CELEX log frequency mean of all words scores, the MTLT, and the *voc-D*.

For the second research question investigating how much vocabulary size contributes to lexical diversity, a bivariate Pearson-product moment correlation was conducted. This statistical analysis was appropriate because both variables were interval data. It was hypothesized that there would be a significant correlation based on Laufer's (1994) results that growth in learners' productive vocabulary size had no significant effect on their variation of vocabulary.

To answer the third and chief research question regarding the relationship between vocabulary size and variation and writing score, a binary logistic regression analysis was selected to reveal any contributions indices of vocabulary size and variation have on the dependent variable of writing score. A logistic regression analysis allows for the study of relationships between a dependent variable and multiple independent variables (Shavelson, 1996). The logistic regression analysis is useful when intending to examine (a) the relationship between continuous and categorical variables, (b) to

calculate the probability of a dichotomized outcome, or (c) to test a particular theory (Franken & Wallen, 2009; Tabachnick & Fidell, 2007).

The writing score resulting from the TOEFL iBT Independent Writing Rubric (a categorical score of 0-5) formed the dependent, categorical variable. The independent variables of the vocabulary size scores from the CELEX log frequency mean of all words, the MTL D , and the voc- D were interval data. These variables were hypothesized to contribute significantly to overall writing score as demonstrated by previous studies (Engber, 1995; Laufer & Nation, 1995; Linnarud, 1986).

Ethical Considerations

Two major ethical considerations arise in the execution of this study relating to participant confidentiality and grade effects due to participation. First, the researcher made every effort to protect the participants' and raters' identities by using coded numbers in place of student names, raters' names, and institutional names. Any identifying information was stored via a password-protected data sheet and stored on a computer that also requires a password only knowledgeable to the researcher to access. Secondly, although a score was assigned to the compositions, the resulting score was used for the purposes of the study only and did not affect the learners' actual course grade. Furthermore, learners' grades were not impacted if a participant chose to decline participation in the study.

Conclusion

This chapter described the research design, participants, operational research questions, and the data collection and analysis procedures. It also provided the ethical considerations governing the provisions made to protect participants' rights within the study. Chapter Four will detail the results from the procedures provided above and reveal the answers to the research questions posed within the present chapter. Following the results, Chapter Five will discuss the implications for research, pedagogy, and future avenues for study.

CHAPTER 4: RESULTS

This chapter presents the findings of the present study investigating the extent to which vocabulary size or lexical variation contributes to writing score. The chapter revisits the research questions, the associated hypotheses, and research design originally presented in Chapter Three. The descriptive statistics of the sample are then presented. The chapter proceeds to describe the data screening and normality checks conducted prior to data analysis. The final section contains the results from a series of MANOVAs, Pearson product moment correlations, and logistic regression analyses appropriate to explore each research question.

Research Questions

The study was designed to examine the relationship of vocabulary size and lexical variation to writing achievement in advanced non-native speaking (NNS) and native speaking (NS) college students. To achieve this aim, three main research questions guided the investigation. These questions are presented below along with their corresponding hypotheses.

1. Is there a significant difference between advanced NNS learners' and NS learners' measures of vocabulary size and lexical diversity (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004; the *voc-D* and MTL*D* in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) as evidenced in academic writing?

It was hypothesized that NS compositions would demonstrate greater vocabulary size and lexical diversity than the same measures in advanced NNS compositions as demonstrated by earlier studies that revealed marked differences in NS and NNS learners' lexical resource (Crossley, Salsbury, McNamara, & Jarvis, 2010; de Haan & van Esch, 2005; Laufer, 1994; Laufer & Nation, 1995; McNamara, Crossley, and McCarthy, 2009; Schoonen, van Gelderen, Stoel, Hulstijn, & de Glopper, 2010; Silva, 1993).

2. Is there a relationship between vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLTD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) as evidenced in NNS and NS essays?

It was posited that vocabulary size would correlate significantly to lexical diversity in both NS and NNS essay groups. Evidence from the literature has suggested that a larger lexicon assists writers to vary words in production (Laufer & Nation, 1995).

3. Is vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser, et al., 2004) or lexical diversity (as measured by the *voc-D* and MTLTD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) a greater predictor of writing score achievement (as measured by the TOEFL iBT Independent Writing Rubric; ETS, 2005) in non-native and native speaking college writing?

Based on evidence from Crossley, Salsbury, McNamara, and Jarvis' (2010) findings, it was hypothesized that NNS and NS essays displaying greater lexical diversity would achieve higher writing scores on college level compositions.

To answer these research questions and test the directional hypotheses, the study followed a quantitative research design. The statistical software package SPSS 21.0 was used to perform analyses on a corpus of authentic writing assignments gathered from NNS and NS learners in composition courses designed to prepare them for college writing tasks.

Sampling Procedures

The data collection process occurred from February 2013 through March 2013. The sample essays that comprised the corpus in this study were assembled from two groups of college students. The first group of essays originated from NNS learners ($n = 112$) classified to be at advanced levels of English language proficiency and enrolled in English as a second language composition courses at six intensive English programs (IEPs) across the United States. The second set of essays was collected from NS learners enrolled in four sections of a first-year composition course at a large public university in the southeastern United States ($n = 71$; Table 12).

The decision to include both groups of college writers within the data served a dual purpose. First, the examination of native speakers' vocabulary size and variation within a typical first-year composition assignment served to create a general threshold for productive lexical proficiency within the mainstream college classroom. Subsequently,

the study of the same lexical measures produced by NNS enrolled in the highest levels of second language (L2) composition courses allowed for the comparison of the NNS preparedness for the lexical demands of college writing. The inclusion of both groups served to add to the body of literature to determine (a) to what extent does the variation of words or use of lower-frequency terms in composition contributes to writing proficiency, and (b) how do advanced NNS' lexical diversity and vocabulary size profiles compare to those of the novice NS writer (Laufer, 1994).

Accordingly, data collection yielded a corpus of $N = 183$ academic essays for lexical analysis in the present study. The sample corpus was reduced to $N = 172$ viable essays for analysis after removing eight NNS compositions that did not meet the minimum text length requirement of 100 tokens needed for meaningful interpretation by study instruments (Koizumi, 2012.) and three NS essays that exceeded the Coh-Metrix's 15,000-character maximum limit for data entry (Graesser, McNamara, Louwense, & Cai, 2004). The resulting sample size of 172 was acceptable for finding a medium effect size ($\alpha = .99$) at the $p < .05$ level (Cohen, 1992).

Table 12

Number of Essays by Language Designation

	Frequency	Percent in Sample
Non-native speakers	104	60.5
Native speakers	68	39.5
Total	172	100.0

Essays' demographic data relating to first languages (L1s) and writing genre was also collected. The sample corpus thus represented 14 different L1s (Table 13). English was the largest L1 group in the sample with 68 (40%) L1 English essays collected. Of

the 104 essays composed by NNS group, the Spanish ($n = 45$, 26%) and Arabic ($n = 36$, 21%) L1 groups comprised the majority of the NNS compositions followed by Mandarin ($n = 7$, 4%), Korean ($n = 4$, 2%), Azerbaijani ($n = 2$, 1%), Portuguese ($n = 2$, 1%), and Tagalog ($n = 2$, 1%). The remaining L1 groups of Japanese, Taiwanese, Thai, Turkish, Ukrainian, and Vietnamese were represented by only one essay (1%) each within the sample corpus.

Table 13

Number of Essays by First Language

L1	Frequency	Percent of Sample
Arabic	36	20.9
Azerbaijani	2	1.2
Japanese	1	.6
Korean	4	2.3
Mandarin	7	4.1
Portuguese	2	1.2
Spanish	45	26.2
Tagalog	2	1.2
Taiwanese	1	.6
Thai	1	.6
Turkish	1	.6
Ukrainian	1	.6
Vietnamese	1	.6
English	68	39.5
Total	172	100.0

The sample corpus also spanned seven different academic writing genres (Table 14). The largest genre group within the sample consisted of the analysis essays composed by the NS learners, yielding 68 (40%) essays. Persuasive essays ($n = 31$, 18%) and compare and contrast texts ($n = 27$, 16%) were the largest categories of genres within the NNS sub-corpus. Narratives ($n=17$, 10%), summaries ($n = 14$, 8%), process texts ($n = 9$, 5%), and cause and effect essays ($n = 6$, 4%) round out the data, each with less than 20 compositions within the sample.

Table 14

Number of Essays by Genre

Genre	Frequency	Percent of Sample
Analysis	68	39.5
Cause and Effect	6	3.5
Compare and Contrast	27	15.7
Narrative	17	9.9
Persuasive	31	18.0
Process	9	5.2
Summary	14	8.1
Total	172	100.0

Two independent raters trained to use the TOEFL Independent Writing Rubric then holistically scored the compositions on the instrument's scale of 0 to 5. Inter-rater reliability was checked using a Pearson's product moment correlation (Table 15). Scores from both raters were highly correlated ($r = .79, n = 172, p < .001$) indicating good inter-rater reliability. In the few cases where the two raters' scores differed, a third rater evaluated the essays and provided a reconciled final score.

Table 15

Inter-Rater Reliability Analysis

	Writing Score Rater 1	Writing Score Rater 2
Writing Score Rater 1	--	.79*
Writing Score Rater 2	.79*	--

*Note. $p < 0.001$ level (2-tailed).

Descriptive Data Results

Description of the Sample

Descriptive data and measures of central tendency indicated that the mean text length of the essays was 870.23 words ($SD = 669.91$; range, 102-2530; Table 16). The mean text length of NNS essays was 407.18 words ($s = 199.47$; range, 102-1003). For NS compositions, the mean text length was 1578.43 words ($s = 492.32$, range, 501-2530).

Table 16

Text Length

	<i>N</i>	Range	Minimum	Maximum	<i>M</i>	<i>SD</i>
Total Corpus	172	2428	102	2530	870.23	669.91
Non-native	104	901	102	1003	407.18	199.47
Native	68	2029	501	2530	1578.43	492.32

The three independent variables comprising the data were scores relating to two measures of lexical diversity and one measure of vocabulary size indexed within the Coh-Metrix computational linguistics program (Graesser et al., 2004). The MTL D and *voc-D* quantified the sample essays' lexical diversity while the CELEX mean log frequency score for all words produced the average vocabulary size of the sample (Table 17). For the MTL D and *voc-D*, higher scores indicate greater lexical diversity. In contrast, a lower CELEX frequency score shows the use of lower-frequency word items, signaling indications of a greater productive lexicon. Therefore the lower the CELEX score, the larger the average vocabulary size of the text. The mean MTL D score for the whole sample was 73.01 (*SD* = 14.98; range, 33.55-117.49). For *voc-D*, the mean score was 85.38 (*SD* = 18.08; range, 30.88-126.68). The mean CELEX score for the sample was 3.07 (*SD* = .09; range, 2.78-3.36).

Table 17

Lexical Diversity and Vocabulary Size

	<i>N</i>	Range	Minimum	Maximum	<i>M</i>	<i>SD</i>
MTL D	172	83.94	33.55	117.49	73.01	14.98
NNS	104	71.87	33.55	105.42	69.12	15.63
NS	68	67.45	50.04	117.49	78.96	11.74
VOC D	172	95.81	30.88	126.68	85.38	18.08
NNS	104	95.81	30.88	126.68	78.17	17.85
NS	68	55.06	66.15	121.21	96.40	11.86
CELEX	172	.57	2.79	3.36	3.07	.09
NNS	104	.57	2.79	3.36	3.09	.10
NS	68	.34	2.86	3.20	3.04	.07

The dependent variable was derived from the raters' holistic assessment of academic English writing preparedness and proficiency using the TOEFL Independent Writing Rubric (ETS, 2008). The mean holistic writing score for all compositions within the sample was 4.04 ($SD = 1.01$; range, 2-5; Table 18). The mean score of the NNS essays was 3.42 ($SD = .83$; range 2-5). NS texts achieved a mean score of 4.99 ($SD = .12$; range, 4-5).

Table 18

Descriptive Statistics for Writing Score

	<i>N</i>	Range	Minimum	Maximum	<i>M</i>	<i>SD</i>
Total Corpus	172	3	2	5	4.04	1.01
Non-native	104	3	2	5	3.42	.83
Native	68	1	4	5	4.99	.12

Since the inclusion of a NS group of essays was to form a baseline group, the NS texts were expected to score near perfectly on the writing score scale designed to measure NNS learners' English writing preparedness. The data confirmed this hypothesis with 67 of the 68 NS texts earning a 5, the score deemed by the TOEFL iBT Independent Writing Rubric to be the highest indicator of academic English writing proficiency (ETS, 2005). Only one essay in the NS group earned a score of 4. Hence, the variation in writing score was found within the 104 NNS compositions where eight essays scored a 5, forty-three earned a score of 4, thirty-eight scored at a level 3, and fifteen essays were rated at a level 2 (Table X). None of the essays garnered scores of 0 or 1. This is consistent with the literature demonstrating that there are marked differences between NS and NNS learners' writing quality (Silva, 1993).

Table 19

Number of Essays by Writing Score

Score	Frequency	Percent in Sample
2	15	8.7
3	38	22.1
4	44	25.6
5	75	43.6
Total	172	100.0

Initial Data Screening of the Independent Variables

Before the data were analyzed, the data set was examined to gauge the fit between the distribution of the variables and the assumptions of the selected statistical analyses, including skewness, kurtosis, and normality checks. The continuous independent variables of the MTLD, *voc-D*, and CELEX frequency scores were run separately from the categorical dependent variable of writing score.

First, an outlier analysis for the independent variables was conducted using trimmed means. The trimmed mean is a robust method that removes the top 5% highest and lowest values and recalculates the mean to check the symmetry of data distribution (Field, 2009). In other words, if there are significant outliers present, the mean changes considerably (2009). Thus, the means of the MTLD, *voc-D*, and CELEX scores were compared with their 5% trimmed means (Table 20). Resulting means did not vary significantly and therefore indicated there were no significant outliers in the data.

Table 20

Data Screening for the MTLD, voc-D, and CELEX Frequency Scores

	<i>M</i>	Trimmed <i>M</i>	Skewness	Kurtosis	95% CI	
					LL	UL
MTLD	73.09	72.95	.08	.17	70.75	75.26
VOCD	85.38	85.83	-.43	-.07	82.66	88.10
CELEX	3.07	3.07	-.16	.62	3.06	3.09

Next, the skewness and kurtosis of the independent variables were checked. To examine if skewness was significant, the standard error of skewness was doubled (i.e., $.185 \times 2 = .37$) creating a range of $-.37$ to $.37$. Skewness for the MTL D and CELEX frequency scores were both non-significant. However, *voc-D*, was slightly negatively skewed, falling just outside the range. Therefore, boxplots were generated and revealed the presence of two mild outliers in the *voc-D* data. The same process was applied to kurtosis (i.e., $.37 \times 2 = .74$) resulting in a range of -0.74 to 0.74 . All three variables' kurtosis statistics were non-significant.

K-S tests of normality also revealed that the *voc-D* ($D_{172} = .07, p < .05$) is significantly non-normal (Table 21). However, in samples larger than 50 the K-S test results must be interpreted with caution as the statistic can be significant even when the scores slightly deviate from a normal distribution (Field, 2009). Therefore, the boxplots, histograms, and Q-Q plots were consulted.

Table 21

K-S Tests of Normality for the MTL D, Voc-D, and CELEX

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	<i>df</i>	<i>p</i>	Statistic	<i>df</i>	<i>p</i>
MTLD	.03	172	.20	.99	172	.84
VOCD	.07	172	.03*	.98	172	.04*
CELEX	.04	172	.20	.99	172	.36

Note. * $p < .05$ level (2-tailed)

Initial Data Screening of the Dependent Variable

An inspection of the distribution statistics of writing score revealed a significant deviation from a normal curve. The non-normality of the writing score variable was expected due to the large number of essays earning a score of five on the holistic rating

scale. Therefore, the skewness (-.61), kurtosis (-.87) fall out of the acceptable range confirming that the variable is negatively skewed with a kurtosis that is slightly leptokurtic (Table 22).

Table 22

Data Screening for Writing Score

	<i>M</i>	Trimmed <i>M</i>	Skewness	Kurtosis	95% CI	
					LL	UL
Writing Score	4.04	4.10	-.61	-.87	3.89	4.19

K-S tests ($D_{172}=.27, p<.05$) were also significant, corroborating the non-normal distribution findings (Table 23). The clustering of scores at the high-end of the writing score scale aligns with NNS learners' advanced English proficiency and NS learners' fluent proficiency. Because there is the potential to violate the assumption of homogeneity of variance, a MANOVA was run (see research question one) to examine the differences between variables.

Table 23

K-S Tests of Normality for Writing Score

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	<i>df</i>	<i>p</i>	Statistic	<i>df</i>	<i>p</i>
Writing Score	.27	172	.000	.81	172	.000

Note. * $p < .001$ level (2-tailed)

Data Analyses for Research Questions and Hypotheses

Initial data screening revealed that the independent variables of the MTLT and CELEX frequency scores did not meaningfully violate assumptions for normality. The non-normal distribution of dependent variable of writing score was expected based on the

demographics of the sample. The current section thus details the statistical analyses of these variables to answer the research questions presented in earlier chapters.

Final Data Screening

A binary logistic regression analysis is based on a few assumptions that ensure the robustness of results. The data were screened and each assumption tested. The first assumption of a logistic regression presumes that there is a significant linear relationship between the continuous predictor variables and the log-odds of the categorical outcome variable (Field, 2009; Pallant, 2011). Therefore, Pillai's Trace within the MANOVAs was consulted as a robust validation of the violation of homoscedasticity (Tabachnik & Fidell, 2007).

Secondly, a logistic regression analysis assumes that none of the cases within a sample are related (Field, 2009; Tabachnik & Fidell, 2007). In other words, the same participants should not be examined at different points in time. A violation of this independence assumption can result in over dispersion of the data which can lead to the regression producing more variance than can accurately be predicted by the model (Field, 2009). The cases in this study were independent of each other, with only one essay collected per participant and included within the sample corpus. As such there was no violation of the independence assumption within the data.

Lastly, multicollinearity of the predictor variables is a final assumption to avoid violating in a logistic regression analysis (Field, 2009; Tabachnik & Fidell, 2007). Multicollinearity occurs when two or more of the predictor variables are highly correlated. Multicollinearity was tested by looking at the correlations between the

independent predictor variables (Table 24). Correlations at $r = .75$ or higher signify the presence of multicollinearity (Pallant, 2011). Although all variables are significantly related at the $p < .05$ level, examination of the correlations between the predictor variables of the MTL D and CELEX word frequency ($r = .44, n = 172, p < .001$) reveals that the correlations do not breach the 0.75 threshold for multicollinearity (Table 24). However, the MTL D and the *voc-D* are highly correlated at $.81 (n = 172, p < .001)$ given that these two variables both measure lexical diversity. However, this was not surprising based on previous studies' recommendations to include both measures when estimating lexical diversity in samples with great variation in text length (Jarvis & McCarthy, 2010). As a result, the *voc-D* remained a variable for analysis of research questions one and two, but the *voc-D* was removed from the logistic regression equation in research question three in order to avoid violating the assumption of multicollinearity.

Table 24

Correlations between the MTL D, Voc-D, and CELEX Frequency Scores

	MTLD	VOCD	CELEX
MTLD	--	.81*	-.44*
VOCD	.81*	--	-.47*
CELEX	-.44*	-.47*	--

Note. * $p < 0.01$ level (2-tailed)

Final data screening thus indicated that it was appropriate to continue with the binary logistic regression for the third research question.

Research Question One

To address the first research question, a multivariate analysis of variance (MANOVA) was performed to examine mean differences between native and non-native

speakers' profiles of lexical diversity and vocabulary size. The grouping variable was language designation (NNS = 1, NS = 2) and the MTL D, voc-*D*, and CELEX frequency scores comprised the dependent variables. Results demonstrated significant differences in the two measures of lexical diversity and the one measure of vocabulary size based on language designation ($F_{3, 168} = 20.30, p < .05, \eta^2 = .27$). Because there is a violation of homogeneity of variance, Pillai's Trace (.27), was used to interpret effect size (Tabachinik & Fidell, 2007). Thus, language designation accounts for roughly 27% of the combined variance in lexical diversity and vocabulary size (Table 25).

Table 25

Multivariate Analysis of Variance between of Lexical Diversity and Vocabulary Size by Language Designation

Effect		Value	<i>F</i>	Hypothesis <i>df</i>	Error <i>df</i>	<i>p</i>	η^2
Language Designation	Pillai's Trace	.27	20.30	3	168	.000	.27
	Wilks' Lambda	.73	20.30	3	168	.000	.27
	Hotelling's Trace	.36	20.30	3	168	.000	.27
	Roy's Largest Root	.36	20.30	3	168	.000	.27

When the MTL D, voc-*D*, and CELEX frequency scores are considered separately

within the MANOVA, all three dependent variables are significantly different by language designation (Table 26). Language designation accounted for about 7% of the variance in CELEX frequency score ($F_{3, 168} = 12.55, p < .05, \eta^2 = .07$), 10% in MTL D variance ($F_{3, 168} = 19.66, p < .05, \eta^2 = .10$), and 25% of the variance in voc-*D* score ($F_{3, 168} = 55.02, p < .05, \eta^2 = .25$).

Table 26

Tests of Between-Subjects Effects for Lexical Diversity and Vocabulary Size by Language Designation

Source	Dependent Variable	Type III SS	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2
Language Designation	MTL D	3976.83	1	3976.83	19.66	.000	.10
	VOCD	13673.29	1	13673.29	55.02	.000	.25
	CELEX	.11	1	.11	12.55	.001	.07

In sum, NS displayed greater lexical diversity in composition as measured by both the MTL D ($M = 78.97, s = 11.74$) and the *voc-D* ($M = 96.40, s = 11.86$) than NNS (MTLD; $M = 69.12, s = 15.63$; *voc-D*; $M = 78.17, s = 17.85$). In terms of vocabulary size, NS also used lower-frequency words ($M = 3.04, s = .07$) in production than their NNS counterparts ($M = 3.09, s = .07$; Table 27).

Table 27

Means for Language Designation

Dependent Variable	Language Designation	<i>M</i>	<i>SD</i>	<i>SE</i>	95% CI	
					LL	UL
MTLD	NNS	69.12	15.63	1.40	66.37	71.87
	NS	78.97	11.74	1.73	75.55	82.36
VOCD	NNS	78.17	17.85	1.55	75.12	81.22
	NS	96.40	11.86	1.91	92.63	100.18
CELEX	NNS	3.09	.10	.01	3.08	3.11
	NS	3.04	.07	.01	3.02	3.07

As a result of significant differences between NS and NNS' data, the data set was split and a second MANOVA run to examine the within-group differences in lexical diversity and vocabulary size on their respective essays (Table 28). Inspection of the individual variables revealed that NNS writers varied significantly from each other in terms of lexical diversity and vocabulary. CELEX frequency scores ($F_{3, 100} = 4.52, p < .01, \eta^2 = .12$) accounted for the least amount of variance. The MTL D ($F_{3, 100} = 13.24, p < .01, \eta^2 = .28$) accounted for 28% of the variance. The *voc-D* ($F_{3, 100} = 15.24, p < .01, \eta^2 = .31$) explained the most variance between NNS texts. However, only the MTL D ($F_{1, 66} = 4.17, p < .05, \eta^2 = .06$) was significantly different for NS texts, but it only accounted for 6% of the variance.

Table 28

Split-File Analysis of Between-Subjects Effects for Lexical Diversity and Vocabulary Size by Language Designation

Source	Dependent Variable	Type III SS	df	MS	F	p	η^2
NNS	MTLD	7154.18	3	2384.73	13.24	.000	.28
	VOCD	10300.03	3	3433.34	15.24	.000	.31
	CELEX	.13	3	.04	4.52	.005	.12
NS	MTLD	548.19	1	548.19	4.17	.045	.06
	VOCD	239.57	1	239.57	1.72	.194	.03
	CELEX	4.002E-005	1	4.002E-005	.01	.929	.00

Research Question Two

For the second research question, a Pearson's product moment analysis revealed a moderate negative correlation to both the MTLD ($r = -.44, p < .001$) and the voc-D ($r = -.46, p < .001$) indicating that essays with greater lexical diversity utilizes lower-frequency words to a degree.

Table 29

Correlations between the MTLD, Voc-D, and CELEX Frequency Scores

	MTLD	VOCD	CELEX
MTLD	--	.81*	-.44*
VOCD	.81*	--	-.47*
CELEX	-.44*	-.47*	--

Note. * $p < 0.001$ level (2-tailed)

Research Question Three

A binary logistic regression analysis was performed to investigate research question three. To prepare the dependent variable for binary logistic regression, the TOEFL writing scores were dichotomized, splitting the scores into proficient (score of 5) and non-proficient (scores 0-4). The decision to set the cutoff at level 5 was based on the threshold established by the NS group of essays ($M = 4.99, s = .12$, range, 4-5).

Furthermore, the Educational Testing Service identifies a score of 5 to be an indicator of

“good” academic English (2005). To validate this cutoff score, a one-way analysis of variance (ANOVA) was conducted. There was a statistically significant difference ($F_{1, 170} = 235.72, p < .001$) in writing score based on language designation; therefore, the dichotomy appeared appropriate.

The predictors of the MTLTD and CELEX frequency scores were entered into Block 1 and the dichotomized TOEFL writing score was placed into the dependent variable field. The *voc-D* was removed due to a large correlation to the MTLTD and to avoid violating the assumption of multicollinearity ($r = .81, n = 172$). The full model containing the MTLTD and CELEX scores was statistically significant, $\chi^2(4, n = 172) = 38.21, p < .001$, suggesting that the model was able to discriminate between essays deemed proficient and non-proficient in terms of writing score. The model overall described between 19.9% (Cox & Snell R^2) and 26.7% (Nagelkerke R^2) of the variance in writing score. The model accurately classified 71.5% of the cases. Results showed that the MTLTD significantly contributed to the model, reporting an odds ratio of 1.07 ($\text{Exp}[B] = 1.07, p < .001$; Table 30). As such, for 1.07 unit increase in MTLTD score, the log odds of scoring a five increases by 0.07. CELEX frequency scores did not contribute significantly to the model ($\text{Exp}[B] = .03, p > .05$).

Table 30

Logistic Regression Predicting Likelihood of Scoring a 5

		B	SE	Wald	df	p	Exp(B)	95% CI for Exp(B)	
								LL	UL
Step 1	MTLTD	.07	.02	19.73	1	.000	1.07	1.04	1.10
	CELEX	-3.41	2.065	2.719	1	.099	.03	.00	1.90
	Constant	5.36	6.67	.65	1	.42	213.66	--	--

A recommended validation check of the results of a logistic regression is to conduct a split-file analysis (Field, 2009). For this second binary logistic regression, the sample was split by language designation. For NNS, the full model with all predictors was significant, $\chi^2(2, n = 104) = 14.55, p < .01$, indicating that the model distinguished essays' ratings by language designation. The full model was also significant for NS, $\chi^2(2, n = 104) = 10.42, p < .01$. This split-file model explained between 13.1% (Cox & Snell R^2) and 31.2% (Nagelkerke R^2) of writing score variance for NNS and correctly predicted 92.3% of NNS cases; for NS, the model accurately predicted 100% of the cases due to 67 of the 68 essays earning a score of 5.

Again, only the MTLTD significantly contributed to the model for NNS essays ($\text{Exp}[B] = 1.11, p < .001$; Table 31). As such, for 1.11 unit increase in MTLTD score, the log odds of a NNS text scoring a 5 increases by .11. In this analysis, CELEX scores were non-significant for NNS compositions. The model was unable to predict significant relationship between the MTLTD and the CELEX frequency scores to writing score for NS due to the lack of variation in the dependent variable.

Table 31

Logistic Regression Predicting NNS Likelihood of Scoring a 5

		B	SE	Wald	df	p	Exp(B)	95% CI for Exp(B)		
									LL	UL
NNS	Step 1	MTLD	.11	.04	8.73	1	.003	1.11	1.04	1.19
		CELEX	1.61	4.69	.12	1	.732	5.004	.00	49561.71
		Constant	-15.75	16.04	.96	1	.326	.00	--	--
NS	Step 1	MTLD	15.11	482.92	.001	1	.975	3631875.48	.00	--
		CELEX	1241.42	40279.86	.001	1	.975		.00	--
		Constant	-4640.84	150112.65	.001	1	.975	.000	--	--

A third MANOVA was performed to further validate the regression results and analyze mean differences in lexical diversity as measured by the MTLT only and the CELEX frequency measure by writing score. Results demonstrated a significant difference in the combined measures of lexical diversity and vocabulary size based on writing score, ($F_{6, 336} = 10.61, p < .001, \eta^2 = .16$), explaining about 16% of the variance in score (Table 32).

Table 32

Multivariate Analysis of Variance of Lexical Diversity and Vocabulary Size by Writing Score

Effect		Value	<i>F</i>	Hypothesis <i>df</i>	Error <i>df</i>	<i>p</i>	η^2
Writing Score	Pillai's Trace	.32	10.61	6	336	.000	.16
	Wilks' Lambda	.55	12.60	6	334	.000	.17
	Hotelling's Trace	.81	14.84	6	332	.000	.19
	Roy's Largest Root	.79	44.42	3	168	.000	.31

Tests of between-subjects effects revealed that both the MTLT ($F_{3, 168} = 21.66, p < .001, \eta^2 = .28$) and the CELEX frequency scores ($F_{3, 168} = 10.20, p < .001, \eta^2 = .15$) were significantly different at each writing score level. The MTLT explained 28% of the variation in writing score while the CELEX accounted for 15% only (Table 33).

Table 33

Tests of Between-Subjects Effects for Lexical Diversity and Vocabulary Size by Writing Score

Source	Dependent Variable	Type III SS	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2
Writing Score	MTLT	10702.21	3	3567.40	21.66	.000	.28
	CELEX	.24	3	.08	10.20	.000	.15

Subsequently, a Bonferroni adjustment was conducted to examine the individual mean differences between each writing score level. Results indicated that for the MTLT, level 3 ($M = 64.78, s = 13.45$), level 4 ($M = 73.28, s = 12.60$), and level 5 ($M = 80.33, s$

= 12.02) significantly differed from each other ($p < .05$). However, score level 2 ($M = 56.46, s = 15.73$) and level 3 were not significantly different from each other in terms of the MTL D. In other words, texts earning higher writing scores exhibited greater lexical diversity.

For the CELEX frequency score, there were no significant differences between writing score level 4 ($M = 3.07, s = .09$) and level 5 ($M = 3.04, s = .07$). Differences were also not significant between score level 3 ($M = 3.10, s = .11$) and level 4. Level 5 was significantly different from both levels 2 ($M = 3.10, s = .09$) and 3. Therefore, essays earning ratings of 4 and 5 used lower-frequency vocabulary words (Table 34).

Table 34

Means for Writing Score Levels

Dependent Variable	Writing Score	<i>M</i>	<i>SD</i>	<i>SE</i>	95% CI	
					LL	UL
MTLD	2	56.46	15.73	3.31	49.92	63.00
	3	64.78	13.45	2.08	60.67	68.89
	4	73.28	12.60	1.94	69.46	77.10
	5	80.33	12.02	1.48	77.40	83.25
CELEX	2	3.17	.09	.023	3.13	3.22
	3	3.10	.11	.014	3.07	3.13
	4	3.07	.09	.013	3.05	3.10
	5	3.04	.07	.010	3.02	3.08

Additional Analyses

To look at the MTL D and CELEX frequency scores in terms of their relationship to each individual score level, the different writing score levels were dummy-coded and Pearson correlations produced (Table 35). Results indicate that the MTL D correlated the highest with each writing score except for level 4 (Table 35). Essays receiving a score of 5 had the highest positive correlation to lexical diversity as measured by the MTL D (r

=.43, $n = 172$, $p < .001$). Texts earning scores of 2 and 3 negatively correlated to the MTLT, indicating that essays displaying less word variation tended to rate lower.

CELEX frequency scores had significant correlations with essays rated at level 2 ($r = .32$, $n = 172$, $p < .001$) and level 5 ($r = -.28$, $n = 172$, $p < .001$) although the correlations were lower than the MTLT. The positive correlation of word frequency score to level 2-rated essays indicated that these texts used higher-frequency words as compared to level 5 essays. CELEX frequency scores did not significantly correlate with essays rated at levels 3 and 4.

Table 35

Correlations between Lexical Diversity and Vocabulary Size with Each Writing Score Level

	MTLD	CELEX	Score 2	Score 3	Score 4	Score 5
MTLD	--	-.44**	-.34**	-.29**	.011	.43**
CELEX	-.44**	--	.32**	.13	-.01	-.28**
Score 2	-.34**	.32**	--	-.17*	-.18*	-.27**
Score 3	-.29**	.13	-.17*	--	-.31**	-.47**
Score 4	.01	-.01	-.18*	-.31**	--	-.52**
Score 5	.43**	-.28**	-.27**	-.47**	-.52**	--

Note. * $p < 0.05$ level (2-tailed); ** $p < .01$

Split-file Pearson correlations were run next to validate the directionality of the whole group correlations within the NNS and NS sub-corpora. For NNS texts, the MTLT again showed stronger correlations at each score level than the CELEX frequency scores (Table 36). The CELEX scores only correlated significantly for essays rated at a level 2 ($r = .32$, $n = 104$, $p < .01$) indicating that the use of higher-frequency words tended to net lower writing scores.

The split-file analysis also demonstrated that for NNS essays, only the MTLT ($r = .24$, $n = 68$, $p < .05$) showed a significant relationship to writing score. The CELEX

frequency scores did not significantly impact ratings on NS compositions. This offers further support for the outcome of the logistic regression.

Table 36

Correlations between Lexical Diversity and Vocabulary Size with Each Writing Score Level by Language Designation

		MTLD	CELEX	Score 2	Score 3	Score 4	Score 5
NNS	MTLD	--	-.40**	-.33**	-.21*	.25*	.37**
	CELEX	-.40**	--	.32**	.02	-.18	-.13
	Score 2	-.33*	.32*	--	-.31**	-.35**	-.12
	Score 3	-.21*	.02	-.31*	--	-.64**	-.22*
	Score 4	.25*	-.18	-.35**	-.64**	--	-.24*
	Score 5	.37**	-.13	-.12	-.22*	-.24*	--
NS	MTLD	1	-.36**	--	--	-.24*	.24*
	CELEX	-.36**	1	--	--	.01	-.01
	Score 4	-.244*	.011	--	--	--	-1.000**
	Score 5	.244*	-.011	--	--	-1.000**	--

Note. * $p < 0.05$ level (2-tailed); ** $p < .01$

Conclusion

In sum, Chapter Four presented the results of the statistical analyses that examined the relationship of measures of lexical diversity and size to academic English writing proficiency as evidenced in compositions by native and non-native speakers. Two indexes of lexical diversity and one measure of vocabulary size were used to investigate the principal research question of “To what extent does lexical diversity and size contribute to advanced NS and NNS academic writing proficiency?” A series of MANOVAs revealed that there are significant differences between NS and NNS texts in terms of lexical diversity and vocabulary size. NS compositions tended to display greater variation of lexis and utilized lower-frequency words to convey meaning. Additionally, logistic regression analyses indicated that texts exhibiting greater lexical diversity are more likely to earn higher writing scores overall. Vocabulary size did not appear to have

a significant effect on writing score within the regression model; however, a moderate Pearson-product coefficient ($r = .44$, $n = 172$) suggests that lexical diversity benefits somewhat from the use of lower-frequency vocabulary words.

Chapter Five will interpret the results from the statistical analyses as well as discuss the limitations of this study. The chapter will then consider the implications for advanced non-native speakers and the field of second language writing. It will conclude with recommendations for further research.

CHAPTER 5: DISCUSSION

This chapter summarizes the present study examining the relationship between two measures of lexical richness, vocabulary variation and size, with native speakers and advanced non-native speakers' English academic writing proficiency. The major findings of the study are synthesized and limitations resulting from the research design are discussed. The chapter closes with implications for pedagogy and recommendations for future research.

Purpose of the Study

Many advanced non-native writers face a number of challenges when it comes to writing in a second language (L2) in the post-secondary classroom. Subsequently, concerns arise when non-native English speaking (NNS) learners are assessed according to and alongside native speaking (NS) peers who have the advantage of composing within their first language (L1). Studies have established that L2 academic writing is a time-consuming and demanding process that produces marked and unique linguistic differences from L1 compositions (Burke & Wyett-Smith, 1996; Casanave & Hubbard, 1992; Cumming, 2001; Knoblauch & Matsuda, 2008; Silva, 1993). Furthermore, many college instructors are not trained in methods of L2 writing instruction and assessment, a concern which often results in NNS texts receiving lower grades and referrals to the university writing center (Ferris, 2007; 2009; Ferris, Brown, Liu, & Arnaudo Stine, 2011; Huang & Foote, 2010). Empirical evidence has revealed that college instructors find

lexical concerns to be the most present and most unforgivable in NNS texts, often correlating to significantly poorer writing scores (c.f., Dordick, 1996; Engber, 1995; Santos, 1988). Hence, there is an assumption that NNS learners' lexical matters such as poor word choice and variation stem from their significantly smaller lexicons than those of NS (Agustín-Llach & Gallego, 2009; Engber, 1995; Laufer & Nation, 1995). However, the specific contribution that vocabulary size lends to writing quality in NS compositions and how advanced NNS text compare has not been adequately researched.

As a result, the purpose of the current study was to investigate to what extent vocabulary size and lexical diversity contribute to NNS and NS learners academic writing proficiency. The research design sought to examine three empirical objectives. The first aim was to compare the differences between NS and NNS academic essays' lexical diversity and vocabulary size. The second objective tested the relationship between vocabulary size and its influence on lexical diversity. The final and main objective of the study was to analyze whether vocabulary size or lexical diversity increases essays' likelihood to earn higher scores on a measure of English writing proficiency.

To achieve these goals, the study utilized four instruments. Lexical diversity was assessed through the Measure of Textual Lexical Diversity (MTLD; McCarthy & Jarvis, 2010) and the *voc-D* (Malvern & Richards, 1997). Vocabulary size was obtained using CELEX word frequency scores (Baayen, Piepenbrock, & Gulikers, 1995). All three lexical indices were available through the Coh-Metrix, an online computational linguistics tool. Finally, the Test of English as a Foreign Language Internet-Based

(TOEFL iBT) Independent Writing Rubric served to guide evaluation of the sample essays.

Summary of the Findings

A sample corpus of 172 authentic advanced NNS and NS essays from post-secondary writing courses was assembled during the Spring 2013 academic semester. The NNS texts ($n = 104$) were gathered from the most advanced L2 writing courses at six different Intensive English Programs (IEPs) in the United States. The NS essays ($n = 68$) were collected from four sections of a first-year composition course at a large public university in the southeastern United States. The essays represented 14 different first languages (see Table 13) and covered seven academic writing genres (see Table 14). Two independent raters (one ESL instructor and one first-year composition instructor) scored the essays on a scale 0 to 5. A Pearson's product moment correlation demonstrated a good inter-rater reliability of .79 ($n = 172, p < .001$). In the few instances where the two raters' scores differed, a third rater scored the conflicting essays and provided a reconciled score. There were no instances where all three raters differed in their scores. Essays within the sample averaged a score of 4 ($M = 4.04, SD = 1.01$; range, 2-5), and no essays earned scores of 0 or 1. Texts by the NS writers scored near perfectly on the scale ($M = 4.99, SD = .12$; range, 4-5) whereas essays by advanced NNS writers averaged a score of 3.42 ($SD = .83$; range 2-5).

After rating, minor errors such as spelling, repeated consecutive words, and extra spaces or punctuation marks were corrected and the essays were entered into the Coh-

Metrix. The resulting data were inserted in the Statistical Package for the Social Sciences (SPSS) 21.0 for analysis. Descriptive data demonstrated that the mean text length of NS compositions ($M = 1578.43$, $s = 492.32$, range, 501-2530) was larger than NNS texts ($M = 407.18$, $s = 199.47$; range, 102-1003) on the whole. The mean text length of the total corpus was 870.23 words ($SD = 669.91$; range, 102-2530).

Prior to the data analyses, the fit between the variables' distribution and the assumptions of the statistical analyses was tested. No major violations of the assumptions were found. Data analyses proceeded to answer the following three research questions relating to the objectives of the study.

1. Is there a significant difference between advanced NNS learners' and NS learners' measures of vocabulary size and lexical diversity (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004; the *voc-D* and MTLTD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) as evidenced in academic writing?
2. Is there a relationship between vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser et al., 2004) and lexical diversity (as measured by the *voc-D* and MTLTD in Coh-Metrix; Duran et al., 2004; McCarthy & Jarvis, 2010) as evidenced in NNS and NS essays?
3. Is vocabulary size (as measured by the CELEX log frequency mean for all words index in Coh-Metrix; Graesser, et al., 2004) or lexical diversity (as measured by the *voc-D* and MTLTD in Coh-Metrix; Duran et al., 2004;

McCarthy & Jarvis, 2010) a greater predictor of writing score achievement (as measured by the TOEFL iBT Independent Writing Rubric; ETS, 2005) in non-native and native speaking college writing?

Results

Research question one. The first research question targeted the mean differences between native and non-native writers' profiles of lexical diversity and vocabulary size. It was hypothesized that NS texts would demonstrate significantly greater lexical diversity and vocabulary size than NNS essays (Laufer, 1994; Silva, 1993). A multivariate analysis of variance (MANOVA) revealed that NS texts exhibited significantly higher levels of lexical diversity and used lower-frequency words than NNS essays ($F_{3, 168} = 20.30, p < .05, \eta^2 = .27$). For the overall sample, the *voc-D* showed the greatest differences between NS and NNS texts ($F_{3, 168} = 55.02, p < .05, \eta^2 = .25$). Vocabulary size accounted for just 7% of the differences between the groups. However, when a separate MANOVA examined intragroup differences between compositions, only the MTLTD was able to detect significant differences among the NS sample ($F_{1, 66} = 4.17, p < .05, \eta^2 = .06$). For NNS group, there were significant differences on all three measures ($p < .001$).

These findings are consistent with previous studies examining lexical diversity. Linnarud's (1986) study of Swedish non-native and native English speakers ($n = 63$) concluded that a considerable portion of the differences between NS and NNS writing can be attributed to the variation of words. De Haan and van Esch (2005) arrived at a similar finding in their comparison NNS and NS Spanish learners. Their study

demonstrated that lexical diversity also strengthens over time as NNS grow their overall L2 proficiency.

Based on the present study's results and both Linnarud (1986) and de Haan and van Esch (2005), NS and NNS essays exhibit significant contrasts in terms of lexical diversity and vocabulary size. As no other studies were found that specifically compared vocabulary size and variation within this context, it is interesting to note that both measures of lexical richness differed significantly between NS and NNS essays; however, the only significant variation among the NS texts only was lexical diversity. This provides some preliminary evidence that when learners' productive vocabulary sizes are relatively the same, the richness of their lexis stems from vocabulary variation.

Research question two. Once differences between the two groups of essays within the corpus were established, the second research question sought to examine the variable of vocabulary size's ability to predict lexical diversity within the sample corpus. It was hypothesized that essays exhibiting the use of lower-frequency words, thus indicating a greater productive lexicon, would indeed correlate to lexical diversity (Daller, Van Hout, & Treffers-Daller, 2003). A Pearson product moment analysis indicated that essays with greater lexical diversity utilized lower-frequency words, but only to a moderate degree (MTLD [$r = -.44, p < .001$]; voc- D [$r = -.46, p < .001$]).

In a study of 48 advanced NNS university students, Laufer (1994) investigated if progress in written vocabulary size leads to lexical diversity. Participants' essays were analyzed once at the beginning of the semester and again at the end of the same semester.

Results demonstrated that although the sample's essays significantly increased in vocabulary size, lexical diversity remained unchanged.

Thus, the current study's findings along with Laufer's (1994) offer substantiation that compositions exhibiting vocabulary variation are not always those containing larger productive vocabulary sizes. Moreover, the moderate correlation offers further evidence to support the results from the MANOVA from research question one that lexical diversity is a stronger indicator of lexical richness than productive vocabulary size. This observation appears to be especially true when productive lexicons are relatively equal in size, as evidenced within the NS sub-corpus.

Research question three. The final and principal main research question examined to what extent the independent variables of lexical diversity and vocabulary size predicts the dependent variable of academic English writing proficiency among native and advanced non-native speakers of English. The directional hypothesis was that both would contribute significantly to writing score, but vocabulary diversity would have a greater correlation than vocabulary size (Crossley, Salsbury, McNamara, & Jarvis, 2010). A binary logistic regression demonstrated that lexical diversity was the only significant contributor to the model for both NS and NNS writings ($\text{Exp}[B] = 1.07, p < .001$). In other words, as the lexical diversity within an essay increased, so did its likelihood of earning a rating of 5. Vocabulary size, surprisingly, did not significantly contribute at all to writing score.

A third MANOVA was conducted with just the MTLT and CELEX word frequency measure by writing score to further validate the regression results and

discriminate the impact of the variables at each individual score level. Although both the MTLTD and CELEX scores significantly differed by each score level ($F_{6, 336} = 10.61, p < .001, \eta^2 = .16$), again, only the MTLTD accounted for a greater amount of the variation in ratings ($F_{3, 168} = 21.66, p < .001, \eta^2 = .28$). A post-hoc Bonferroni adjustment revealed that the essays earning scores of 4 and 5 significantly displayed greater lexical diversity than essays that scored at levels 2 and 3 ($p < .05$; Figure 3). Furthermore, there were no significant differences found in vocabulary diversity between essays rated at levels 2 and 3. Therefore, it can be deduced that lexical diversity significantly advances writing score from a level 3 and beyond. This finding aligns with Crossley et al.'s (2010) conclusion.

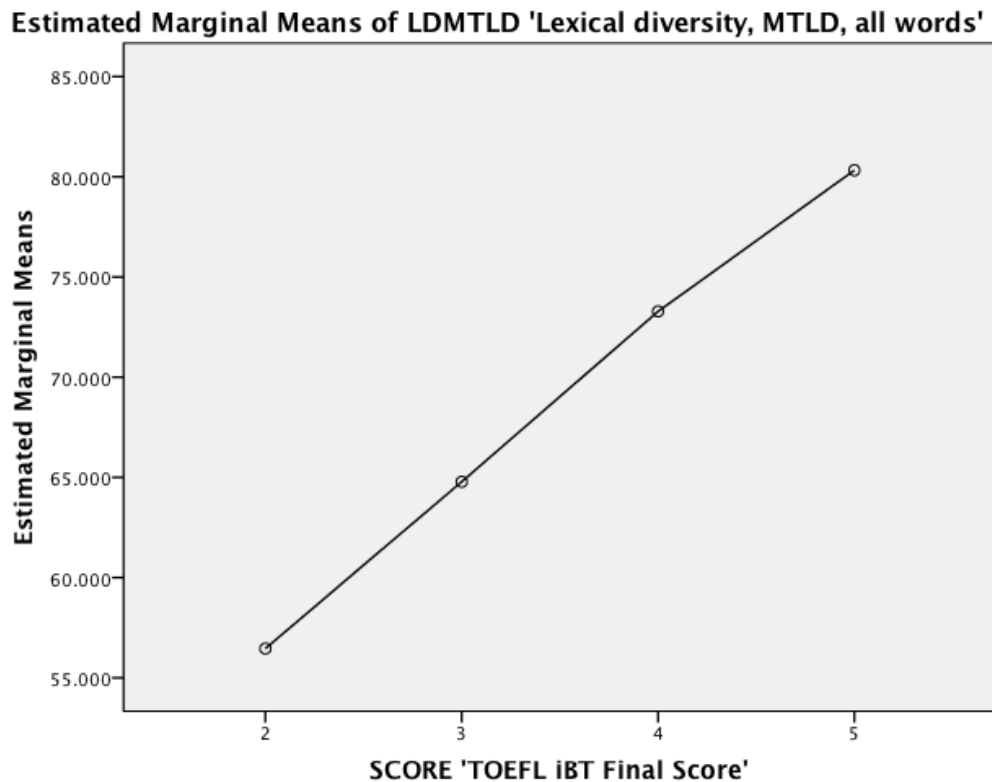


Figure 3. Lexical diversity by writing score

In contrast, vocabulary size demonstrated the opposite trend. Lower-rated essays tended to exhibit greater differences in their use of lower-frequency words, with the largest jump occurring between levels 2 and 3 ($p < .05$; Figure 4). An interesting observation, however, was that there were no significant differences in vocabulary size between levels 3 and 4 or between levels 4 and 5. This result indicates that vocabulary size initially facilitated the sample's writing quality, but its influence began to decline at score level 3. The fact that vocabulary size dropped at the same point where lexical diversity increased is a key finding. This trend implies that a writer's vocabulary size helps in the beginning, but it is their ability to diversify lexis that pushes their composition's quality into the 4 to 5 range.

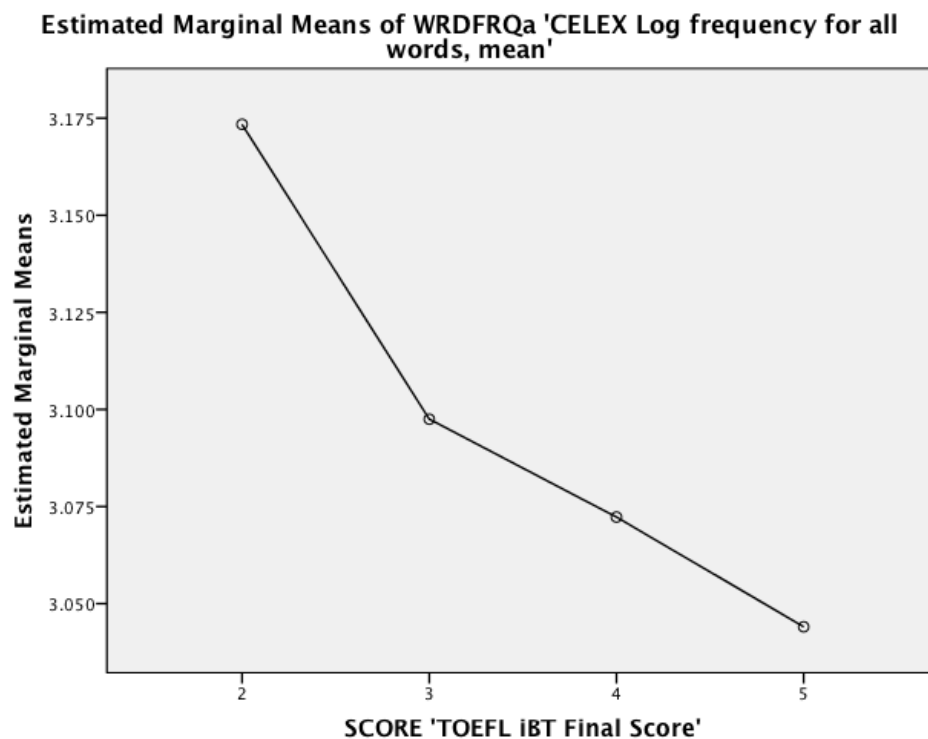


Figure 4. Means for vocabulary size by score level

Consequently, the logistic regression and post-hoc MANOVA results offer further support of Laufer's (1994) observation that writers who are best able to vary their lexis within their essays are not necessarily those who produce the rarest, most sophisticated words. The findings further validate previous studies' conclusions that lexical variation has the highest correlation to writing achievement (Grobe, 1981; Linnarud, 1986; McNamara, Crossley, & McCarthy, 2009; Schoonen, van Gelderen, de Glopper, Hulstijn, Snellings, Simis, & Stevenson, 2003).

Significance of the Findings

The primary contribution of this study is that it offers a baseline of the lexical richness devices that novice native speaking college writers employ within a typical composition assignment for credit. Through this model, a clearer profile and definition of proficient productive vocabulary use has emerged. To the researcher's knowledge, previous lexical richness studies have not focused solely on the comparison of NNS writers at the highest levels of L2 composition study to novice NS college writers. The express purpose of this examination is to investigate what distinguishes NS writers' productive lexical proficiency from NNS learners'.

Secondly, the study lends empirical support to anecdotal assertions that "big" words do not always equal proficient writing. All too often L2 vocabulary instruction and research focuses on the outer fringes of the word frequency bands and ignores the influences of mid-range vocabulary (Cobb, 2013; Schmitt, 2010). The quantitative results from this study indicate that proficient NNS and NS writers scoring at a level 5

averaged a word frequency score of 3.04, which falls close to the middle of the CELEX frequency score range of 0 to 6. Further study of the influences of medium- versus low-frequency words on writing quality may prove beneficial.

Additionally, the fact that only the MTLT explained the variance among NS essays while the *voc-D* was more reliable for NNS texts provides some validation of McCarthy and Jarvis' (2010) claim that the MTLT reliably resists text length effects. Since NS texts in this study were significantly longer than those produced by NNS learners, it can be inferred that the MTLT is a more discerning diversity measure for a wider range of text lengths, while the *voc-D* only performs well on shorter essays. Further support for the MTLT's hardiness against text length effects also appeared in post-hoc Pearson correlations that demonstrated that the *voc-D* had higher correlations to text length ($r = .54, n = 172$) than the MTLT ($r = .29, n = 172$). The fact that the MTLT appears to have solved the fatal flaw that has plagued other lexical diversity instruments opens research up to studying authentic essays produced in natural learning contexts without worrying about text length. Moreover, its inclusion within the Coh-Metrix also makes it easier and faster to analyze larger corpora in order to increase the generalizability of findings.

As a result of this study's use of the more robust MTLT, the outcome that lexical diversity played a significant role in achieving writing proficiency strengthens and validates similar conclusions from previous lexical richness studies (de Haan & van Esch, 2005; Grobe, 1981; Linnarud, 1986; Laufer, 1994, McNamara et al., 2009). As such, the

present study achieved its research goals to situate lexical diversity as an essential component of academic writing proficiency.

Limitations of the Study

Two fundamental limitations of the methodology of this study relate to the challenges text length, task topic, and writing genre pose for any study of lexical diversity. Different results may have been produced had word count, topic, and genre been controlled for either as covariates within data analyses or during data collection. Also, writing topic and genre could impact the generalizability of the results to essays outside of the seven genres contained within the sample corpus.

The final limitation relates to the generalizability of findings due to the demographics of the sample population as a general consequence of research. Data was collected from advanced NNS learners and NS university freshmen in an ESL context only. Thus, it is unknown how these results would apply to NS learners with lower levels of English proficiency or to NS students at higher stages of post-secondary education. Therefore, caution is advised when applying the outcome of this study to an English as a foreign (EFL) environment or to other student populations.

Implications for Practice

The results of this study predominantly highlight the importance of targeted vocabulary instruction within the L2 composition classroom. The finding that lexical diversity significantly impacts writing proficiency contains implications for methods of teaching and assessing L2 composition. Explicit vocabulary instruction for NNS

academic writers should not only focus on expanding learner lexicons, but also how to stylistically diversify those words in production.

This lexical expertise is especially vital for advanced NNS learners on the cusp of entering the mainstream college classroom where they will perform and be assessed alongside NS students. The ability to suitably vary words during composition and revision processes tends to be a hidden skill that may not be outright noticeable to learners without direct practice and attention. Furthermore, rubrics that assess L2 writing proficiency use blanket terminology within scoring benchmarks such as “appropriate word choice,” “sufficient range of vocabulary,” or “control of lexical features.” It can be argued that each of these lexical standards includes lexical diversity as a rating norm. Consequently, L2 writing instructors are charged with teasing out and defining the individual components comprising word choice, range, and control. Moreover, assessment practices in the L2 writing classroom need to provide learners with clear feedback that goes beyond the aforementioned blanket terms to state if a composition lacks lexical diversity and provide practice and textual examples in how it is achieved. The findings from this study suggest that mastery of lexical diversity can help better prepare NNS learners for the lexical demands of post-graduate writing tasks.

Lastly, the finding that vocabulary size did not contribute significantly at all to the likelihood of writers earning a proficient score of 5 is noteworthy. Current practices advocate for instructors to teach more vocabulary. Such lexical practices are evidenced in the emphasis on disciplinary vocabulary and writing in the Common Core State Standards for public K-12 schools and the popularity of word lists such as the Academic

Word List (AWL; Coxhead, 2000) in textbooks and instruction for adult academic English study.

Consequently, teachers are being told to explicitly teach vocabulary and often times the decision of which words to teach is largely left to intuition. The selection of which words to spend instructional time on therefore takes two approaches: (a) to follow the conventional recommendation to teach only the first 2,000 frequent words because words beyond that learners do not encounter often or will learn naturally, and/or (b) to pick out terms that are intuited to be the most “difficult” or content-related words. The former method assumes that lower-frequency vocabulary items can be acquired incidentally and ignores them, which research has shown is not always the case (Folse, 2004), while the latter focuses too heavily on the low-frequency, technical terms. Such approaches of teaching vocabulary ignore the wide-range of mid-frequency vocabulary that, as demonstrated in this study’s analysis, is necessary to allow learners to achieve the lexical diversity needed for writing quality.

Schmitt (2010) echoes this call for the teaching of mid-frequency terms. In reference to Nation’s (2006) recommended 98% text coverage, he states “if the learning target [for receptive proficiency] is 6,000 to 9,000 word families, it is clearly not realistic for learners to acquire lexis beyond the 2,000 level without a great deal of help” (p. 70). In terms of the productive language mode of writing, learners not only need 100% text coverage in order to avoid lexical errors, but also possess a targeted lexicon that allows them to vary word choices by employing synonyms, hypernyms, etc. As a result, it may be more beneficial for instruction to focus on a large number of mid-frequency words in

the 3,000- to 9,000-word range rather than teaching only a few low-frequency items as evidenced by the finding that vocabulary size only aided writing proficiency up to a level 3 on the TOEFL scale. In other words, the best vocabulary to teach and practice is not the most difficult or most rare vocabulary. Instead, a wiser use of instructional time might be to emphasize the use of mid-range vocabulary and related synonyms within good writing.

Recommendations for Future Research

The results from this study are first steps towards a larger exploration of how specific lexical richness features contribute to writing quality. Because of the inherent interlaced relationship between all measures of lexical richness, findings indicate for future research to further operationalize and quantify the moderate correlation of vocabulary size to lexical diversity. It remains unclear to what extent the usage of low-frequency words facilitates the diversification of lexis. It could be beneficial therefore to include measures of lexical density and sophistication within the models.

Another suggestion for further study would be to include measures of lexical error within the analysis. Nation and Webb (2011) point out that lexical richness factors could be obscured by errors and thus impact findings. Appropriately handling lexical errors in data analysis would further corroborate Engber's (1995) result that error-free lexical variation yields higher writing scores. As such, future investigations may choose to add variables relating to the types and frequencies of errors within a text (Nation & Webb, 2011).

The inclusion of an independent measure of general receptive and productive vocabulary size along the lines of Nation's Vocabulary Levels Test (1983) and Laufer and Nation's (1999) Productive Vocabulary Levels Test may further illuminate the differences between NS and NNS learners' lexical resource. Native speakers may demonstrate higher lexical variation in writing simply because they contain more words within their lexicons to achieve this important feature of lexical richness. Adding a measure of the size of learners' lexicons outside of their writing could shed light on how many words NNS writers need to sufficiently perform well on college writing assignments. Thus far, such a productive lexical threshold has yet to be identified.

A final recommendation to validate and strengthen the conclusions drawn in this study would be to replicate the study within intact first-year writing courses that have NNS writers alongside the NS peers. This replication would be useful in order to compare the writers' performance on the same topic and task. It also offers the opportunity to examine the contribution of lexical richness factors on authentic student grades. Likewise, the addition of instructors' feedback analysis could provide qualitative insights into their judgments of writing quality.

Conclusion

The present study examined the extent of the contributions that vocabulary size and lexical diversity impart to English writing proficiency in non-native and native academic compositions. Quantitative analyses revealed that lexical diversity has a greater impact on writing achievement than vocabulary size; however, vocabulary size

does initially help increase writing scores at the lower proficiency levels. These findings indicate that it is not enough to simply teach vocabulary words in the L2 composition classroom, but to also guide learners in how to employ these words in a varied manner during the composition process.

APPENDIX: IRB LETTER OF APPROVAL



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1
FWA00000351, IRB00001138**

To: **Melanie C Gonzalez**

Date: **April 09, 2013**

Dear Researcher:

On 4/9/2013, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review: Exempt Determination
Project Title: The Intricate Relationship between Lexical Richness and Academic Writing Proficiency
Investigator: Melanie C Gonzalez
IRB Number: SBE-13-09246
Funding Agency:
Grant Title:
Research ID: n/a

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 04/09/2013 09:15:40 AM EDT

A handwritten signature in cursive script that reads "Joanne Muratori".

IRB Coordinator

LIST OF REFERENCES

- Agustín Llach, M. (2007). Lexical errors as writing quality predictors. *Studia Linguistica*, 61(1), p. 1-19.
- Agustín Llach, M. (2011). *Lexical errors and accuracy in foreign language writing*. Bristol, UK: Multilingual Matters.
- Agustín Llach, M. & Gallego, T. (2009). Examining the relationship between receptive vocabulary size and reading and writing of primary school learners. *Atlantis*, 31(1), 11-29.
- Alderson, J.C. (2005). *Diagnosing foreign language proficiency*. London: Continuum.
- Arnaud, P. (1984). The lexical richness of L2 written productions and the validity of vocabulary tests. In T. Culhane, C. Klein-Bradley, & D.K. Stevenson (Eds.), *Practice and Problems in Language Testing*. Colchester: University of Essex.
- Astika, G.G. (1993). Analytical assessments of foreign students' writing. *RELC Journal*, 24(1), 61-70. doi: 10.1177/003368829302400104
- Baayen, R. H., Piepenbrock, R. & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Bardovi-Harlig, K. & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11(1), 17-34.

- Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, 16(1), 15-34.
- Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62(2), 384-414.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL quarterly*, 35(2), 331-336.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Bloom, B. (1956). *Taxonomy for educational objectives, handbook 1: The cognitive domain*. New York: McKay.
- Burke, E. & Wyatt Smith, C. (1996). Academic and non-academic difficulties: Perceptions of graduate non-English speaking background students. *TESL-EJ*, 2(1). Retrieved from <http://www.tesl-ej.org/wordpress/issues/volume2/ej05/ej05a1/>
- 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(1), 31-64.
- Campbell, D. T., & Stanley, J. C. (1963) *Experimental and quasi-experimental design for research*. Boston, MA: Houghton Mifflin.

- Casanave, C. P., & Hubbard, P. (1992). The writing assignments and writing problems of doctoral students: Faculty perceptions, pedagogical issues, and needed research. *English for Specific Purposes, 11*(1), 33–49.
- Chastain, K. (1980). Native speaker reaction to instructor-identified student second language errors. *The Modern Language Journal, 64*(2), 210-215. doi: 10.1111/j.1540-4781.1980.tb05187.x
- Chastain, K. (1981). Native speaker evaluations of student composition errors. *The Modern Language Journal, 65*(3), 288-294. doi: 10.2307/324155
- Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2), 103-116.
- Cobb, T. (n.d.). Compleat lexical tutor. Retrieved from <http://www.lextutor.ca/>
- Cobb, T. (2008). Commentary: Response to McQuillen and Krashen. *Language Learning and Technology, 12*(1), 109-114.
- Cohen, J. (1992). Power primer. *Psychological Bulletin, 112*, 155-159.
- Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly, 34*(3), 548-560.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.
- 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.

- Crossley, S.A. & McNamara, D.S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing, 18*(2), 119-135.
- Crossley, S.A., Salsbury, T., McNamara, D.S., & Jarvis, S. (2010). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing, 28*(4), 561-580.
- Cumming, A. (2001). Learning to write in a second language: Two decades of research. *International Journal of English Studies, 1*(2), 1-23.
- Cumming, A. (2006). Writing expertise and second language proficiency. *Language learning, 39*(1), 81-135. doi: 10.1111/j.1467-1770.1989.tb00592.x
- Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/research/policy_research_reports/rm-00-05_toefl-ms-18
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*(1), 5-43.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(1), 222-251. doi: 10.3102/0034654-3049002222
- Cummins, J., & Swain, M. (1986). *Bilingualism in education: Aspects of theory, research and practice*. London: Longman.

- Daller, H., Van Hout, R., & Treffers- Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197-222.
- De Groot, A.M.B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1001-1018.
- De Haan, P., & van Esch, K. (2005). The development of writing in English and Spanish as foreign languages. *Assessing Writing*, 10(2), 100-116.
- Dordick, M. (1996). Testing for a hierarchy of the communicative interference of errors. *System*, 24(3), 299-308.
- Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics*, 25(2), 220-242.
- Educational Testing Service. (2005). *Helping your students communicate with confidence*. Princeton, NJ: Educational Testing Service.
- Edwards, R. & Collins, L. (2011). Lexical frequency profiles and Zipf's law. *Language Learning*, 61(1), 1-30. doi: 10.1111 1/j.1467-9922.2010.00616.x
- Ellis, R. (2008). *The study of second language acquisition*. New York: Oxford University Press.
- Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139-155.
- Ferris, D. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, 28(2), 414-420.
- Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly*, 29(1), 33-53.

- Ferris, D. (2007). Preparing teachers to respond to student writing. *Journal of Second Language Writing, 16*(3), 165-193.
- Ferris, D. (2009). *Teaching college writing to diverse student populations* Ann Arbor: University of Michigan Press.
- Ferris, D. (2011). *Treatment of error in second language student writing* (2nd Ed.). Ann Arbor: University of Michigan Press.
- Ferris, D., Brown, J., Liu, H. S., & Stine, M. E. A. (2011). Responding to L2 students in college writing classes: Teacher perspectives. *TESOL Quarterly, 45*(2), 207-234.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.
- Fitzpatrick, T. & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing, 27*(4), 537-554. doi: 10.1177/0265532209354771
- Folse, K. (2004). *Vocabulary myths*. Ann Arbor: University of Michigan Press.
- Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*(2), 273-293.
- Folse, K. (2008). Myth 1: Teaching vocabulary is not the writing teacher's job. In J. Reid (Ed.) *Writing myths: Applying second language research to classroom teaching* (pp.1-17). Ann Arbor: University of Michigan Press.
- Folse, K. (2010). Is explicit vocabulary focus the reading teacher's job? *Reading in a Foreign Language, 22*(1), 139-160.
- Fraenkel, J.R. & Wallen, N.E. (2009). *How to design and evaluate research in education*. (7th ed.). New York: McGraw-Hill.

- Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(1), 341-363.
- Grabe, W., & Stoller, F. L. (2002). *Teaching and researching: Reading*. Boston: Allyn & Bacon.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123–145.
- Graesser, A. C., McNamara, D. S., Louwse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193–202.
- Green, D., & Meara, P. (1995). Guest editorial. *Computer Assisted Language Learning*, 8(2-3), 97-101.
- Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English*, 15(1), 75-85.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Harklau, L., Siegal, M., & Losey, K. (2009). *Generation 1.5 meets college composition: Issues in the teaching of writing to U.S. educated learners of ESL*. Mahwah, NJ: Taylor & Francis.
- Hawkey, R., & Barker, F. (2004). Developing a common scale for the assessment of writing. *Assessing Writing*, 9(2), 122-159.

- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure?. *Reading in a Foreign Language*, 8, 689-689.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.
- Hu, M. & Nation, I.S.P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 23(1), 403-430.
- Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments? A generalizability theory approach. *Assessing Writing*, 13, 201–218.
- Huang, J. (2009). Factors affecting the assessment of ESL students' writing. *International Journal of Applied Educational Studies*, 5(1), 1–17.
- Huang, J. & Foote, C. (2010). Grading between the lines: What really impacts professors' holistic evaluation of ESL graduate writing. *Language Assessment Quarterly*, 7(3), 219-233. doi: 10.1080/15434300903540894
- Institute of International Education. (2011). International student increase: Press release. Retrieved from <http://www.iie.org/Who-We-Are/News-and-Events/Press-Center/Press-Releases/2011/2011-11-14-Open-Doors-International-Students>
- International English Language Testing System. (2013). *IELTS task 1 and 2 band descriptors* [PDF document]. Retrieved from http://www.ielts.org/pdf/UOBDS_WritingT1.pdf

- Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Boston: Newbury House. Retrieved from http://seltmedia.heinle.com/resource_uploads/downloads/1424051010_35982.pdf
- Janapoulos, M. (1993). Comprehension, communicative competence and construct validity: Holistic scoring from an ESL perspective. In M.W. Williamson and B. Hout (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 303-325). Cresskill, NJ: Hampton Press.
- Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57-84.
- Johns, A. M. (2008). Genre awareness for the novice academic student: An ongoing quest. *Language Teaching*, 41(2), 237-252.
- Kinneavy, J. L. (1983). Writing across the curriculum. *Profession*, 13-20.
- Knoblauch, A. & Matsuda, P. (2008). First-year composition in the 20th century U.S. higher education: An historical overview. In P. Friedrich (Ed.) *Teaching Academic Writing* (pp. 4-22). New York: Continuum.
- Koizumi, R. (2012). Relationships between Text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction*, 1(1), 60-69.
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, 75(4), 440-448.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension. *Vocabulary and Applied Linguistics*, 3, 16-323.

- Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal*, 25(2), 21-33.
- Laufer, B. (1997). The lexical plight in second language reading. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy* (pp. 20–34). Cambridge: Cambridge University Press.
- Laufer, B. (1998). Development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B. & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B. & Nation, I.S.P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Laufer, B. & Nation, I.S.P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Laufer, B. & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15-30.
- Laufer, B., & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign language annals*, 18(5), 405-411.
- Leki, I. (1995). Coping strategies of ESL students in writing tasks across the curriculum. *TESOL Quarterly*, 29(2), 235-260.

- Leki, I., & Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28(1), 81-101.
- Lennon, P. (1991). Error and the very advanced learner. *International Review of Applied Linguistics*, 29(1), 31-44.
- Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Malmö, Sweden: Liber Förlag Malmö.
- Lovejoy, L. (1991). Cohesion and information strategies in academic writing: Analysis of passages in three disciplines. *Linguistics and Education*, 3(4), 315-343.
- Malvern, D.D., & Richards, B.J. (1997). A new measure of lexical diversity. *British Studies in Applied Linguistics*, 12, 58-71.
- Matsuda, P. (2008). Myth 8: International and U.S. resident ESL writers cannot be taught in the same class. In J. Reid (Ed.) *Writing myths: Applying Second Language Research to classroom teaching* (pp.159-176). Ann Arbor: University of Michigan Press.
- McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459-488.
- McCarthy, P.M., & Jarvis, S. (2010). MTLT, voc-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381-392. doi: 10.3758/BRM.42.2.381
- McNamara, D.S., Crossley, S.A., & McCarthy, P.M. (2009). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86. doi: 10.1177/0741088309351547

- Meara, P. (1990). Some notes on the Eurocentres vocabulary tests. Retrieved from <http://www.lognostics.co.uk/vlibrary/meara1990b.pdf>
- Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics*, 26(1), 32-47. doi:10.1093/applin/amh037
- Meara, P. & Bell, H. (2001). P_Lex: a simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5-19.
- Meara, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(1), 142-154.
- Melka, F. (1997). Receptive vs. productive aspects of vocabulary (pp. 84-102). In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press.
- Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System* 30(2), 225–235.
- Nation, I.S.P. (n.d.). English vocabulary levels test. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/vocabulary-size-test.pdf>
- Nation, I.S.P. (1983) Testing and teaching vocabulary. *Guidelines* 5(1), 12-25.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Heinle and Heinle.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82.

- Nation, P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language, 12*(2).
- Nation, I.S.P. & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schimitt & M. McCarthy (Eds.) *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge: Cambridge University Press. Retrieved from http://www.lex tutor.ca/research/nation_waring_97.html
- Nation, I.S.P. & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle-Cengage.
- Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). Crows Nest, Australia: Allen & Unwin.
- Paribahkt, T. & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 9-29.
- Ratner, N. & Silverman, S. (2000). Parental perceptions of children's communicative development at stuttering onset. *Journal of Speech, Language, and Hearing Research, 43*, 1252-1263.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal, 19*, 12-25.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355-371.

- Read, J. (1998a). Validating a test to measure depth of vocabulary knowledge. In A.J. Kunnan (Ed.), *Validation in Language Assessment* (pp. 41-60). Mahwah, NJ: Lawrence Erlbaum.
- Read, J. (1998b). Word associates test. Retrieved from <http://www.lex tutor.ca/tests/associates/>
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press. Retrieved from http://www.cup.es/servlet/file/store6/item2466355/version1/item_9780521627412_frontmatter.pdf
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.) *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (pp. 209-227). Amsterdam: John Benjamins.
- Roberts, F. & Cimasko, T. (2008). Evaluating ESL: Making sense of university professors' responses to second language writing. *Journal of Second Language Writing, 17*(3), 125-143.
- Ruegg, R., Fritz, E., & Holland, J. (2011). Rater sensitivities to qualities of lexis in writing. *TESOL Quarterly, 45*(1), 63-80.
- Santos, T. (1988). Professors' reactions to the academic writing of non-native-speaking students. *TESOL Quarterly, 22*(1), 69-90.
- Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly 18*(1), 199-219.

- Scarcella, R. (2002). Some key factors affecting English learners' development of advanced literacy. In M.J. Schleppegrell & M.C. Colombi (Eds.), *Developing advanced literacy in first and second languages: Meaning with power*, (pp. 209-226). Mahwah, NJ: Lawrence Erlbaum.
- Scarcella, R. (2003). Academic English: A conceptual framework. Technical Report 2003-1. *University of California Linguistic Minority Research Institute*.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave MacMillan.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, 18(1), 55-89.
- Schmidt, N. & Zimmerman, C.B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36(2), 145-171.
- Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, 53(1), 165-202.
- Schoonen, R., van Gelderen, A., Stoel, R.D., Hulstijn, J., & de Glopper, K. (2010). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning*, 61(1), 31-79. doi: 10.1111 1A.1 467-9922.2010.00590.x

- Shavelson, R. J. (1996). *Statistical reasoning for the behavioral sciences* (3rd ed.). Boston: Allyn and Bacon.
- Silva, T. (1992). L1 vs. L2 writing: ESL graduate students' perceptions. *TESL Canada Journal*, 10(1), 27-47.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657-677.
- Simpson-Vlach, R. & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Singhal, M. (2004). Academic writing and generation 1.5: Pedagogical goals and instructional issues in the college composition classroom. *The Reading Matrix*, 4(3). 1-13.
- Staehr, L.S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152.
- Sweedler-Brown, C. O. (1993). ESL essay evaluation: The influence of sentence-level and rhetorical features. *Journal of Second Language Writing*, 2(1), 3-17.
- Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.
- Tidball, F. & Treffers-Daller, J. (2008). Analyzing lexical richness in French learner language: What frequency lists and teacher judgments can tell us about basic and advanced words. *Journal of French Language Studies*, 18(3), 299-313. doi: 10.1017/S0959269508003463

- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011), Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569–613. doi: 10.1111/j.1467-9922.2010.00622.x
- Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, 32(5), 323-352.
- UCF Writing and Rhetoric Program. (n.d.). First-year writing program. Retrieved from <http://writingandrhetoric.cah.ucf.edu/composition.php>
- U.S. Census Bureau. (2011). American community survey. Retrieved from <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>
- Vermeer, A. (1992). Exploring the second language learner lexicon. In L. Verhoeven & J. de Jong (Eds.), *The construct of language proficiency*, (pp. 147-162). Amsterdam: John Benjamins.
- Wang, L. (2003). Switching to first language among writers with differing second-language proficiency. *Journal of Second Language Writing*, 12(4), 347-375.
- Wang, W. & Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, 11(3), 225-246.
- Waring, R. (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata*, 1(1), 53-68
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30, 79-95. doi: 10.1017/S0272263108080042

- Weigle, S. C., Boldt, H., & Valsecchi, M. I. (2003). Effects of task and rater background on the evaluation of ESL writing: A pilot study. *TESOL Quarterly*, 37(2), 345–354.
- West, M. (1953). *A general service list of English Words*. London: Longman, Green.
- Xue, G. & Nation, I.S.P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215-229.
- Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31(2), 236-259.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27(4), 567-595.