UNIVERSITY OF CALIFORNIA,
IRVINE

Unpacking Adolescent Writers' Texts: A Systematic Investigation of the Language Features in the Academic Writing of Linguistically Diverse Students

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Undarmaa Maamuujav

Dissertation Committee:
Professor Emerita Carol Booth Olson, Chair
Professor Emerita Robin Scarcella
Professor Penelope Collins

2022

DEDICATION

To

My father,

I dedicate this work to my father,
who taught me the value of hard work, commitment to cause beyond your own,
and compassion for those less fortunate.

Аавдаа зориулав

Энэхүү эрдмийн бүтээлээ хайрт аав таньдаа зориулав.
Таныхаа хөдөлмөрч бүтээлч байдал, хүнд тусархаг зан, хязгааргүй сайхан сэтгэлийг
дээдлэн хэвшүүлж явъа.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work would not have come to fruition without the support, guidance, care, generosity, and trust of the many individuals, including mentors, colleagues, friends, and family members whom I have been fortunate to have in my life.

I would like to express my sincere gratitude to the committee of my dissertation and dissertation prospectus — Dr. Carol Booth Olson, Dr. Robin Scarcella, Dr. Penelope Collins, Dr. Lisa Pearl, and Dr. Bahiyyih Hardacre. It has been an honor to present my work to the committee of these brilliant scholars and inspirational women. I am blessed to have you all in my committee, and I am indebted for your time, support, and encouragement without which this scholarly endeavor would not have come to completion.

My deepest appreciation goes to my advisor, mentor, and the role model, Dr. Carol Booth Olson, whose support and guidance have been instrumental in not only this project but the entire journey of my Ph.D. pursuit and my development as a scholar and researcher. The success of my scholarly endeavor is because of the tremendous support and guidance that you have generously provided. It was also because you believed in me. Your trust and encouragement have been the impetus for my resolve to keep going, to overcome challenges, and to persevere in spite of many setbacks. Beyond the encouragement, guidance, and support, your understanding and empathy have had a profound effect. When I was in a vulnerable state grieving the loss of my father, you came in with a beautiful white orchid plant, gave me a warm hug, and shared your own experience of losing your father at a young age. I will never forget this moment. Your commitment to cause beyond your own and your drive to improve the educational attainment of the marginalized youths have been inspirational. I am blessed to be mentored by such an amazing woman who embodies the kind of compassion, commitment, and zeal that many of us can only aspire to.

I express my profound gratitude to Dr. Huy Chung for the support on this project and on my research and work in general. Dr. Chung has been instrumental in the data collection of this dissertation by recruiting teachers from multiple states and connecting them with me. Without your support and help not only on the data collection but on many other tasks (IRB compliance, piloting surveys, preparing packages to teachers to name a few), this dissertation research would not have been possible. Dr. Chung has served as my immediate supervisor for the graduate student researcher position that I have held throughout my time in the Ph.D. program. I have been inspired by the remarkable sense of personal integrity and intellectual tenacity, as well as humility, you have modeled. I have learned so much from you and look forward to many opportunities of collaboration in the future.

I am tremendously grateful to Dr. Penelope Collins whose guidance, collaboration, and support have been extremely influential in my academic journey. I have had the opportunity to collaborate with Dr. Collins on multiple projects during my time in the Ph.D. program. I have taken multiple classes you have taught in the School of Education, and these classes have been pivotal in guiding me to the direction of this dissertation research. In fact, it was in your Adolescent Literacy course in which I drafted the research proposal for this dissertation. I have greatly benefited from your expertise in literacy research, technological savviness, and your

generous mentorship. Thank you so much for extending an invitation to collaborate, providing guidance, feedback, and encouragement, and modeling acts of kindness and courteousness.

My sincere gratitude also goes to Dr. Robin Scarcella, who has supported me not only through my Ph.D. journey but through my teaching career prior to my doctoral pursuit. Dr. Scarcella was my supervisor in the Academic English program at UCI and wrote a very strong recommendation letter in support of my application to the Ph.D. in Education. Without your support and encouragement, I would not have been able to join the program. I have been fortunate to continuously benefit from your counsel, guidance, and support on this research and on many other collaborative works. As a matter of fact, Dr. Scarcella was one of the coders for the analytic scoring of the essays for this dissertation research, spending countless hours coding hundreds of essays and meeting with me to discuss ways to resolve discrepancies and improve the analytic scoring framework for language features. I am immensely grateful for your support.

I also thank Dr. Bahiyyih Hardacre, my faculty mentor for the Chancellor's Doctoral Incentive Program Fellowship. Dr. Hardacre and I have collaborated on various professional endeavors, including serving as a college and university level chair and co-chairs for CATESOL, a professional organization serving teachers of English in California. Working with you on the many ventures we have taken, I have learned a great deal from you. Thank you for agreeing to serve as my faculty mentor and for your collaboration and friendship. I look forward to continuing my collaboration with you to improve teaching and learning of English and academic literacy to diverse learners.

A special thank you to Dr. Lisa Pearl, who has graciously agreed to serve on my dissertation prospectus committee and who invited me to join the awesome QuantLang Lab at UCI. Being part of the Language Science community has been such a rewarding experience as I am learning so much from the brilliant scholars and students in the Department of Language Science. The remarkable presentations and vibrant discussions that happen during QuantLang Lab meetings have been helping me to expand my knowledge about computational linguistics, cognitive modeling, and corpus-based linguistics research. So thank you for extending an invitation to join the lab and welcoming me into the language science community.

I thank Dr. Jenell Krishnan for the mentorship and collaboration since I was a first-year doctoral student. Your wholehearted advice on how to navigate the doctoral journey and your willingness to share from your own experience have made you my go-to person for big and small questions. I am fortunate to have collaborated with you on several projects and manuscripts. I have always enjoyed our collaborative writing sessions as working with you has made the daunting task of writing a manuscript a less lonely and more pleasurable undertaking. I have also learned a great deal from your amazing leadership and organizational skills.

The support of my colleagues in the Pathway Lab, including Jacob Steiss, Minkyung Cho, Jane Nazzal, Vicky Chen, and Lauren Godfrey, has been instrumental in my development as a writer and researcher. I sincerely appreciate the time you have spent reading long, verbose, badly written first drafts of the many manuscripts. The detailed comments and constructive feedback you provided have been extremely valuable in revising and refining the manuscripts and in improving my writing skills.

I thank the faculty, staff, and the community of the School of Education at UCI, including the UCI Writing Project staff whose behind-the-scenes support has been integral in my journey. I thank all members of the UCI community who are striving to sustain a professional, positive, and inclusive community. Being part of the community has been a truly wonderful experience. The faculty members whom I have met and taken classes with, the staff who are behind the scenes, and students whom I have had conversations with have been extremely cordial, professional, and willing to listen and help.

I also would like to express my sincere gratitude to the teachers and students who participated in this dissertation research, as well as the site leaders and literacy coaches. I am inspired by the resolve and the resilience of the many educators and teachers who are working with diverse learners with different needs, backgrounds, and experiences. Your work is important in advancing education and improving the lives of many students.

To my father and mother, I am forever indebted. Without your love, sacrifice, and support, I would not have been here. My accomplishments are the fruit of your unwavering devotion to the education, well-being, and happiness of your children. Both of you didn't have the opportunity to go to college due to life's adversity and a corrupt system, but you worked so hard and made many sacrifices for the education of your children. You have taught me the value of hard work, resilience in the face of adversity, and the virtue of selflessness. This degree is for you.

To my husband, the love of my life, the father of my children, I cannot thank you enough. Without your sacrifice and support, I could not have done it. When you return home after the hard labor and many days on the road, you roll your sleeves to clean the dishes piled up in the sink, to do the laundry that had been waiting for you, and to cook our meals. You have never complained about it. I depend on you and your love and support. To your children, you are an awesome dad. To me, you are a loving husband, devoted father, and very kind man. Thank you for loving me so.

To my brothers and sisters-in-law, I am grateful for your love and support. You have provided me with emotional support when I was going through the dark time of depression. During a tough year of pandemic, grief, and financial insecurity, my big brother took us under his wing and provided us a home in his house for almost a year. I thank my sister-in-law for caring for my children and cooking our meals during this time. I am blessed with a close bond with my brothers and their wonderful spouses who are always there to support me in any way they can.

Finally, I am grateful for the love and understanding of my three beautiful children—Anu, Maika, and Todd. For these past several years, your mom has been busy glued to her chair and desk, but you have shown so much maturity and independence over the years. You have been kind and considerate. Thank you for appreciating what you have, for not complaining about what you cannot have, for striving to be the best you can be. You make my life complete.

I gratefully acknowledge that this dissertation research is part of the Pathway to Academic Success Project that was funded by the U.S. Department of Education's Education Innovation and Research Grant (grant number U411A180004).

VITA

Undarmaa Maamuujav

School of Education
University of California, Irvine

EDUCATION

Ph.D., Education 2022
University of California, Irvine

M.S., Public Administration 2010
California State University, Los Angeles

M.A., TESOL (Teaching English to the Speakers of Other Languages) 2005
California State University, Los Angeles

M.A., Linguistics 2002
University of the Humanities, Ulaanbaatar, Mongolia

B.A., English and Education 2000
University of the Humanities, Ulaanbaatar, Mongolia

AWARDS, GRANTS, and FELLOWSHIPS

Chancellor's Doctoral Dissertation Fellowship, $5,000 2021
Chancellor's Doctoral Incentive Program, California State University

Summer Inclusive Excellence Grant, $5,000 2021
Graduate Division, University of California, Irvine

Otto W. Shaler Scholarship, $6,000 2021
Graduate Division, University of California, Irvine

Pedagogical Fellow, $2,500 2021
Division of Teaching Excellence and Innovation, University of California, Irvine

CSU Chancellor's Doctoral Incentive Program Fellowship, $30,000 (loan) 2020
California State University

Teaching as Research Fellowship, $1,000 2020
Teaching and Learning Research Center, University of California, Irvine

David E. Eskey Memorial Award, $1,000 2020
CATESOL Education Foundation

Education Research Initiative Grant, $30,000 2019
University of California, Irvine

TEACHING AND PROFESSIONAL CERTIFICATES

Certificate of Completion: Structural Equation Modeling and MPlus, June 2021
Stats Camp, Statistics Training Seminar

Certificate in Course Design: Transforming Your Research into Teaching, July 2020
Center for the Integration of Research, Teaching, and Learning (CIRTL)

Certificate of Teaching Excellence, March 2020
In Recognition of Advanced Training in Evidence-based Pedagogy
UCI Division of Teaching Excellence and Innovation

Course Design Certificate in Student-centered Course Design, March 2020
UCI Division of Teaching Excellence and Innovation

Certificate of Completion, March 2020
Center for the Integration of Research, Teaching, and Learning Associate Level
UCI Division of Teaching Excellence and Innovation

PUBLICATIONS

Maamuujav, U., & Hardacre, B. (under review). The Politics and Praxis of Academic English: Towards Antiracist Language Pedagogy. *TESOL Journal.*

Olson, C. B., Maamuujav, U., Steiss, J., & Chung, H. (under review). Examining the impact of a cognitive strategies approach to teaching arguments of literary analysis on the writing of mainstreamed English learners in secondary school. *Written Communication.*

Maamuujav, U., Olson, C. B., & Scarcella, R. (2022). Sentence boundaries in academic writing: Helping L2 students develop sentence fluency in secondary school. *The TESOL Encyclopedia of English Language Teaching.* Hoboken, NJ: John Wiley & Sons, Inc.

Maamuujav, U., Krishnan, J., & Collins, P. (2022). Beyond words: Info & graphic meshing. *Communicator,* Winter 2022, The Institute of Scientific and Technical Communicators.

Maamuujav, U., Olson, C. B., & Chung. H. (2021). Syntactic and lexical features of adolescent L2 students' academic writing. *Journal of Second Language Writing.*

Maamuujav, U. (2021). Examining lexical features in adolescent L2 students' text-based analytical writing. *Assessing Writing.*

Krishnan, J., Maamuujav, U., & Collins, P. (2021). Multiple utilities of infographics in undergraduate students' process-based writing. *Writing and Pedagogy.*

Warschauer, M.**,** Jacob, S., & Maamuujav, U. (2020). Online Englishes. In A. Kirkpatrick (Ed.), *The Routledge Handbook of World Englishes*.

Maamuujav, U., & Olson, C. B. (2019). Meeting the linguistic needs of adolescent multilingual writers for academic writing. *The TESOL Encyclopedia of English Language Teaching.* Hoboken, NJ: John Wiley & Sons, Inc.

Maamuujav, U. (2019). Developing autonomous self-editors: An alternative approach to written corrective feedback. *CATESOL Journal*.

Maamuujav, U., Krishnan, J., & Collins, P. (2019). The utility of infographics in L2 writing classes: A practical strategy to scaffold writing development. *TESOL Journal,* 484, 1-13.

CONFERENCE PRESENTATIONS IN 2022

Maamuujav, U. (March 20-22, 2022). Examining Linguistic and Rhetorical Features in Text-Based Analytical Writing of Adolescent Multilingual Students, *AAAL 2022 Annual Conference*, Pittsburgh, PA.

Maamuujav, U. & Hardacre, B. (March 19-22, 2022). The Antiracist Language Pedagogy (ALP) Framework for Academic English Development: Critical, Inclusive, and Multiliteracies Approach, *AAAL 2022 Annual Conference*, Pittsburgh, PA.

Maamuujav, U. (March 22-25, 2022). Of Academic English: The Politics and Praxis. Conceptual paper presentation accepted at *TESOL International Convention*; Pittsburg, PA.

Maamuujav, U. (March 22-25, 2022). Digital Tools to Scaffold Multilingual Students' Multimodal Writing. "Advances in Technology for Language Teacher Preparation" Panel at *TESOL International Convention*; Pittsburg, PA.

Maamuujav, U. (March 22-25, 2022). Examining linguistic features of multilingual students' academic writing. Invited to the TESOL Research Agenda Panel at *TESOL International Convention*; Pittsburg, PA.

Eslami, M., Maamuujav, U., Collins, P. (April 20-22, 2022). WISE Project: Improving Undergraduate Instruction in Writing and Communication — Research presentation accepted at AMPS (Architecture, Media, Politics, Society), A Focus on Pedagogy: Teaching, Learning, and Research in the Modern Academy, Virtual Conference.

EMPLOYMENT

Graduate Student Researcher, 2019—2022
School of Education, University of California, Irvine

Lecturer, 2017—2018
Courses taught: Academic English; Critical Reading and Rhetoric
Academic English Program & Composition Program
University of California, Irvine

Lecturer, 2005—2018
Department of English, Undergraduate Studies, English Language Program
California State University, Los Angeles

Writing Tutor & Consultant, 2003—2005
University Writing Center, California State University, Los Angeles

Lecturer, 2000—2002
Courses taught: English, Linguistics, Teaching Pedagogy
Department of English, University of the Humanities, Ulaanbaatar, Mongolia

Mentor Teacher, 2001—2002
National Teacher Training Project
Ministry of Education, Science and Technology, Ulaanbaatar, Mongolia

AREAS OF INTEREST

Academic Writing, Adolescent Literacy, Composition and Rhetoric, Corpus Linguistics, Second Language Acquisition, Second Language Writing, Digital and Multimodal Tools for Writing

ABSTRACT OF THE DISSERTATION

Unpacking Adolescent Writers' Texts: A Systematic Investigation of the Language Features in the Academic Writing of Linguistically Diverse Adolescent Students

By

Undarmaa Maamuujav

Doctor of Philosophy in Education

University of California, Irvine, 2022

Professor Emerita Carol Booth Olson, Chair

With a premium placed on academic writing in the Common Core State Standards (CCSS, 2010), adolescent students in secondary school are expected to develop advanced skills to analyze, interpret, and produce complex texts in a variety of content areas. To be able to accomplish the higher-order tasks of analyzing, interpreting, and producing academic texts, students need to develop proficiency in the specialized language of academic written discourse. While the expectations and needs are high, the results of the writing assessment conducted by the National Assessment of Educational Progress (NAEP) presents a bleak picture of secondary students' writing performance (U.S. Department of Education, National Center for Education Statistics, 2012, 2017). Despite the emphasis on academic literacy in middle and high school and the continued underperformance on national assessments, little research has been conducted on the language and writing features of linguistically diverse adolescent students' academic writing. This dissertation research aims to fill this gap in the literature and address the pressing issue pertaining to the academic language and literacy skills of adolescent writers by systematically examining the linguistic features of text-based analytical essays written by diverse students (7^{th}–12^{th} grades) attending public schools in the United States.

The dissertation consists of three related studies that each address specific linguistic features of students' text-based analytical writing. Study 1 focuses on the lexical features and academic vocabulary use in a sample of text-based analytical essays (n=70) written by multilingual students whose first language (L1) is Spanish. I analyze lexical density, diversity, and sophistication and examine how these three lexical features predict human-judged writing quality. In Study 2, I analyze lexical and syntactic features and their relations to writing quality using the same learner corpus, but with a slightly larger sample (n=86). Study 3 takes a more holistic approach and examines lexical, syntactic, and rhetorical features and their relations to one another and to writing quality. This study uses a much larger learner corpus of text-based analytical essays written by linguistically diverse students in grades 7–11 (n=410) from public schools in 5 different states. I use the same lexical and syntactic features as in the first and second studies and add rhetorical features, including text structure and cohesion, source-use and integration, and balance of summary, evidence, and commentary.

Collectively, these three studies: 1) provide a descriptive picture of middle- and high-school students' academic English use and academic writing proficiency within the genre of text-based analytical writing; and 2) contribute to our understanding of the complex relations among lexical, syntactic, and rhetorical features and writing quality. Studying the language features of student texts that are produced as part of an academic writing genre and their relation to rhetorical features and writing quality provides insight into understanding students' linguistic needs as they strive to meet the demands of academic writing. Having a better understanding of the linguistic characteristics of multilingual students' writing can inform pedagogy that aims to address the academic literacy of this student population in a secondary school context. The infographic below visually presents the abstract and outlines the three studies of the dissertation.

INFOGRAPHIC OF THE DISSERTATION

Unpacking Adolescent Writers' Texts: A Systematic Investigation of the Language Features in the Academic Writing of Linguistically Diverse Students

The Principal Goals of the Dissertation Research:

- **To identify language use and linguistic patterns exhibited in diverse adolescent students' (grades 7-12) academic writing**
- **To investigate whether and how the identified linguistic features relate to rhetorical features and human-judged writing quality**

Study 1
Lexical Features and Academic Vocabulary Use in Adolescent L2 Students' Text- based Essays

Research Questions:

2. What degree of lexical density, diversity, and sophistication are observed?
3. To what extent do the students use academic vocabulary?
4. To what extent do the three lexical features predict writing quality?

Methodology:

- Unit of Analysis: Text-based analytical essays (n=70) written by Spanish-speaking L2 students
- Measures: Lexical density, Lexical diversity, and Lexical sophistication
- Analytic Strategies: Coh-Metrix, VocabProfilers, Descriptive statistics, Pearson Correlation, Multiple Regression

Results:

- Low lexical density and diversity.
- Reliance on basic and frequently-used words.
- Lexical density and lexical diversity do not predict writing quality.
- Lexical sophistication predicts holistic writing score.
- 44% of the AWL words in students' texts are from the source and the prompt.

Study 2
Syntactic and Lexical Features of Adolescent L2 Students' Academic Writing

Research Questions:

1. What syntactic and lexical characteristics do students' text-based analytical essays exhibit?
2. To what extent do syntactic and lexical features predict writing quality?

Methodology:

- Unit of Analysis: Text-based analytical essays (n=86) written by Spanish-speaking L2 students
- Measures: Syntactic complexity & Lexical sophistication
- Analytic Strategies: SFL approach, Coh-Metrix, Descriptive statistics, Pearson Correlation, Multiple Regression

Results:

- Unconventional sentences and basic coordination and subordination are common syntactic patterns.
- Low use of academic vocabulary.
- Syntactic features explain 44% of the variance in writing quality.
- Lexical features explain 23% of the variance in writing quality.
- Syntactic complexity and lexical sophistication predict quality.

Study 3
Linguistic and Rhetorical Features in Text-based Analytical Writing of Diverse Adolescent Students

Research Questions:

1. What lexical, syntactic, and rhetorical features are observed?
2. To what extent do the identified features vary among L1 groups?
3. How do the lexical, syntactic & rhetorical features relate to one another and predict writing quality?

Methodology:

- Unit of Analysis: Text-based analytical essays (n=410) of linguistically diverse students
- Measures: Lexical, Syntactic, and Rhetorical features
- Analytic Strategies: SFL approach, NLP tools, Analytic Scoring, Descriptive statistics, MANOVA, CFA, Structural Regression

Results:

- Low use of advanced and academic words.
- Less syntactic variety at clausal and phrasal levels.
- Summary and summary dominant structures are common patterns.
- Variability among students with different L1s.
- Rhetorical features predict writing quality while lexical and syntactic features are strongly correlated with rhetorical features.

The Significance & Contribution of the Dissertation Research:

- addresses the research gap by examining lesser-studied writing of underexplored student population
- has the potential to shed light on how to best support diverse adolescent students for academic writing

CHAPTER 1

Introduction

Developing proficiency in academic writing and language is central to educational attainment and career success. The significance of academic writing is emphasized by Graham and Perin (2007), who assert that writing effectively in an academic context is "not just an option for young people—it is a necessity" as it is "a predictor of academic success and a basic requirement for participation in civic life and in the global economy" (p. 3). In addition, students' proficiency in academic language is cited as one of the crucial factors affecting their academic success (Francis et al., 2006). Hence, an ability to effectively convey thoughts and ideas using academic language is a requisite skill for students who strive to pursue a path of higher education. According to the *Framework for Success in Postsecondary Writing* (2011) developed jointly by the Council of Writing Program Administrators (CWPA), National Council of Teachers of English (NCTE), and National Writing Project (NWP), students' success in college and beyond depends heavily on their ability to write and communicate well in an academic context. Given the importance of academic writing and language proficiency to college and career success, the curriculum and instruction for literacy development in a secondary education context have placed considerable emphasis on students' academic writing and language development.

In the United States, the importance of academic language and literacy development are reflected in the Common Core State Standards (CCSS, 2010) and the Next Generation Science Standards (NGSS, 2013) as the goals of these standards are to prepare America's youth for college and career readiness and to help them succeed in the global economy. These standards emphasize the need for students to develop skills to effectively communicate in academic,

educational, and professional contexts. Consequently, the Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects (CCSS-ELA) for grades 6-12 have set a high bar for students, requiring them to be able to produce clear, coherent, and well-developed writing (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). By the end of 6th grade, students are expected to develop a claim about a complex topic and substantiate it with relevant examples and evidence using credible sources while maintaining a formal style of writing. They must demonstrate "the ability to analyze and interpret challenging texts and to write about those texts using academic discourse in extended pieces of writing" (Olson et al., 2015, p. 5). In addition, students are expected to use precise language, domain-specific vocabulary, and nuanced transitional devices to show relations between ideas. In short, the CCSS-ELA's emphasis on academic writing demands higher levels of analytical writing proficiency and academic English fluency. But are the students in middle and high school performing at the level of proficiency called for in the CCSS?

As indicated in the National Assessment of Educational Progress (NAEP; 2012, 2019), only about one-quarter of graduating high school students show solid performance in academic writing. The results of the first computer-based writing assessment conducted by NAEP indicate that only about 20% of students in 8th grade, 18% of students in 12th grade, and only 1% of English Learners (ELs) in both grades perform at or above the proficient level in writing (U.S. Department of Education, Institute of Educational Sciences, National Center for Education Statistics, 2012). The NAEP report shows that the majority of the students in middle and high school in the United States struggle as academic writers and in meeting the expectations set forth by CCSS. The alarming 1% proficiency level for ELs indicates that the challenge and demands

of academic writing are much greater for many multilingual students who are in process of developing their proficiency in English. Consequently, a large number of adolescent students in the United States graduate from high school "unable to write at the basic levels required by colleges or employers" (Graham & Perin, 2007, p. 3). As many students in the United States have shown poor performance in writing skills over time, the National Commission of Writing (2003) has declared writing as a neglected skill. In general, adolescent students entering college have not had enough practice in their secondary education to develop sophisticated writing skills and to meet the demands of academic writing (Eberly Center, Teaching Excellence & Educational Innovation, Carnegie Mellon University, 2021).

Students' under-preparedness for academic writing and language is an equity issue as the lack of preparation is pervasive among English learners, language-minority students, underrepresented racial and ethnic populations, first-generation college students, and low-income immigrant students from underserved communities (Bowen et al., 2005; Kanno & Cromley, 2015; Kim & Diaz, 2013). Multilingual students who have not developed proficiency in academic literacy are less likely to graduate from high school, let alone enroll in college (Núñez et al., 2016). Examining the patterns of students' access to college and attainment of academic degrees, Kanno and Cromley (2015) found that ELs lagged far behind their non-EL counterparts in both college enrolment and degree attainment. They concluded, "Four-year college access and bachelor's degree attainment were beyond the reach of many ELLs" (p. 2). Many of the English learners and language-minority students have K-12 experiences that leave them less prepared for the writing demands of higher education. Differences in social and educational opportunities, institutional resources, and instructional contexts within schools contribute significantly to the inequitable development of writing and language skills, giving rise to inequity in educational

attainment. In general, lack of equity, educational barriers, and academic literacy challenges are pressing issues facing many multilingual students who represent underrepresented populations, including racial and ethnic minority, low-income immigrant, and first-generation college students.

English Learners as Multilingual Students in Process of Developing Proficiency in English

There is no uniform or agreed upon definition of English learners as they are called many names including "English language learners," "multilingual students," "language-minority students," and "second language learners." In general discussions, I will refer to them as adolescent multilingual students. In the context of my research, I will refer to them as developing multilingual writers who are in process of building proficiency in English or "developing multilingual writers" in short. This is: 1) to recognize that these students use multiple languages in different communicative contexts; 2) to indicate that the first language (L1) or the home language they were exposed to as a child is not English; and 3) to situate them in the context of the dissertation research which focuses on academic writing. This growing student population today represents a diverse, heterogeneous group with differing linguistic, cultural, and educational experiences and backgrounds (Harklau et al., 1999; Matsuda et al., 2006). As of Fall 2017, about 10% of students attending public schools in the United States were classified as English learners (U.S. Department of Education, NCES, 2020), and the number of English learners has been continuously increasing. Many of these students have a dominant home language that is not English and are still in process of developing their English language proficiency. In addition, they often have limited knowledge of academic registers in any of their languages and, thus, the challenge of engaging in highly demanding academic tasks called for by the CCSS are enormous for this student population (Bunch et al., 2012; Olson et al., 2017).

While enrollment of these students in U.S. public schools has been increasing, literacy and opportunity gaps remain, and "the challenges posed by the Common Core State Standards make those gaps more glaring" (Goldenberg, 2013, p. 10).

Despite their growing rate in the U.S. secondary education context, adolescent multilingual students are an underexplored student population in the field of writing in second or additional languages (L2 writing), as well as in English for Academic Purposes (EAP) research and literature (Cumming, 2016; Harklau & Pinnow, 2009; Ortmeier-Hooper & Enright, 2011; Yi et al., 2018). Although academic literacy development of adolescent multilingual students in the United States has been gaining attention among scholars at the intersection of language and education, there is a paucity of studies that systematically examine language features of academic texts produced by adolescent multilingual students in a secondary education context. Much of the research in L2 writing, more specifically on the linguistic features of multilingual students' writing, has focused on postsecondary students (Ortega & Carson, 2010). Based on the research review, Fitzgerald (2017) pointed out that research involving U.S. adolescent multilingual students' compositions remains thin as little attention has been devoted to this area and student population. This gap in the research, thus, highlights a need to investigate how adolescent multilingual students use language in their academic writing, what linguistic needs they have, and what challenges they may encounter when writing within an academic context. Such investigation is crucial in addressing the glaring disparity in the academic writing and language development among students in a secondary education context.

Alongside the critical need for research to address the opportunity gaps in academic literacy for this population, there also remain gaps in pedagogy to address their linguistic needs. Writing instruction in secondary schools is not adequate to meet the high demands of academic

writing (Applebee & Langer, 2011; Graham et al., 2014). Applebee & Langer (2011) surveyed teachers across the United States and observed how writing is taught in 260 middle and high school classes in five different states. Their study found that writing practice and instruction given to middle school students were minimal, with only 7.7% of class time spent on composing extended texts. Similarly, Graham et al. (2014) reported that the amount of writing practice students engage in and the language instruction they receive are not sufficient in a typical middle school classroom. This is associated with the fact that many of the teachers surveyed did not receive adequate training to teach writing, especially to multilingual students who haven't developed proficiency in English. The study concluded that there is a discrepancy between writing practices in secondary school and writing standards as defined by the CCSS (Graham et al., 2014). Both of these studies suggest that students in secondary school, and multilingual students who are learning English specifically, are not receiving sufficient writing practice and are not exposed to high-quality writing and language instruction, in general. Thus, the low performance of students in middle and high school, particularly of English learners, on NAEP assessment is not surprising.

For adolescent multilingual students developing their English proficiency, engaging in academic writing in English is a challenging and cognitively demanding task. Effective written communication in an academic context requires an in-depth understanding of genre-conventions, domain knowledge, and linguistic principles. It demands a nuanced skill to synthesize concrete details and abstract concepts, as well as an ability to present ideas in a clear and coherent way, using language appropriate for an academic audience (Defazio et al., 2010). When composing analytical essays in an academic context, students often need to move beyond summary or what Bereiter and Scardamalia (1987) refer to as "knowledge telling" and compose at an advanced,

"knowledge-transforming" level that includes analysis and requires higher-order thinking skills. These higher-order writing tasks demand higher-level linguistic competence, including syntactic fluency and lexical capacity. In short, the skill to write well in an academic context demands mastery of academic English and requires students to develop an in-depth understanding of the specialized lexical, syntactic, and discursive features of academic texts and discourse. However, developing multilingual writers may lack the linguistic competence and language proficiency necessary to meet the demands of academic writing (Harklau & Pinnow, 2009; Scarcella, 2003). In order to address the academic literacy of adolescent multilingual students, it is crucial to understand how these students use language in their academic writing and what linguistic needs they have in order to be able to write successful academic essays.

Overview of the Dissertation

In this dissertation, I analyze text-based analytical essays written by linguistically diverse adolescent students attending public middle and high schools in the United States. Text-based analytical writing is a school based, academic writing genre that requires students to construct and develop an argument of literary analysis. In systematically analyzing students' academic texts, I focus on the linguistic features of the texts and their relations to the writing quality and rhetorical features. The principal goals of the dissertation research are: 1) to identify language use and linguistic patterns exhibited in linguistically diverse adolescent students' text-based analytical essays; and 2) investigate whether and how the identified linguistic features relate to human-judged writing quality and rhetorical features. Studying the language features and linguistic characteristics of adolescent writers' texts that are produced as part of a school-based, academic genre and their relation to rhetorical features and writing quality will provide insight into how they use language when composing in an academic context. It can also elucidate our

understanding of the linguistic needs of these students as they strive to meet the demands of academic writing. Having a better understanding of these characteristics can inform interventions and curricular approaches that aim to address the academic literacy of linguistically diverse adolescent students in a secondary school context.

The dissertation consists of three studies that each address specific linguistic features of students' text-based analytical writing and their relations to writing quality. These three studies are related and expand on each other to provide an in-depth understanding of the linguistic characteristics observed in students' academic texts and the extent to which these features contribute to the quality of their writing. The foci of the studies move from a specific linguistic component to more comprehensive linguistic and rhetorical features of students' text-based analytical writing. Study 1 focuses exclusively on lexical features and their relation to writing quality, and its findings inform the selection of lexical measures in the consecutive studies. Expanding on the first study, I add syntactic features to the select lexical features and examine how syntactic and lexical features independently and together contribute to writing quality. Study 3 takes a more holistic approach and examines lexical, syntactic, and rhetorical features of student texts and how these features relate to one another and to writing quality. The first two studies inform the selection of lexical and syntactic variables in study 3. In what follows, I provide a brief overview of these three studies.

The first study focuses on lexical features and academic vocabulary use in a sample of the text-based analytical essays (n=70) written by Spanish-speaking multilingual students in grades 7-12 from a large urban school district in Southern California. I specifically investigated lexical density, lexical diversity, and lexical sophistication, measured by computational Natural Language Processing (NLP) tools Coh-Metrix (McNamara et al., 2014) and VocabProfilers

(Cobb, 2018). I further examined how these three lexical features predict human-judged writing quality and the extent to which students in the sample used academic vocabulary in their literary analysis. I employed descriptive statistics to analyze the degree of lexical density, diversity, and sophistication in the learner corpus under investigation. I then used Pearson correlation and multiple regression analyses to examine the extent to which the three lexical features predict the holistic writing quality score. The results of this study indicated that the sample of text-based analytical essays written by Spanish-speaking multilingual students had low lexical density and diversity and exhibited reliance on basic and frequently used words. While lexical density and lexical diversity did not predict writing quality, lexical sophistication was a predictor of the holistic writing score. An important finding of the study was that many of the academic words students used in their writing came from the source text and prompt, which points to the value of using academic texts and genres to scaffold multilingual students' vocabulary development.

The second study systematically examines lexical and syntactic features and their relations to writing quality using the same learner corpus, but with a slightly larger sample (n=86). Employing manual sentence coding and quantitative measures of selected syntactic features from a NLP tool Coh-Metrix, I analyzed lexical and syntactic patterns focusing on lexical diversity and frequency, psycholinguistic properties of words, sentence boundary issues (run-on sentences, fragments, and faulty sentences), syntactic variety and complexity (simple, compound, complex, and compound-complex), clausal density and types (finite and nonfinite clauses, subordinated and coordinated clauses), and the connectives used to combine clauses. Using descriptive statistics, I identified common lexical and syntactic features and patterns in the sample of text-based analytical essays written by Spanish-speaking multilingual students. I then conducted Pearson correlation and multiple regression analyses to examine the extent to which

the syntactic and lexical features separately and together predict writing quality. The results of the study demonstrated that unconventional sentences and basic coordination and subordination were the common syntactic patterns. Syntactic features explained 44% of the variance in writing quality while lexical features explained 23% of the variance in writing quality. The finding that both syntactic and lexical features were crucial indicators of writing quality attest to the need for adolescent multilingual students to develop their language resources to engage in complex academic writing tasks.

In the third study, I analyze lexical, syntactic, and rhetorical features to identify common patterns and features, examine how the identified features vary among three different L1 groups, and investigate the relations among the lexical, syntactic, and rhetorical features and writing quality. This study used a much larger and more representative learner corpus of text-based analytical essays written by linguistically diverse students in grades 7-11 (n=340) from public school districts in 5 different states (AZ, MN, TX, UT, and WI). In this study, I used the select lexical and syntactic features based on the results of study 1 and 2 of the dissertation, as well as the findings of extant research and the relevant theorization in literature. I added rhetorical features, including measures of text cohesion, source-use and integration, and balance of summary, evidence, and commentary. For measures of lexical, syntactic, and rhetorical features, I utilized a combination of manual coding, human-judged analytic scoring, and three NLP tools—Tool for the Automatic Analysis of Cohesion (TAACO; Crossley et al., 2016), Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Kyle & Crossley, 2015), and Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016). Using descriptive statistics, I first identified common linguistic and rhetorical patterns of texts written by culturally and linguistically diverse adolescent writers. I then conducted multiple

analysis of variance (MANOVA) to examine whether the identified lexical, syntactic, and rhetorical features varied among three different L1 groups. To investigate the relations among the linguistic and rhetorical features and writing quality, I used structural equation modeling (SEM) that includes confirmatory factor analysis (CFA) and structural regression. For SEM analyses, I added the sample of 70 essays written by Spanish-speaking multilingual students from previous studies. Thus, the sample used for CFA and structural regression is 410 text-based analytical essays written by linguistically diverse students in 6 states (CA, AZ, MN, TX, UT, WI). The results of the study indicated that student texts, in general, had low use of advanced and academic words and less syntactic variety at both clausal and phrasal levels. Summary and summary-dominant structures were common rhetorical patterns. No significant differences were found in rhetorical features among the three different L1 groups. However, there was some variability in the lexical and syntactic features among students with different L1. Based on CFA results, the three-factor model in which the lexical, syntactic, and rhetorical features are distinct but related latent constructs was the best-fitting model and thus was chosen as the final model used for structural regression. The latent construct of rhetorical features was a strong predictor of writing quality. While the latent construct of lexical and syntactic features did not significantly predict writing quality, they were strongly correlated with rhetorical features, indicating shared variance.

Significance of the Dissertation Research

Together, these three studies help to paint a descriptive picture of middle- and high-school students' academic English use and academic writing proficiency within the genre of text-based analytical writing. Gaining a deeper understanding of these characteristics has important implications for writing research and pedagogy. The findings of this study have the potential to

inform researchers and practitioners of the linguistic needs of diverse adolescent multilingual students and the language skills they need to meet the demands of academic writing. This is particularly important since there is a dearth of research that focuses on the language use and linguistic features of adolescent multilingual students' academic writing despite it being a crucial skill for college and career readiness.

Much of the research exploring the linguistic features of student writing and text analysis of learner corpora have focused on student writing at the collegiate and pre-collegiate levels in a postsecondary context. As mentioned earlier, adolescent multilingual students in the secondary education setting, despite their growing rate, are an underexplored group in the field of L2 writing (Yi et al., 2018) and have remained "outside the purview" of L2 writing research (Ortmeier-Hooper & Enright, 2011, p. 167). In addition, most of the extant studies that examine linguistic features and their relation to writing quality have focused on independent writing tasks, as opposed to source- or text-based writing tasks. Examining linguistic features in a variety of written texts is important as different writing tasks (e.g., independent vs. source-based), modes (e.g., narrative, expository, argumentative) and genres (memoir, literary analysis, opinion editorial) elicit different linguistic features and, thus, may demand different linguistic skills (Crossley, 2020; Plakans & Gebril, 2013). To my knowledge, no study has examined the linguistic features of text-based analytical writing and the complex relationship between linguistic elements, rhetorical components, and writing quality. Collectively, the three studies of the dissertation address the gaps in research by examining the linguistic features in a lesser studied yet important type of academic writing produced by an underexplored student population.

Importantly, the intent of this dissertation research is to inform writing pedagogy in the secondary education context. Scarcella (2003) notes, "By the time children reach high school, academic tasks require students to have acquired an extensive range of competencies" as school tasks become more complex and academic in nature. (p. 8). However, for many multilingual students across different school districts and states, school is the only place where they are exposed to academic language. Thus, instructional practice that provides opportunities for developing multilingual writers to build their language skills and expand their linguistic repertoires is consequential in the U.S. education context in which many multilingual students are considered as "struggling" writers. In order to develop effective academic language instruction and integrate it into academic writing pedagogy and curriculum, it is pivotal to: 1) understand the language use in linguistically diverse adolescent students' academic writing; and 2) examine how the linguistic patterns and features of their writing relate to rhetorical features and human-judged holistic writing quality. In light of our collective effort to support diverse multilingual students' academic writing and language development, this dissertation research makes an important contribution to the writing pedagogy geared toward meeting the linguistic needs of the unique student population with differing exposure to and experiences of navigating the complex landscape of US secondary education.

References

Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal*, 100, 14–27.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates.

Bunch, G. C., Kibler, A. K., & Pimentel, S. (2012). *Realizing opportunities for English learners in the common core English language arts and disciplinary literacy standards*. Paper presented at the Understanding Language Conference, Stanford, CA. Retrieved from http://ell.stanford. edu/papers.

Bowen, W. G., Kurzwell, M. A., & Tobin, E. M. (2005). *Equity and excellence in American higher education*. University of Virginia Press.

Carnegie Mellon University, Eberly Center, Teaching Excellence & Educational Innovation. (2021). Why are students coming into college poorly prepared to write? Retrieved from https://www.cmu.edu/teaching/designteach/teach/instructionalstrategies/writing/poorlyprepared.html.

Cobb, T. (2018). *Compleat lexical tutor.* Retrieved from http://www.lextutor.ca.

Council of Writing Program Administrators., National Council of Teachers of English., & National Writing Project. (2011). *Framework for success in postsecondary writing.* Creative Commons. Available at http://wpacouncil.org/aws/CWPA/asset_manager/get_file/350201?ver=7548.

Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415–443.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods,* 48(4), 1227–1237.

Cumming, A. (2016). Writing development and instruction for English language learners. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 364–376). The Guilford Press.

Defazio, J., Jones, J., Tennant, F., Hook, S.A. (2010). Academic literacy: The importance and impact of writing across the curriculum—a case study. *Journal of the Scholarship of Teaching and Learning,* 10, 34–47.

Fitzgerald, J. (2017). How practice-based research informs adolescent English-language learners' composing and compositions. In K. A. Hinchman & D. A. Appleman (Eds.). *Adolescent literacies: A handbook of practice-based research.* (pp. 357-378). Guilford Press.

Francis, D. J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Research-based recommendations for instruction and academic interventions.* Center on Instruction.

Goldenberg, C. (2013). Unlocking the research on English learners: What we know—And don't yet know—About effective instruction. *American Educator,* 37, 4–11.

Graham, S., Capizzi, A., Harris, K. R., Heber, M., & Morphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading and Writing,* 27, 1015–1042.

Graham, S. & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools—A report to Carnegie Corporation of New York.* Alliance for Excellent Education.

Harklau, L., Losey, K. M., & Siegal, M. (Eds.). (1999). *Generation 1.5 meets college composition: Issues in the teaching of writing to U. S.-educated learners of ESL*. Erlbaum.

Harklau, L., & Pinnow, R. (2009). Adolescent second-language writing. In L. Christenbury, R. Bomer., & P. Smagorinsky (Eds.). *Handbook of adolescent literacy research.* Guilford Press.

Kanno, Y., & Cromley, J. G. (2015). English language learners' pathways to four-year colleges. *Teachers College Record*, 117, 1–44.

Kim, E., & Díaz, J. (2013). Immigrant students and higher education. *ASHE Higher Education Report*, 38(6), 1–170.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Doctoral Dissertation). Retrieved from http://scholarworks.gsu.edu/alesl_diss/35.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757–786.

Matsuda, P. K., Ortmeier-Hooper, C., & You, X. (Eds.). (2006). *The politics of second language writing: In search of the promised land*. Parlor Press.

McNamara, D. S., Graesser, A. C. McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

National Commission on Writing. (2003). *The neglected "R": The need for a writing revolution.* College Board.

National Governors Association Center for Best Practices & Council of State School Officers. (2010). *Common core state standards for English language arts and literacy in history/social studies, science, and technical subjects.* Washington, DC: Authors.

Next Generation Science Standards. (2013). Fact sheet. Retrieved from https://www.nextgenscience.org.

Núñez, AM., Rios-Aguilar C., Kanno, Y., & Flores S. M. (2016). English learners and their transition to postsecondary education. In M. Paulsen (Eds.) *Higher education handbook of theory and research.* Springer.

Olson, C. B., Matuchniak, T., Chung, H. Q., Stumpf, R., & Farkas, G. (2017). Reducing achievement gaps in academic writing for Latinos and English learners in Grades 7–12. *Journal of Educational Psychology,* 109, 1–21.

Olson, C. B., Scarcella, R., & Matuchniak, T. (2015). *Helping English learners to write: Meeting common core standard, grades 6-12.* Teachers College Press.

Ortega, L., & Carson, J. (2010). Multicompetence, social context, and L2 writing research praxis. In T. Silva & P.K. Matsuda (Eds.). *Practicing theory in second language writing*. (pp. 48-71). Parlor Press.

Ortmeier-Hooper, C., & Enright, K. A. (2011). Mapping new territory: Toward an understanding of adolescent L2 writers and writing in US contexts. *Journal of Second Language Writing,* 20, 167–181.

Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing,* 22, 217–230.

Scarcella, R. (2003). Academic English: A conceptual framework. *The University of California Linguistic Minority Research Institute Technical Report 2003-1,* Retrieved from https://escholarship.org/uc/item/6pd082d4.

U.S. Department of Education. National Center for Education Statistics. (2020). English language learners in public schools. Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2012). *The nation's report card: Writing 2011* (NCES 2012-470). Washington, DC.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2019). *Technical Summary of Preliminary Analyses of NAEP 2017 Writing Assessments.* Available at https://nces.ed.gov/nationsreportcard/writing/2017writing.aspx.

Yi, Y., & Kao, C., & Kang, J. (2018). Middle and high school writing. In J. I. Liontas (Ed.). *The TESOL Encyclopedia of English Language Teaching.* John Wiley & Sons, Inc.

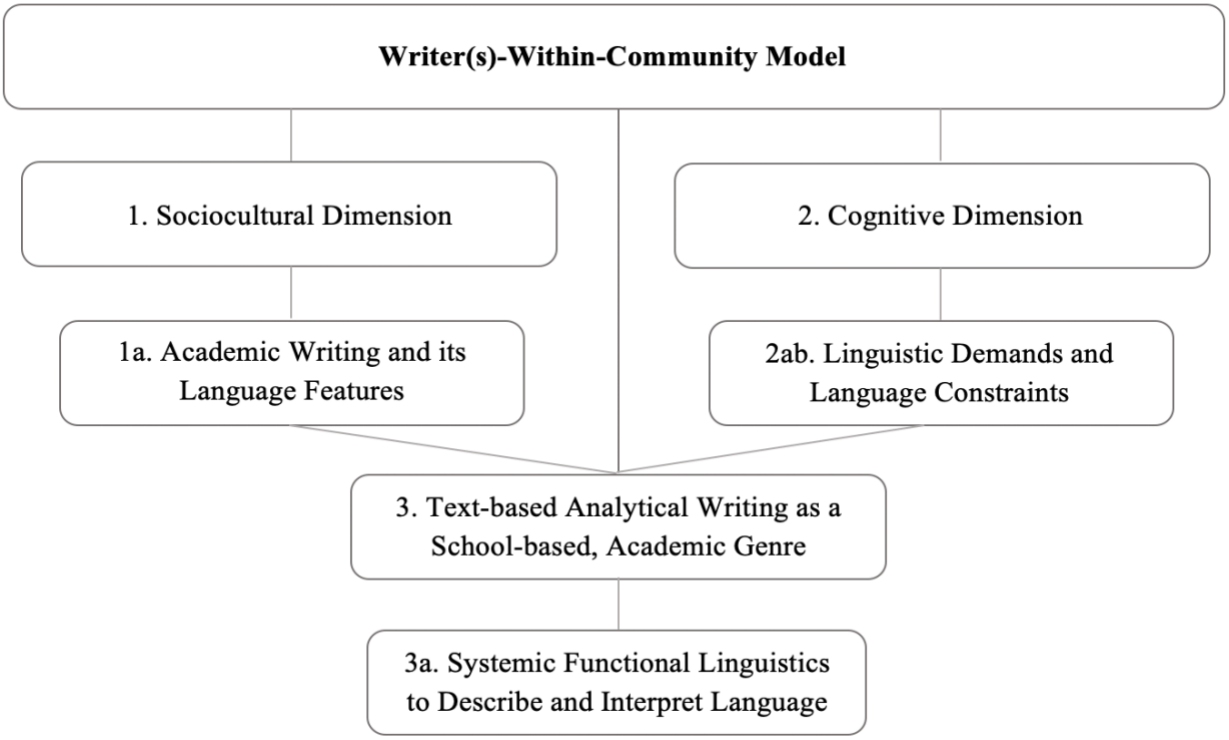
CHAPTER 2

Conceptual Framework

A broad, all-encompassing framework that informs this dissertation research is the Writer(s)-Within-Community (WWC) model of writing proposed by Graham (2018a, 2018b). The WWC model (see Appendix A) unifies sociocultural and cognitive perspectives and situates the act and activity of writing within communities that are shaped by social, cultural, historical, institutional, and political factors. As Graham (2018b) points out, "The model proposes that writing is simultaneously shaped and bound by the characteristics, capacity, and variability of the communities in which it takes place and by the cognitive characteristics, capacity, and individual differences of those who produce it" (p. 258). The graphic representation presented in Figure 1 below outlines the concepts that I draw on and the structure that organizes this section.

Figure 2.1

Graphic Representation of the Conceptual Framework



Within the WWC model, I first discuss the sociocultural dimension of writing, highlighting the situatedness of academic writing in a social context and examining language features of academic written discourse that are shaped by shared purposes and audience expectations within the academic writing community. I then explore the cognitive dimension, focusing on the processing demands of writing to explain the language resources students draw on and the linguistic constraints they may encounter when composing. Finally, I zero in on the text-based analytical essays (the unit of analysis in this dissertation research) as a school-based, academic writing genre. In doing so, I draw on the functional-semantic approach to language grounded in Halliday's (1994) systemic functional linguistics as a systematic strategy to describe and interpret language use as a meaning-making resource.

1. Sociocultural Dimension

Writing is a complex task and a generative social activity, and writing development entails a multidimensional process involving interaction of social, cultural, historical, and cognitive dimensions. From the vantage point of the sociocultural perspective, writing is framed as a dynamic social participatory performance (Bazerman, 2016) and a socially situated activity whose object is socially productive practice (Engeström, 2015). The sociocultural dimension of writing "foregrounds the social practices student writers acquire through learning to adopt and adapt in constructing social and cultural contexts for their writing" (Beach et al., 2016, p. 90). Thus, students' participation in the activity of writing is situated in and shaped by social circumstances.

Writers develop their writing skills by engaging in socially situated activities within communities of practice that are influenced by social, cultural, and historical forces (Graham, 2018b; Lave & Wenger, 1991). Participation in a particular community of practice and

disciplinary culture requires writers to acquire knowledge of norms and expectations, become articulate in the discursive practice of the community, figure out how meaning is constructed and communicated within this particular rhetorical context, and learn to use and negotiate the resources that are available within the community. Thus, an academic writing task is a socially situated activity that fosters "enculturation and socialization into the norms, practices, and action goals of a community" for successful communication (Bazerman, 2016, p. 16). By actively participating in different types of academic reading and writing tasks, writers acquire knowledge of the specialized language features, develop understanding of genre conventions, and build awareness of discursive practices.

1a. Academic Writing and its Language Features

Writing in an academic context is guided by and aligned with communicative purposes, established norms, and language conventions sanctioned by the members of particular discourse communities (Swales, 1990). Regardless of the discipline in which it is situated, academic writing is marked by a well-informed and focused topic/argument, evidence-based development, textual coherence based on logical order and relations between ideas, and formal tone. The expression and the development of complex ideas in an academic context exhibit common features because of the similarity in purposes of academic genres (Schleppegrell, 2001). Academic writing is "characterized by precision, and economy of expression, variety in sentence structure, support for claims, conceptual/abstract treatment of topics, and adherence to the expectations of specific types of writing such as arguments" (Olson et al., 2015, p. 13).

Academic written register, which is characterized by its highly informational purpose, impersonal style and absence of narrative features, exhibits salient patterns of language use and grammatical complexity that need to be acquired by writers composing within an academic

context (Biber, 2006; Biber et al., 2011). The highly specialized language features of academic written discourse "create the qualities of precision, detachment, and density that fit the context of academic tasks" (Ranney, 2012, p. 565). To accomplish higher-order analytical tasks, academic writers heavily rely on their linguistic resources; namely, syntax and vocabulary are two fundamental linguistic resources used to construct meaning and communicate ideas. Further, the lexical features and syntactic forms that are common in academic texts are associated with their communicative functions (Halliday, 1994). In essence, academic written genres exhibit common syntactic and lexical features, and understanding "the linguistic elements that make up the [formal] registers of schooling" is crucial to students' academic writing development (Schleppegrell, 2001, p. 431).

Syntactic Features of Academic Writing. Grammatical structures of academic writing tend to be highly specialized (Biber et al., 2011). One prominent syntactic and stylistic feature that is common across various academic written genres is the application of and dependence on hypotactic syntactic constructions that are more sophisticated and less common than those found in casual conversations. For example, Biber et al. (2011) found that nonfinite relative clauses, WH relative clauses (post-modifiers using *who, that, which, where, when)*, noun complement clauses with *that* and *to,* and "dependent clauses functioning as constituents in a noun phrase" are much more prevalent in academic writing. On the other hand, finite dependent clauses with *if* and *because* conjunctions and finite complement clauses (verb + *that* clause as in "I think that ...") are more common in conversations than academic writing. Sophisticated forms of hypotaxis or subordination of clauses are common in academic writing for the dual purpose of achieving efficiency/economy and establishing a hierarchical order of ideas. Syntactic efficiency is achieved through embedding a subordinate idea within the main sentence rather than expressing

it in a separate, independent sentence. Hypotactic syntactic construction also makes the logical relations and interdependence of ideas explicit, showing arrangements of clauses and phrases that are subordinate to and dependent on main clauses and phrases.

Other specific syntactic features of academic writing include embedded phrases (participial and absolute phrases, and apposition embedded within sentences), complex phrasal structures (noun and prepositional phrases), and hierarchical structure (phrase and clause subordination). For instance, post-modifying prepositional phrases, pre-modifying nouns, and attributive adjectives in noun phrases are particularly common in academic writing (Biber et al., 2011). The sentences below, which are extracted from the newspaper articles that students read and responded to in an on-demand pretest for a large randomized field trial from which the writing samples in this dissertation were obtained and analyzed, exhibit some of the clause embedding and other hierarchical syntactic features mentioned above.

Excerpt 1: From "The Man in the Water" by Roger Rosenblatt

Lenny Skutnik, *a 28-year-old employee of the Congressional Budget Office*, said: "It's something I never thought I would do" – *referring to his jumping in the water to drag an injured woman to shore.*

- *[a 28-year-old employee of the Congressional Budget Office]* — apposition embedded in a subject-verb split position
- *[referring to his jumping in the water to drag an injured woman to shore]* — present participial phrase linked to the subject of the main clause

Excerpt 2: From "Sometimes the Earth is Cruel" by Leonard Pitts Jr.

That is ultimately the fundamental lesson here, <u>as children wail, families sleep out of doors, and the dead lie unclaimed in the rubble</u> <u>that once was Port-au-Prince</u>.

- *[as children wail, families sleep out of doors, and the dead lie unclaimed in the rubble]* — subordinated, adverbial clause with a connective *as*
- *[that once was Port-au-Prince]*— relative clause with relative pronoun *that*

In the sentences above, the subordinate clauses (underlined) make explicit the logical connections between ideas and show the hierarchical structure of clauses. In addition, the use of apposition and participial phrases (italicized) embedded in the sentences is a way to efficiently combine related ideas. In general, academic writing is characterized by complex grammatical structures, interdependent ideas expressed in hierarchical structures, and embedding and nesting of clauses and phrases. These clause-combining strategies are used by academic writers to achieve efficiency and to build logico-semantic relationships.

Lexical Features of Academic Writing. Academic writing across various genres and disciplines also shares common lexical features. A linguistic feature that is conspicuous in academic writing is "the vocabulary of academic language, often described as comparatively large, precise, and formal" (Ranney, 2012, p. 563). Comparing the lexical features in spoken interactions and school-based texts, Schleppegrell (2001) notes that words used in academic writing tend to be specific and technical, and academic texts have higher lexical density. The density, precision, and technicality of academic lexicons are associated with informational density, technical concepts, and content-specific words that are required to convey complex ideas.

Corpus-based research investigating the lexical features of academic writing has led to the development of vocabulary typologies. For example, Scarcella (2003) identified three major categories of academic vocabulary: (1) general words used across academic disciplines and in everyday situations, (2) academic words that are common across different disciplines, and (3)

technical words found in specific academic fields. The words in the latter two categories tend to be highly specialized. In order to effectively communicate in an academic context, writers use words in all three categories to express their ideas clearly, avoiding ambiguity and vagueness. The communicative purpose and the clarity of expressions in an academic text are realized through appropriate and precise lexical choices that draw on a large repertoire of words and lexical resources.

Taken together, understanding the common linguistic features of academic writing is crucial in writing instruction that is geared towards developing diverse multilingual students' academic language and writing proficiency. In a comprehensive review of the literature on Academic English (AE) in K-12 settings, Anstrom and colleagues (2010) call for a framework of AE to guide instructional practice, professional development, and assessment decisions after concluding that there is lack of clarity on academic English and how it is conceptualized. In short, to become effective academic writers, students need to not only develop knowledge, awareness, and acquisition of the specialized language features of academic writing but also understand the discursive norms of academic written genres that are shaped by social, cultural, and historical forces.

2. Cognitive Dimension

Although writing is a social act with a communicative purpose, it is also a cognitively demanding task bound by cognitive characteristics, processing capacity, and experiential differences of the individuals who produce it (Graham, 2018a; 2018b). A writing task, whether it involves crafting an initial draft or revising a subsequent draft, presents a set of demands, and writers often have to juggle between a variety of mental activities (Olson et al., 2018). Flower and Hayes (1980) contend, "As a dynamic process, writing is the act of dealing with an excessive

number of simultaneous demands or constraints…, [and] a writer in the act is a thinker on full-time cognitive overload" (p. 33). In their initial drafting phase, writers move back and forth between two major cognitively demanding processes: 1) the ideation process which involves sub-processes of generating new ideas and organizing those ideas; and 2) the transcription process which includes translating ideas into words and sentences to form coherent paragraphs (Flower & Hayes, 1981; Torrance & Galbraith, 2006). Both ideation and transcription place heavy demands on cognitive resources (Flower & Hayes, 1981). When students attempt to perform these two major cognitive tasks simultaneously while lacking adequate linguistic knowledge, diminished performance of one or both tasks is inevitable (Torrance & Galbraith, 2006). When linguistic resources are limited, writing is a laborious and cognitively challenging task. Thus, knowledge of language is a necessary long-term memory (LTM) resource that writers draw on in order to effectively communicate their ideas.

2a. Linguistic demands of Academic Writing

Academic writing requires the ability to analyze, interpret, synthesize, and explain complex ideas. Performing these higher-order analytical tasks successfully depends on linguistic resources and requires a higher level of language proficiency. Kellogg (2008) asserts, "Without knowledge [of language and discourse] being accessible and creatively applied by the writer, it remains inert during composition and unable to yield the desired fluency and quality of writing" (p. 3). This is especially true at an advanced level of writing that is beyond what Bereiter and Scardamalia (1987) referred to as "knowledge telling" which relies primarily on retelling or summary. Thus, writing at the level of "knowledge transformation" (Bereiter & Scardamalia, 1987) that is characterized by complex rhetorical problem-solving processes demands advanced

language skills and a mastery of comprehensive language knowledge (Bereiter & Scardamalia, 1987).

To communicate well in an academic context, students need to have comprehensive knowledge of sentence structures, types, and functions. This includes the ability to coordinate and subordinate sentences, as well as the ability to make a variety of syntactic choices appropriate for the genre and audience. Further, syntactic knowledge is necessary to avoid sentence boundary errors. The National Writing Project's (2010) Analytic Writing Continuum, which is used as a rubric for literary analysis, requires proficient writers to: (1) "demonstrate effective phrasing so that each sentence flows easily into the next"; (2) "include sentences that vary in structure and length, creating an extremely effective text"; and (3) construct sentences that are "consistently logical and clear so that the relationships among ideas are firmly and smoothly established" (p. 2). Because academic writing demands complexity and variety in sentence structure, as well as conventionalized clause organization, developing writers need to learn the formalized syntactic structure (Schleppegrell, 2001).

Academic writing also demands a large lexical capacity and an in-depth knowledge of words, including their morphological, semantic, and pragmatic properties. While knowing more words is important, knowledge about the multi-dimensionality of vocabulary is vital in learning to write in an academic context (Kieffer & Lesaux, 2012). In addition, students need to learn to make appropriate lexical choices considering the academic audience and genre. For students, then, writing in an academic context requires an ability to differentiate between the distinct genre conventions and to use "discipline-specific rhetorical and linguistic conventions to serve their purposes as writers" (Berkenkotter et al., 1991, p. 191). Given these points, an assumption can be made that without knowledge of the specialized linguistic features of academic written discourse,

formal register, and discursive conventions, developing writers are constrained as they engage in complex writing tasks that require higher-order analytical skills.

2b. Language Constraints of Developing Multilingual Writers

For many multilingual writers, developing proficiency in academic writing is constrained by various factors affecting language development, including their exposure to academic language and formal instruction, previous development of linguistic knowledge, and the extent to which they pay attention to words, phrases, and sentence structures (Olson et al., 2015). Generating an extensive text at an advanced and academic level involves the "language system" and "cognitive systems for memory and thinking" (Kellogg, 2008, p. 2). However, students who haven't developed academic English proficiency are often "cognitively overloaded," facing the "dual challenge of learning how to write while at the same time they are still developing proficiency in the English language" (Olson et al., 2015, p. 9). In other words, the cognitive overload that results from juggling simultaneous demands is magnified for developing multilingual writers because of the linguistic constraints they may experience.

As postulated by the capacity theory of writing, composing is a complex and dynamic task that exerts considerable processing demands on cognition and working memory (McCutchen, 1996). For developing multilingual writers, who may prioritize simpler cognitive tasks over more demanding cognitive activities, transcribing complex ideas into words and sentences to form coherent writing is a challenging task, especially at the advanced levels of "knowledge transformation" (Bereiter & Scardamalia, 1987; Kellogg, 2008) and "knowledge crafting," which is characterized by the complex interaction of the author, text, and the reader (Kellogg, 2008). Multilingual students' cognitive load is reduced when the linguistic resources they draw on are readily available and easily accessible. Thus, developing proficiency in the

language of academic written discourse is integral in composing at an advanced level of knowledge transformation. In support of this, Kellogg (2008) contends that writers are unable to reach "the desired quality of writing" if different types of knowledge related to language use, text content, and discourse are unavailable and inaccessible for retrieval from long-term memory (p. 3). Thus, having a rich and advanced vocabulary, comprehensive knowledge of sentence structures, and an in-depth understanding of academic registers are needed to successfully engage in academic discourse.

3. Text-based Analytical Writing as a School-based, Academic Genre

The studies in this dissertation focus on text-based analytical essays written by linguistically diverse adolescent students in a context of formal schooling. To complete this writing task, students engage in reading a nonfiction article, analyzing its theme, message, purpose, and author's craft, and writing an interpretive, analytical essay in which they make a claim about the theme and develop their claim with evidence from the article. As the task prompts students to construct an argument of interpretive analysis and to use the source text for evidence, this type of writing can be considered as a subgenre of a broader source-based argumentative writing that is emphasized in the Common Core State Standards and that has gained prominence in school-based writing. However, unlike typical source-based writing that often requires integration of multiple sources, this type of writing task draws on a single source and requires an interpretive analysis of the source text in a form of literary explication. This text-based analytical writing task is aligned with the CCSS's emphasis on developing the ability to analyze, argue, interpret, and use evidence to support ideas about the meaning and message of a literary work (CCSS, Appendix C). It is also similar to the AP Literature Assessment that many students who strive for college entrance may take. Because of its emphasis on analytical and

academic writing skills necessary for college and career success, this type of writing can be considered as a school-based, academic writing genre. Identifying the genre of a text is vital in a text analysis since genre reflects "the cultural purpose" of the text (Eggins, 2004, p. 54). Thus, in describing and interpreting the rhetorical features and the language use in texts, the genre they represent and the context within which they are produced need to be regarded (Halliday & Matthiessen, 2014).

Research on writing indicates that the rhetorical and linguistic features of texts vary greatly by the type of writing task, purpose, and genre (Bouwer, et al., 2015; Crossley, 2020). Unlike independent or non-source-based writing tasks, including narrative and expository writing, text-based analytical writing can be more complex since it requires writers to analyze and interpret a text, make a claim, integrate textual evidence, and provide commentary on the evidence selected to advance the claim (Olson et al., 2020). Successful completion of these higher-order analytical tasks heavily depends on the writers' linguistic resources. In addition, as the text-based analytical writing is a school-based, academic writing genre, students are expected to make syntactic and lexical choices that are appropriate for an academic audience, adhering to the language conventions of academic written discourse.

3a. Systemic Functional Linguistics to Describe and Interpret Language

In analyzing the text-based analytical essays produced by adolescent multilingual writers, I draw on the theory of Systemic Functional Linguistics (SFL) as a functional-semantic approach to describe and interpret language and its use within a specific context. SFL, initially developed by Halliday (1973), is often used as an analytical approach of text linguistics and deals with language use and patterns within a text in relation to the contexts in which it appears. The SFL approach to text and discourse analysis is to determine "the functions of language and its features

in context, rather than from the abstract syntactic categories of language" (Hinkel, 2002, p. 16). Halliday (1978) contends that certain patterns of language use within a specific discursive context can be detected through a systemic approach and functional analysis of text construction. According to Eggins (2004), SFL is "increasingly recognized as a very useful descriptive and interpretive framework for viewing language as a strategic, meaning making resource" (p. 2). SFL deals with two important questions in regard to functions of language, with particular attention given to the context in which language is used: "How do people use language?" and "How is language structured for use as a [semiotic system]?" (Eggins, 2004, p. 3). In an effort to answer these questions, the SFL approach uses text as its primary unit of analysis (Coffin & Donohue, 2012) and aims to describe and explain language use and patterns in relation to the concomitant contextual and social dimensions.

The systemic approach and functional grammar that deal with structural features related to English phrases, clauses, and sentences serve as a useful tool and an analytical strategy that guides my research aim to examine the language use and detect language patterns in students' written texts. Since the systemic functional approach to text and language analysis is concerned with "how grammatical and lexical choices a writer makes are related to the context of situation and culture in which the writer produces text" (Beck, 2009, p. 316), the language patterns observed in students' texts should be interpreted and understood in relation to the text genre, in this case the text-based analytical essays that are produced as part of a formal, school-based writing activity. Corpus-based research shows that academic writing genres that require advanced analytical skills are characterized by complex grammatical structures (Beck, 2009); corpus analysis of student texts "can provide quantitative descriptions of children's linguistic productions, which is one type of behavior that can indicate the underlying grammar" (Pearl,

2021, p. 7). Thus, analyzing the grammatical structures and lexical use in student texts is an important step in the direction of understanding students' linguistic resources and the extent to which they reflect the language conventions of academic writing. To this end, the three studies of the dissertation systematically examine the lexical and syntactic features and productive language use in a learner corpus of text-based analytical writing.

References

Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., Rivera, C. (2010). *A review of the literature on academic English: Implications for K-12 English language learners.* The George Washington University Center for Equity and Excellence in Education.

Bazerman, C. (2016). What do sociocultural studies of writing tell us about learning to write? In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 11–23). Guilford Press.

Beach, R., Newell, G. E., & VanDerHeide, J. (2016). A sociocultural perspective of writing development: Toward an agenda for classroom research on students' use of social practices. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 11–23). Guilford Press.

Beck, S. W. (2009). Composition across secondary and postsecondary contexts: Cognitive, textual and social dimensions. *Cambridge Journal of Education*, 39, 311–327.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates.

Berkenkotter, C., Huckin, T., & Ackerman, J. (1991). Social context and socially constructed texts: The initiation of a graduate student into a writing research community. In C. Bazerman & J. Paradis (Eds.). *Textual dynamics of the professions* (pp. 191–215). University of Wisconsin Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* John Benjamins.

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

Bouwer, R., Béguin, A., Sanders. T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing,* 32, 83–100.

Coffin, C., & Donohue, J. (2012). Academic literacies and systemic functional linguistics: How do they relate? *Journal of English for Academic Purposes,* 11, 64–75.

Common Core State Standards Initiatives. (2021). *English Language Arts Standards: College and Career Readiness Anchor Standards for Writing.* Retrieved from http://www.corestandards.org/ELA-Literacy/CCRA/W/.

Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research,* 11(3), 415–443.

Eggins, S. (2004). *An introduction to systemic functional linguistics (2nd ed.).* Bloomsbury Academic.

Engeström, Y. (2015). *Learning by expanding: An activity-theoretical approach to developmental research.* (2nd ed.). Cambridge University Press.

Flower, L. S., & Hayes, J. R. (1980). The dynamic of composing: Making plans and juggling constraints. In L.W. Gregg & E.R. Steinberg (Eds.). *Cognitive processes in writing* (pp. 31–50). Lawrence Erlbaum Associates.

Flower, L. S., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication,* 32, 365–387.

Graham, S. (2018a). A writer(s)-within-community model of writing. In C. Bazerman, V. W. Berninger, D. Brandt, S. Graham, J. Langer, S. Murphy, ... M. Schleppegrell (Eds.*), The lifespan development of writing* (pp. 272–325). National Council of English.

Graham, S. (2018b). A Revised writer(s)-within-community model of writing. *Educational Psychologist,* 53, 258–279.

Halliday, M. A. K. (1994). *Introduction to functional grammar* (2nd ed.). Edward Arnold.

Halliday, M. A. K. (1978). *Language as social semiotic.* Edward Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research,* 1(1), 1–26.

Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing*, 25(2), 347–373.

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate Peripheral Participation*. Cambridge University Press.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review,* 8, 299–325.

National Writing Project (2010). *Analytical Writing Continuum*. Available at https://archive.nwp.org/cs/public/download/nwp_file/21466/NWP_Analytic_Writing_Continuum.pdf?x-r=pcfile_d.

Olson, C. B., Chen, V., Krishnan, J., & Nazzal, J. (2018). Helping English Learners Navigate the Challenges of Academic Writing. In D. Basaraba, D. Baker, & C. Richards-Tutor (Eds.). *Second language acquisition: Methods, perspectives, and challenges*. Nova Science Publishing.

Olson, C. B., Scarcella, R., & Matuchniak, T. (2015). *Helping English learners to write: Meeting common core standard, grades 6-12.* Teachers College Press.

Olson, C. B., Woodworth, K., Arshan, N., Black, R., Chung, H. Q., D'Aoust, C., Dewar, T., Friedrich, L., Godfrey, L., Land, R., Matuchniak, T., Scarcella, R., & Stowell, L. (2020). The pathway to academic success: Scaling up a text-based analytical writing intervention for Latinos and English learners in secondary school. *Journal of Educational Psychology*, 112(4), 701–717.

Pearl, L. (in press). Modeling syntactic acquisition. In J. Sprouse (ed.). *Oxford Handbook of Experimental Syntax*.

Ranney, S. (2012). Defining and teaching academic language: Developments in K-12 ESL. *Language and Linguistics Compass*, 9, 560–574.

Scarcella, R. (2003). Academic English: A conceptual framework. *The University of California Linguistic Minority Research Institute Technical Report 2003-1*, Retrieved from https://escholarship.org/uc/item/6pd082d4.

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, 12, 431–459.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67–80). The Guilford Press.

CHAPTER 3

Study 1. Examining Lexical Features and Academic Vocabulary Use in Adolescent Multilingual Students' Text-based Analytical Essays

Abstract

Having rich and complex vocabulary is a crucial component that contributes to the quality of writing for academic purposes. However, use of academic vocabulary can be challenging for adolescent multilingual students who are developing their academic language proficiency. Thus, understanding the lexical needs of multilingual writers in composing academic essays is pivotal in supporting this population in their endeavor to become proficient academic writers. This study investigates the lexical features of adolescent multilingual students' text-based analytical essays and analyzes the extent to which lexical density, lexical diversity, and lexical sophistication predict the quality of their writing. Computational tools, Coh-Metrix and VocabProfilers, were used to obtain quantitative measures of lexical density, diversity, and sophistication. The results of the study indicate that the students' text-based analytical essays (n=70), on average, have: (1) low lexical density; (2) more repetition of words indicating less diversity compared to grade-level estimates; and (3) a higher percentage of basic and frequent words and lower percentage of academic words. 44% of the words covered by the Academic Word List (AWL) in the essays come from the source text and prompt. The results of multiple hierarchical regression indicate that the use of academic vocabulary is a predictor of writing quality. The study has important pedagogical implications for classroom practice at secondary school.

Keywords: Lexical density, Lexical diversity, Lexical sophistication, Academic writing, Text-based analytical writing, Adolescent multilingual students

Introduction

Developing proficiency in academic writing is a requisite for college and career success (Scarcella, 2003), and students who strive to pursue a path of higher education must learn to use academic language effectively in a variety of content areas. Academic language is defined as "a repertoire of language forms and functions that co-occur with school learning tasks across disciplines" (Uccelli et al., 2014, p. 1077). For academic language development, students' vocabulary knowledge and lexical competence play an important role, and the deployment of academic vocabulary is an important dimension of academic language skills. As the school tasks become increasingly complex throughout middle and high school, adolescent writers are expected to know and use academic vocabulary effectively in order to be successful in school (Nagy & Townsend, 2012). However, productive use of advanced, academic vocabulary can be particularly challenging for adolescent multilingual writers who are developing their language proficiency. Having rich and complex vocabulary, as well as an in-depth knowledge of academic register, is needed to engage in academic discourse. Both the breadth and depth of students' academic vocabulary, therefore, are crucial components that contribute to the quality of their academic writing. In a study that examined readers' judgment of multilingual writers' essays, Santos (1988) found that the readers were less tolerant of lexical errors and concluded that lexical weakness is a language component that "impinges directly on content" (p. 84). Although multiple components determine the overall quality of a written text within academic discourse, the general consensus is that "lexicon is a significant component in both the construction and interpretation of meaningful text" (Engber, 1995, p. 141).

Over the past several decades, there has been considerable learner-corpora based research that has examined the lexical features of students' writing. The development of linguistic

analysis tools and software programs, such as Coh-Metrix (McNamara et al., 2014), VocabProfilers (Cobb, 2018) and TAALES or Tool for the Automatic Analysis of Lexical Sophistication (Kyle & Crossley, 2015; Kyle et al., 2018**)** have made important contributions to the lexical analyses of corpus data. These tools have been used in numerous studies to analyze the lexical features of learner-generated texts and how various lexical features relate to language proficiency and writing quality (Crossley & McNamara, 2012; Crossley et al., 2011; 2014; Durrant et al., 2019; Guo et al., 2013; Kim & Crossley, 2018; Kyle & Crossley, 2016; McNamara et al., 2010; Morris & Cobb, 2004; Vogelin et al., 2019; Yu, 2010). These studies demonstrate that vocabulary measures and lexical analysis of students' writing are crucial and useful indicators of students' writing and language development and offer a window into how one's lexicon affects the quality of writing.

However, much of the research exploring the lexical features of student writing and text analysis of learner corpora have focused on student writing at the collegiate and pre-collegiate levels in a post-secondary context. In fact, adolescent multilingual students in a secondary education setting, despite their growing rate, are an underexplored group in the field of second language (L2) writing (Yi et al., 2018) and have remained "outside the purview" in L2 research (Ortmeier-Hooper & Enright, 2011, p. 167). This study focuses on adolescent multilingual students in mainstream classrooms in a US public secondary education setting to analyze their productive vocabulary use within the context of academic writing. Adolescent multilingual students in this context are often faced with cultural, cognitive, linguistic, communicative, contextual, and affective constraints while grappling with the demands of academic writing and learning an additional language (Olson et al., 2015). Thus, these students constitute a unique

population with differing exposure to and experiences of navigating the complex landscape of US secondary education.

In addition, most of the extant studies that examine how lexical features relate to writing quality have focused on independent writing tasks, as opposed to source- or text-based writing tasks. This is an important factor as different writing tasks (e.g., independent vs. source-based), modes (e.g., narrative, expository, argumentative) and genres (memoir, literary analysis, opinion editorial) elicit different linguistic features, and thus, may demand different linguistic skills (Crossley, 2020; Plakans & Gebril, 2013). Previous studies (Guo et al., 2013; Kyle & Crossley, 2016) indicate that "independent and source-based writing tasks differ in the lexical features that are predictive of writing proficiency scores" (Kyle & Crossley, 2016, p. 13). Other studies have found that expository writing exhibited less lexical repetition compared to narrative writing (Berman & Verhoeven, 2002) and that argumentative essays often exhibit more linguistic complexity (Yoon & Polio, 2017). Therefore, it may be the case that different linguistic skills are required for different writing tasks (Plakans, 2008; Plakans & Gebril, 2013).

Few studies have analyzed lexical features of text-based analytical writing in which students respond to a non-fiction text and analyze its theme. This is a writing task and genre that is distinct from a typical source-based, argumentative writing task. Unlike typical source-based writing that integrates multiple sources, this type of writing task draws on a single source—the text the students are analyzing. Crossley (2020) contends that "the majority of information available to the field about the interactions between writing quality/development and linguistic features is derived from a single task (independent writing), shedding some confidence on the generalizability of the findings to other tasks" (p. 432). Although recent studies have begun to focus on source-based writing (Guo et al., 2013; Kyle, 2020; Kyle & Crossley, 2016; Gebril &

Plakans, 2016; Plakans, 2009), to my knowledge, no study that examined lexical features of multilingual writers' texts has been conducted on text-based analytical writing. Therefore, the present study contributes to the extant literature by examining the lexical features in a lesser studied writing task or genre. Finding whether and how texts that represent different writing tasks and genres exhibit commonality and variability in lexical features is crucial for building a more comprehensive and complete understanding of the relation between linguistic features and writing quality.

To examine lexical attributes and academic vocabulary use in text-based analytical essays written by adolescent multilingual students, this study focuses on lexical density, lexical variation or diversity, and lexical sophistication, which are three main types of measure suitable for quantitative analysis of productive vocabulary (Durrant et al., 2019; Read, 2000). The goal of the present study is: (1) to analyze broadly the lexical features of text-based analytical writing on the aforementioned vocabulary measures; (2) to examine specifically the extent to which adolescent multilingual students incorporate academic vocabulary in their writing; and (3) to investigate how these three lexical features (density, variation, and sophistication) predict the human-judged quality of text-based analytical essays. In examining the relation between lexical features and writing quality, I conducted multiple hierarchical regression and added control variables of text length, which is found to have a large effect on writing scores (Guo et al., 2013), and demographic variables of English language proficiency (ELP) designation and gender. These demographic variables, which are often neglected in L2 research examining the relation between linguistic features and writing quality, are controlled in regression analyses for any potential confounding effect they may have. The following research questions guide the study:

1. What degree of lexical density, variation, and sophistication are observed in adolescent multilingual writers' text-based analytical essays?
2. To what extent do these students use academic vocabulary when writing analytical essays in response to an informational text?
3. To what extent do the three lexical features (density, variation, and sophistication) predict writing quality when controlling for essay length and demographic control variables of ELP designation and gender?

The findings of the study will inform researchers and practitioners of not only the relation between lexical features and writing quality in text-based analytical essays but also the lexical challenges developing multilingual writers encounter, as well as the skills they need to meet the demands of academic writing.

Literature Review

1. Lexical Features of Academic Writing

Academic writing across various genres and disciplines shares common lexical features. A linguistic feature that is conspicuous in academic writing within and across various disciplines is the vocabulary of academic language, which is often described as "comparatively large, precise, and formal" (Ranney, 2012, p. 563). Comparing the lexical features in spoken interactions and school-based texts, Schleppegrell (2001) found that words used in academic writing tend to be specific and technical and that academic texts have higher lexical density. The density, precision, and technicality of academic lexicons are associated with informational density, technical concepts, and content-specific words that are required to convey complex ideas. Vande Kopple (1994) contends that typical academic texts consist of a higher proportion of nouns, and this is an important lexical characteristic necessary to achieve informational

density. In addition, the highly structured nature of academic genres contributes to the lexical density in academic writing (Halliday, 1993).

Corpus-based research investigating the lexical features of academic writing has led to compilation of vocabulary lists and development of typologies of academic vocabulary based on frequency occurrences in different texts across different disciplines. The earliest compilation of a vocabulary list is the General Service List (GSL) developed by West (1953) and contains the most frequent 2000 English words in the corpus of written English. The GSL covers "up to 76% of the academic corpus" developed by Coxhead (2000) for the study of academic word list (p. 214). Coxhead's (2000) Academic Word List (AWL), which is compiled from "a corpus of 3.5 million running words of written academic texts" from 28 subject areas across four broad disciplines, excludes the GSL and identifies general academic words that are commonly found in different texts across different disciplines (p. 213). The AWL consists of 570 frequently-used academic words across disciplines and covers approximately 10% of words in academic texts (Coxhead & Nation, 2001). The vocabulary lists and the corpus data contributed to the development of vocabulary typologies. For example, Scarcella (2003) identified three major categories of academic vocabulary: (1) general words used across academic disciplines and in everyday situations; (2) academic words that are common across different disciplines; and (3) technical words found in specific academic fields. The words in the latter two categories tend to be highly specialized.

2. Lexical Features and Learner Corpora

In studies of learner corpora, lexical usage is operationalized in a variety of ways. While there is no agreed-upon method of operationalizing lexical features in writing, lexical density, diversity, and sophistication have been three types of measure that "lend themselves well to

automated, quantitative analysis of vocabulary use" (Durrant et al., 2019, p. 35). Several studies investigating lexical features of student writing found that texts that show higher academic language proficiency have more tokens and types and include a higher percentage of content words, exhibiting more information load (Shaw & Weir, 2007; Vidakovic & Barker, 2010). In addition, studies that investigate lexical richness and sophistication in learner corpora based on the lexical frequency profile (LFP) and vocabulary profile (VP) suggest that students with higher academic language proficiency tend to use less frequent words and a higher percentage of academic words (Biber & Gray, 2013; Crossley & McNamara, 2012; Laufer & Nation, 1995; Morris & Cobb, 2004).

In a study analyzing argumentative essays of multilingual students, Vogelin et al. (2019) experimentally manipulated the level of lexical diversity (measured by MTLD and D) and sophistication (measured by word range) to examine how the varying lexical features influence human judgment of essay quality. They found that the human raters assessed the overall quality more positively for essays that have greater lexical diversity and sophistication. Further, Hou et al. (2016) examined lexical density, variation, and sophistication in learner corpora and found that texts produced by students with different proficiency levels exhibit different lexical features. They concluded that "learners at different stages of proficiency make progress in different variables of the language" (p. 60). In general, L2 writing research indicates that lexical features are predictive of writing quality and proficiency (Crossley, 2020). Examining lexical features of students' texts within an academic writing context can help us understand the lexical needs and challenges developing writers have when composing academic essays. This understanding is crucial in writing and language instruction geared toward adolescent multilingual students' academic language development.

2.1. Lexical Density

Lexical density refers to the proportion of content words (i.e., nouns, verbs, adjectives, and adverbs) in a text (Laufer & Nation, 1995). Lexical density is closely associated with the notion of *information packaging* as content words in a text, unlike functional words (prepositions, determiners, pronouns, conjunctions etc.), carry information (Johansson, 2009). Thus, texts with a higher proportion of content words are considered to be dense as they package more information as opposed to texts that have a higher proportion of function words (Johansson, 2009). Since academic texts are informationally and lexically dense, it can be assumed that high-quality academic essays written by students may exhibit a higher percentage of content words. However, there is a lack of empirical evidence that demonstrates a correlation between lexical density and writing quality. Several studies that examined the relationship between lexical density and writing quality found that lexical density was not significantly correlated with writing quality (Engber, 1995; Uccelli et al., 2012). In a study examining the association between lexical richness and quality of writing in an independent writing task completed by university students, Engber (1995) included lexical density as one of the measures. Her study found that the association between the percentage of content words (r = .23) and the human-rated writing quality was weak. Similarly, Uccelli et al. (2012) found no significant correlation between lexical density and writing quality in an independent writing task completed by high school seniors.

Several concerns arise in using a lexical density measure to predict the quality of writing. One concern is that lexical density is dependent on "syntactic and cohesive properties of the composition," meaning "fewer function words in a composition may reflect more subordinate clauses, participial phrases and ellipsis" (Laufer & Nation, 1995, p. 309). These syntactic and

structural features of sentences may confound the measure of lexical density in assessing writing quality. Another concern is that lexical density is highly correlated with lexical diversity, with correlations as high as $r = .74$ (Johansson, 2009). Despite these concerns, exploratory investigation of lexical density in texts written by multilingual students at different proficiency levels may provide some insight into the lexico-grammatical features of writing. For example, Hou et al. (2016) found that texts written by multilingual students with different proficiency levels exhibit variability in lexical density. It is also important to note that the earlier studies (Engber, 1995; and Uccelli, 2012) examined lexical density in independent writing tasks, as opposed to source-based writing. To my knowledge, no study has examined lexical density in text-based analytical writing.

2.2. Lexical Diversity/Variation

Lexical diversity, also referred to as lexical variation, indicates the variety and the range of productive vocabulary used in a text and is often measured by type-token ratio, that is the ratio of the total number of unique words (type) and the total number of running words (token) in a text (McCarthy & Jarvis, 2007). The computational tool Coh-Metrix provides four different types of LD measures: type-token ratio for content word lemmas, type-token ratio for all words, *vocd* and Measure of Textual Lexical Diversity (MTLD). Simple type-token ratio is sensitive to text length as longer texts have more words that are repeated (Jarvis, 2013). The latter two LD measures — *vocd* and MTLD — are adjusted using estimation algorithms to reduce the impact of length (McNamara et al., 2014). Lexical diversity is one of the commonly used variables associated with the deployment of productive vocabulary. It is found to be a useful indicator of writing quality (Crossley & McNamara, 2012).

L2 writing studies have found a positive relationship between lexical diversity and writing proficiency and demonstrated that essays that are scored higher by human raters show greater lexical diversity (Crossley et al., 2014; Engber, 1995; Yu, 2010). For example, Engber's (1995) study, which includes measures of lexical variation (using type-token ratio) with and without errors, found moderate associations between the lexical variation with error (r = .45) and without error (r = .57) and the holistic writing score. Similarly, Crossley et al. (2014), who used MTLD as their LD measure in analyzing journals written by university-level students (with an age ranging from 18-27), reported a strong correlation between lexical diversity and the holistic score of lexical proficiency (r = .70). Using *D* (vocd) as their LD measure in analyzing multilingual students' compositions on both personal and impersonal topics, Yu (2010) found a positive correlation with writing quality. Yu's (2010) study also revealed a significant difference in lexical diversity between different writing topics (personal vs. impersonal topics). Overall, L2 writing studies that investigate lexical variation in learner corpora suggest lexical diversity to be predictive of writing proficiency and, thus, an important measure to be included in lexical analysis.

2.3. Lexical Sophistication

Another important variable that is used to analyze lexical features of learner corpora is lexical sophistication. Although lexical sophistication, in some studies, is loosely referred to the percentage of 'advanced' words and less frequent words in a text (Laufer & Nation, 1995), there is no current consensus on its definition. While word frequency is "the prototypical measure of lexical sophistication" (Crossley & Kyle, 2018, p. 48), recent studies have proposed and used various measures and indices for lexical sophistication such as word sequences, semantic relationships, and psycholinguistic properties (Crossley et al., 2013; Crossley & McNamara,

2012; Guo et al., 2013; Kyle & Crossley, 2015; 2016). For example, Kyle and Crossley (2016) expanded lexical sophistication by using word range, bigram and trigram frequency (commonly co-occurring two- or three-word sequences), hypernymy, and psycholinguistic measures (word familiarity, imageability, and age of acquisition) of words in a text. The variations of measures observed in extant studies attest to the complexity of operationalizing lexical sophistication, indicating that the construct of lexical sophistication is not clear-cut. Despite the variation in its measures, lexical sophistication is an important component of language and a strong predictor of writing quality based on the notion that proficient writers with greater vocabulary skills use sophisticated words and expressions (Crossley & Kyle, 2018).

Several studies that analyze lexical sophistication use a word frequency count based on the notion that proficient writers use less frequent and more advanced words, showing lexical sophistication (Crossley et al., 2013; Laufer & Nation, 1995; Morris & Cobb, 2004). For example, Laufer and Nation's (1995) study analyzed university-level multilingual students' texts using the Lexical Frequency Profile. They found that texts written by more proficient writers used less frequent words in their texts. In addition, Crossley and McNamara (2012) found a moderate association between word frequency (r = -.34) and writing proficiency in their study that analyzed essays written by high school seniors. Their findings suggest that students with higher writing proficiency use less frequent, more advanced words. Recent studies have examined lexical sophistication in source-based academic writing in comparison to independent writing tasks (Guo et al., 2013; Kim & Crossley, 2018). Guo et al. (2013) found that integrated and independent tasks exhibited different lexical features. For example, word frequency was a significant predictor of writing quality in integrated tasks, but not in independent tasks. Kim and Crossley's (2018) study that also examined both independent and integrated writing tasks,

however, found that “word frequency measures were correlated with independent writing scores, but not with source-based writing scores” (p. 51). This indicates that the findings of the studies that analyzed source-based writing tasks have not been consistent. Despite this inconsistency, L2 writing research, in general, suggests that as learners develop higher language proficiency, they use longer and less familiar words (Crossley et al., 2011), greater use of infrequent and advanced words (McNamara et al., 2010), and less concrete words (Crossley et al., 2014; Crossley & McNamara, 2012; Salsbury et al., 2011).

Taken together, lexical density, variation, and sophistication have been used in a number of studies that illuminate our understanding of the relation between lexical features and writing quality. However, there are still gaps and inconsistencies in extant literature. One important gap in research is tied to the scarcity of studies that explore lexical features and their relations to writing quality in source-based writing. As noted by Guo et al. (2013), “fewer studies have explored how linguistic features predict scores of integrated writing tasks” (p. 221). In addition, lexical features of texts produced by adolescent multilingual writers in a secondary education context have been underexplored. This study addresses these gaps in the research by examining the lexical features of text-based analytical essays written by adolescent multilingual students in secondary education context.

Methodology

In exploring the lexical features of adolescent multilingual students’ text-based analytical essays, three major procedural steps were undertaken. The first step was to prepare the corpus for analysis, which involved typing the essays, correcting all spelling errors, and removing essays that did not qualify for the analysis of lexical features (see “Procedures” section for a description of exclusion criteria). The second step included conducting lexical analyses of the selected

essays using two computational and analytical tools: Coh-Metrix (McNamara et al., 2014) and VocabProfilers from *Compleat Lexical Tutor* (Cobb, 2018). The final step involved statistical analyses of the selective quantitative data pertinent to lexical density, variation, and sophistication.

Corpus Description

The initial corpus comprised 86 essays selected from the pool produced by middle and high school multilingual students who participated in a reading and writing intervention developed by the UCI Writing Project during the 2016-2017 academic year. The students are from an urban school district that has a large percentage of multilingual students whose first and home language is Spanish. All essays were written by Spanish-speaking multilingual students (in grades 7 through 12) who are classified either as an English Learner (EL) or Reclassified Fluent English Proficient (RFEP). These classifications are based on California English Language Development Test (CELDT) that students whose primary language is not English have to take. The designation EL is given to students who score at 50th percentile or below on CELDT, while RFEP students are reclassified from English learner status to fluent English proficient status based on four criteria: subsequent English language assessment, teacher evaluation, parental recommendation, and strong academic performance (California Department of Education, 2020). This English language proficiency (ELP) designation is considered a demographic variable rather than English proficiency score as the reclassification is not dependent solely on a test score.

From the initial 86 essays, 16 essays were excluded from the study for two reasons. First, essays that are too short (less than 100 words) were not included considering the fact that comparable length is desirable for the analysis of lexical features. The suggested text length,

which is 100 to 1000 words, for using the computational tool Coh-Metrix (McNamara et al., 2014) was also taken into consideration. Second, essays predominantly comprised of large chunks of quotes and sentences that are copied verbatim from the reading text were not included in the study. The quoted and copied sentences have a confounding effect since the purpose of the study concerns students' lexicon and productive vocabulary use in their analytical writing. Thus, the final corpus used in the study consisted of 70 essays. In the final corpus, 81% of the essays are written by students who are classified as RFEP, and the remaining 19% are written by those classified as EL. Table 1 provides descriptive details, including length measures indicated by word, sentence, and paragraph counts, of the learner corpora by students' grade levels. As seen in Table 1, the average length of the essays in the full sample is 189.24 (SD = 73.24) words. An average number of sentences is 10.6, and the average number of paragraphs is 2.59 in the sample. As indicated by word and paragraph counts, the mean scores of the length of the essay and number of paragraphs increase as the grades advance (with the exception of 11th grade), indicating that students in higher grades write longer essays and more paragraphs. Word length (letters) and sentence count, however, do not follow the same ascending trend.

Table 3.1

Descriptive Statistics of the Learner Corpora: Length Measures

	Full sample		7th Grade		8th Grade		9th Grade		10th Grade		11th Grade		12th Grade	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Word count	189.2	73.84	160.4	63.00	171.5	54.92	196.1	75.54	202.7	106.0	192.2	56.38	221.6	80.72
Word Length	4.19	0.25	4.19	0.20	4.18	0.26	4.13	0.31	4.30	0.16	4.14	0.23	4.22	0.30
Sentence count	10.60	5.04	9.09	4.35	9.00	3.34	9.77	5.17	13.25	6.44	10.62	3.48	12.62	6.82
Paragraph	2.59	1.43	1.45	0.93	2.77	1.69	2.85	1.52	2.83	1.27	2.23	1.01	3.62	1.41
Observations	70		11		13		13		12		13		8	

Text-based Analytical Writing Task

All essays in the sample were from pretest or before the implementation of the intervention. At the beginning of the school year before the intervention, the participating students completed reading and analytical writing tasks over two class periods. During the first period, the students read a literary nonfiction text (a short newspaper article) and completed a post-reading packet focusing on summarizing, visualizing, analyzing a theme and author's craft, and reflecting on the purpose and significance of the article. During a second period, the students wrote analytical essays responding to a writing prompt that asked them to select and analyze a theme in the article, make a claim about the theme they selected, and support their claim with evidence from the text (see Appendix B for prompts). The writing task was designed to prompt students to write interpretive responses in an academic context.

Procedures

The initial phase of the research method involved preparing the learner corpus for lexical analysis. This included typing the essays since all of them were handwritten, correcting all spelling errors, and deleting sentences and large chunks of phrases (three or more consecutive words) that are either copied verbatim or quoted from the informational text to which students responded. The copied sentences and phrases from the essays were deleted prior to the analysis because they reflected the lexical features of the source material rather than students' language use. Spelling errors were corrected for the purpose of obtaining accurate measures of lexical density, diversity, and sophistication using computational tools Coh-Metrix (McNamara et al., 2014) and VocabProfilers (Cobb, 2018).

Holistic Scoring of Writing Quality: All essays were scored holistically using 6-point scale rubric developed by UCI Writing Project to assess the writing quality (see Appendix D for

rubric). As indicated in the analytical writing assessment rubric, the writing quality was assessed based on the quality of interpretation, the clarity of thesis/claim, the organization of ideas, the integration of evidence, syntactic variety and sentence fluency, diction and word choice, and language conventions. Each essay was scored by two trained raters, scoring independently, and the two scores were added together for a final score between 2 and 12. If there was a discrepancy of two or more points between the raters, the essay was scored by a third rater. The third rater's score and the closest score given by either first or the second raters were added together for a final score. If the third reader's score was in the middle of the two raters' scores, the average score of the first and second raters was added to the third rater's score for a final score. Raters for all but one paper were in exact agreement or differed by one point. The first- and second-rater scores were highly reliable with an alpha of 0.87. Raters agreed within a single score point or better for 99% of the papers; 56% of the papers had exact agreement between the two raters.

Measuring Lexical Features: The vocabulary profile for each paper was generated using Coh-Metrix (McNamara et al., 2014) and VocabProfilers provided by *Compleat Lexical Tutor* (Cobb, 2018). Lexical density was measured by the percentage of content words (nouns, verbs, adjective, and adverbs) for each essay using Web VP Classic v.4 version of VocabProfilers. The frequency of content words is a commonly used density measure in L2 writing studies (Uccelli et al, 2012). For lexical diversity, Measure of Textual and Lexical Diversity (MTLD) provided by Coh-Metrix was used. MTLD was chosen because type-token ratio is confounded by text length as longer texts have more words that are repeated (Jarvis, 2013; McCarthy & Jarvis, 2007). MTLD attempts to "overcome that confound by using estimation algorithms" (McNamara et al., 2014, p. 67). McCarthy & Jarvis (2010), in their validation study, found that "MTLD performs well with respect to all four types of validity and is, in fact, the only index not found to vary as a

function of text length" (p. 381). For lexical sophistication, I used six measures—word length, word frequency, and age of acquisition provided by Coh-Metrix, percentages of k1 words (first 1000 GSL words), k2 words (second 1000 GSL words), and words covered by AWL provided by VocabProfilers. These LS measures were chosen based on the findings of previous studies that indicate higher quality academic writing contains longer and less familiar words (Crossley et al., 2011) and greater use of infrequent and advanced words (McNamara et al., 2010).

Data Analytic Strategies

To analyze the lexical features and their relation to writing quality, both descriptive and statistical analyses were conducted. Descriptive statistics of the lexical features provided by Coh-Metrix and VocabProfilers was employed to answer the first two research questions. This is to identify the lexical patterns, as well as prominent features of lexicon, that are present in the adolescent multilingual writers' text-based analytical writing. To answer the last research question about the association between the lexical features and holistic writing score, Pearson Correlations were first conducted between all lexical variables and the holistic writing quality. Following the correlation analyses, multiple hierarchical regression was conducted to examine the extent to which the lexical density, diversity, and sophistication predict the human-judged writing quality. In the regression models, demographic variables of ELP designation (EL vs. RFEP) and gender were included as control variables. Gender was added as a control variable since gender-based differences in writing performance have been documented in NAEP report and observed in writing research (Pajares & Valiante, 2001). ELP designation was also added as a control variable because multilingual students are diverse, with differing levels of linguistic exposure and experiences. In addition, even RFEP students with fluent English proficient status

but without sufficient exposure to academic English may produce texts that are scored low when writing text-based analytical essays that require academic English proficiency.

Results

The goal of the study was to investigate what lexical features the text-based analytical essays of adolescent multilingual writers exhibit and to analyze the extent to which lexical density, diversity, and sophistication predict writing quality.

1. Lexical Features Observed in Text-based Analytical Essays

In general, the text-based analytical essays in this study have: (1) low lexical density, with lower percentage of content words and higher percentage of function words; (2) more repetition of content and all words indicating less diversity compared to grade-level norms approximated based on Common Core State Standards; and (3) a higher percentage of basic and frequently used words and lower percentage of academic words. Table 2 shows the correlations among the mean percentages of content words for lexical density, MTLD for lexical diversity, and the mean scores of lexical sophistication measures — word length, word frequency, age of acquisition (AoA), k1 words, k2 words, and AWL-covered words.

Table 3.2

Correlation Matrix: Lexical Features and Writing Quality Score

		M	*SD*	Skew	Kurt	1	2	3	4	5	6	7	8
1.	Score	5.74	1.80	0.28	3.14	—							
2.	CW	46.47	3.79	0.40	2.79	0.13	—						
3.	MTLD	62.54	18.36	0.44	2.56	0.22	0.17	—					
4.	WL	4.19	. 24	0.32	3.01	0.12	0.66***	0.21	—				
5.	WF	2.57	.15	0.01	2.50	0.01	-0.43***	-0.30*	-0.51***	—			
6.	AoA	308.57	26.92	-0.35	3.36	0.34**	0.10	0.10	0.18	-0.20	—		
7.	k1	89.23	3.80	-0.08	2.00	0.09	-0.42***	-0.27*	-0.40***	0.58***	-0.01	—	
8.	k2	4.51	2.30	0.68	3.19	-0.19	0.31*	-0.03	-0.00	-0.23	-0.16	-0.57***	—
9.	AWL	1.58	1 .28	0.71	2.67	0.29*	0.24*	0.16	0.314**	-0.39***	0.30*	-0.37**	-0.15

Note. n=70. * p<0.05. ** p<0.01. *** p<0.001; Score=Writing score, CW=content words, WL=Word length, WF=Word frequency, Skew=Skewness, Kurt=Kurtosis

In general, the text-based analytical essays written by adolescent multilingual students in the sample have a higher percentage of function words than content words. In the sample, the mean percentage of content words was 46.47% (SD = 3.79), showing less lexical density. On the lexical diversity measure, the essays in the study, on average, have lower score of MTLD, indicating less diversity and more repetition of content and function words. A higher score of MTLD indicates more variety of words used in a text while a lower score suggests more words that are repeated. Grade-level norms approximated according to Common Core State Standards' grade bands show that the MTLD mean score is 88.35 (SD=25.31) in grades 6-8 and 95.30 (SD=25.51) in grades 9-10 (McNamara et al., 2014). However, the essays in the sample have a substantially lower score of MTLD (M=62.54; SD=18.36).

For lexical sophistication measures, the essays exhibited a higher percentage of more frequent words. The words from GSL 1-1000 (k1 words) covered 89.23% and words from GSL 1001-2000 (k2 words) covered 4.51% of the text on average. Thus, the overall GSL coverage of the study corpus was 93.74%, whereas words from AWL, on average, covered only 1.57% of the text. GSL 1-2000 words tend to be basic words commonly used in everyday communication while words covered by AWL include more advanced words typical in academic texts. Coxhead's corpus of academic texts has 76.1% GSL coverage and approximately 10% AWL. In general, the AWL coverage in the corpus was far below the estimated figures (about 10%) for the AWL words in academic texts. The Coh-Metrix measures of word frequency (M=2.57; SD=.15), on average, had a higher score compared to grade level estimate, suggesting the use of more frequent words. The mean AoA score (M=308.57; SD=26.92) indicates that the vocabulary level, on average, is in the lower tiers of vocabulary. Higher scores of AoA (600-700) indicate academic and technical words that are not common in everyday speech.

2. The Use of Academic Vocabulary

The second research question was to what extent adolescent multilingual students use academic vocabulary when writing text-based analytical essays. As mentioned above, the essays in the corpus generally have a lower percentage of words from AWL. I created a list of academic words covered by AWL in all the essays. I then checked which of these academic words appear in the prompt and reading text. In completing a text-based writing task, students use and rely on the source text to provide summary and evidence in an effort to develop and support their main ideas. Table 3 provides academic words (covered by AWL) present in students' essays. Among the academic vocabulary in students' texts, 44% were words that appeared in the prompt (11%) and the information text (33%) they read and responded to.

Table 3.3

Academic Words Present in Students' Text-based Analytical Writing

Academic vocabulary covered by AWL in the study corpus						
affect	conflict	fundamental	instance	occur	relax	tropical
aid	contact	grant	involve	occurrence	respond	ultimately
author	create	hemisphere	item	paragraph	response	unidentified
aware	despite	highlight	job	perspective	section	unique
capacity	drug	huge	label	phrase	significant	vehicle
cell	economic	hurricane	location	positive	survive	visualize
collapse	emphasize	ignore	magazine	process	survivor	
community	environment	impact	magnitude	quote	team	
communicate	evaluate	imagery	majority	reach	text	
conclude	finally	incident	maximum	react	theme	
conclusion	focus	injure	normal	reaction	topic	

Words not in reading and prompt (56% productive words)			**Words in the reading text** (33% of productive words)		**Words in the prompt** (11% of productive words)
affect	finally	maximum	capacity	reaction	author
aid	focus	occur	conclusion	relax	create
aware	grant	occurrence	despite	section	magazine
cell	highlight	paragraph	economic	survive	process
collapse	huge	perspective	fundamental	survivor	response
community	imagery	phrase	hemisphere	team	significant
communicate	incident	positive	hurricane	tropical	text
conclude	instance	quote	ignore	ultimately	theme
conflict	involve	reach	impact	unidentified	
contact	item	react	injure	unique	
drug	job	respond	location	vehicle	
emphasize	label	topic	magnitude		
environment	majority	visualize	normal		
evaluate					

More proficient multilingual students with higher writing scores deploy academic words not only from the source text but also from their own lexicons; on the other hand, less proficient writers' essays exhibit no or little use of academic vocabulary and avoidance of academic words from the source text. In Table 4, an essay that received a holistic score of 10 (out of 12) and an essay that received a score of 4 are juxtaposed and the academic words that are present in both essays are highlighted. The high-scored essay has 4.44% of AWL-covered words (shown in bold) and uses a variety of academic words that do not appear in the prompt and the source text. In contrast, the low-scored essay has only one AWL-covered word (*theme*) that also appears in the prompt, constituting only 0.37% AWL coverage.

Table 3.4

Comparison of Academic Words in High- and Low-scored Essays

An essay with a writing score 10 (out of 12)	An essay with a writing score 4 (out of 12)
In the article "the Man in the Water" the **author** Rosenblatt describes the **admirable** and **significant** act of one man to save people. In his article Rosenblatt describes how a man's decision saved lives, and what he hopes to tell us is that a **significant** act sometimes needs to remind us that we do have power. I am not saying **superpowers**, but the power to give life to someone. For example, in the third **paragraph** it talks about a park police helicopter **team**, and the last two sentences, something none of us really thought about was said. Which means we choose to go help someone and that is power. Moreover, in the fourth and fifth **paragraphs** Rosenblatt describes the man in the water and his actions. Rosenblatt **communicates** that when the plane crashed he managed to hang on and help other passengers. Also one of the helicopter **team** tells us that the man in the water was willing to sacrifice himself in order to save others, not just himself. **Furthermore**, Rosenblatt **emphasizes** how the man in the water was calm and like any other ordinary passenger. Rosenblatt shows us how a decision made in a matter of seconds can change everything. He also shows us how he was desperate to live but decided to give others life. In **conclusion**, Rosenblatt describes the relationship between the man in the water and nature. So Rosenblatt's message is to know we do have power in the world we are not powerless.	It takes courage for what a man did that night. Willing to die for the safety of others, the man did something remarkable. The **theme** of the story "The man in the Water" is the greatest gift you can give to anyone is your life. The man in the water stayed unidentified, no one knows till this day, the hero that saved lives. This man had more than courage to do what he did. Every time the helicopter would throw a life vessel at him, he would pass it down to another person. He was willing to put his life at danger to save others. He didn't even know these people but he still did what he did. To whoever he passed down the life vessel to, he gave them the gift of life. People nowadays don't have the courage to do what he had done. They would rather have their safety first. Which in other words is true. But this man didn't care. He sacrificed his life for others. Honestly, if I was in his position I would have saved my life first. I would feel guilty but blessed at the same time because I would still be alive. The man in the water gave the greatest gift you can give to anyone, your own life. He sacrificed his life for the safety of others. He showed strength and courage to do what he had done. Staying unknown sucks because we don't know who was the hero that day. At the end the man gave the greatest gift anyone can receive a second chance.

Relation between Lexical Features and Writing Quality

To examine the relation between lexical features and writing quality, Pearson correlations between all lexical features and writing score were conducted (Table 2). Normality of the distribution for the selected indices and multicollinearity were checked. The values for skewness between - 2 and +2 and values for kurtosis between -7 and +7 were considered to indicate normality (Curran et al, 1996; Hair et al., 2010). Multicollinearity was not detected since the r value for each variable fell below 0.70 threshold. The results of Pearson Correlation indicated that only two lexical sophistication measures — AWL words (r = .29, p =. 014) and age of acquisition (r = .36, p =. 005) — were significantly correlated with writing quality. All other lexical features were not significantly correlated with writing quality. In multiple hierarchical regression analyses, AWL words and age of acquisition, which were the two variables correlated with writing quality, were used as predictor variables.

Regression models 1 and 2 show an association between age of acquisition and holistic writing score. In Model 1 where only demographic variables were controlled for, the regression coefficient of age of acquisition is .024 (p = 0.002), indicating that it was a significant predictor of writing quality. In model 2 in which text length was added as a control variable, the coefficient of age of acquisition slightly decreased (B = .023, p = 0.001) and remained significant. Regression models 3 and 4 show the relationship between the percentage of academic words and writing quality score. In Model 3, in which only demographic variables were controlled for, the regression coefficient of AWL words was .39 (p = 0.021), indicating that it was a significant predictor of writing quality. In Model 4 where text length was added as a control variable, the coefficient of lexical sophistication decreased (B = .34, p = 0.022) and remained statistically significant. Model 5 includes all five variables and shows age of

acquisition (B = .020, p = 0.005) and text length (B = .011, p < 0.001) as significant predictors of holistic writing quality score (Table 5).

Table 3.5

Regression Models: Lexical Sophistication Predicting Writing Quality

	Model 1	Model 2	Model 3	Model 4	Model 5
Age of Acquisition (content words)	0.02**	0.02***			0.02**
	(0.01)	(0.01)			(0.01)
AWL words (percentage)			0.39*	0.34*	0.21
			(0.16)	(0.14)	(0.14)
Text length		0.01***		0.01***	0.01***
		(0.00)		(0.00)	(0.00)
EL designation	-0.60	-0.54	-0.44	-0.42	-0.42
	(0.51)	(0.44)	(0.54)	(0.47)	(0.44)
Gender	-0.84*	-0.60	-0.66	-0.44	-0.59
	(0.41)	(0.35)	(0.41)	(0.36)	(0.35)
Intercept	-1.28	-2.99	5.50***	3.36***	-2.40
	(2.32)	(2.02)	(0.40)	(0.57)	(2.04)
R-squared	0.185	0.410	0.130	0.353	0.430

Note. * p<0.05, **p<0.01, *** p<0.001; Unstandardized Regression Coefficients, standard errors (*se*) in Parenthesis.

Overall, age of acquisition explained 19% of the variance and percentage of AWL words explained 13% of the variance in human-judged writing score when controlling for gender and ELP designation. These variances of lexical features in writing quality are important since multiple components, including both global and local features, determine the overall quality of writing. The results suggest that both of these measures of lexical sophistication contribute to writing quality, indicating essays scored higher by two human raters had a higher percentage of academic vocabulary and words with higher age of acquisition scores. The findings from previous studies (Laufer & Nation, 1995; Crossley & McNamara, 2012) have reported that

proficient writers used less frequent and more academic words. Thus, the results of the current study support a significant association between academic word use and writing quality.

Discussion

The present study investigated the lexical density, diversity, and sophistication and their relation to writing quality in text-based analytical texts produced by adolescent multilingual students. The findings of the study revealed that the adolescent multilingual students' academic essays had low lexical density and diversity. The study also demonstrated that the essays in the study corpus had a much higher percentage of basic and frequently used words and a much lower percentage of words covered by AWL. The percentage of academic vocabulary in the students' texts showed a significant correlation with holistic writing score. Essays scored low predominantly consisted of high-frequency and basic words that are typical in everyday communication and low use of academic vocabulary while essays scored high had a higher percentage of academic vocabulary.

The results of this study demonstrated that lexical sophistication based on age of acquisition and the percentage of words covered by AWL were significant predictors of the holistic writing score. The other two lexical features (density and diversity) did not exhibit significant correlations with the writing score. While the results of this study are in accord with the findings of the foregoing studies on the relation between lexical density and writing quality, as well as lexical sophistication and writing quality, the results of the present study contradict the findings of previous studies that reported a statistically significant association between lexical diversity and writing quality. The inconsistent results can be attributed to the possibility that the lexical features of writing quality may vary depending on the type of writing task (i.e., source-based vs. independent) and mode (i.e., argumentative, expository, narrative). In this regard,

Crossley (2020) reminds us that assessing linguistic features of student writing needs to consider the possibility that different linguistic skills may be required for different writing tasks. In support of this, studies have shown significant differences in lexical features between source-based writing and independent writing (Kyle & Crossley, 2016) and between narrative and expository texts (Berman & Verhoeven, 2002).

The finding of this study that the essays showing lower proficiency exhibit low use of sophisticated and academic words and that lexical sophistication predicts writing quality are consistent with the findings of previous studies. The contribution of the present study, however, is that similar features of lexical sophistication are observed in text-based analytical essays in which students analyze and use a single source. An important finding of the present study that was not much explored in previous studies is the extent to which students used words from the text and the prompt. The finding that 44% of the AWL words in students' texts appeared in the source text and the prompt points to the value of using academic texts and genres to scaffold developing multilingual writers' lexical development and to expand their lexical resources. As students engage in close reading and critical analysis of an academic text and source, the vocabulary they aspire to acquire may naturally occur.

This study is not without limitations. First, the present study utilized quantitative measures of three lexical features and did not account for lexical errors in the productive use of words. Quantitative analysis of lexical features provides us with an insight into general characteristics and patterns of students' lexicon and productive vocabulary use in their writing; however, a qualitative study that investigates the way the words are used in context is crucial in gaining a comprehensive understanding of lexical features in students' writing. For example, the presence of academic words alone is not adequate in determining lexical sophistication; whether

the words are used appropriately within a given context needs to be further explored to identify the level of sophistication. Second, the sample in this study consists of only Spanish-speaking multilingual students from one school district. Adolescent multilingual students constitute a diverse group with unique backgrounds and differing linguistic needs. In order to see whether the linguistic needs identified in this study pertain to other language-minority students, a much larger sample that represents diverse L1 language backgrounds is needed. Future research should address these limitations and use a more inclusive sample that represents a linguistically heterogeneous mixture of multilingual students.

Pedagogical Implications

This study has important pedagogical implications. As students enter middle school and advance to high school, academic reading and writing tasks become increasingly complex (Scarcella, 2003). Successfully completing academic tasks that involve higher-order analytical and critical thinking skills require a higher-level of lexical proficiency and awareness of lexical choices. This study demonstrated that the use of academic vocabulary is an important indicator of writing quality. It also showed that adolescent multilingual students, especially low-proficient writers, depend on basic, high-frequency words that are common in informal conversation and spoken interactions. Given these findings, it is crucial for secondary classroom teachers to immerse learners in texts and activities that have rich academic vocabulary. Based on the results of the study, several recommendations can be drawn for instructional practices geared toward developing multilingual students' lexical needs for academic writing. Writing pedagogy in a secondary education context needs to: 1) provide developing multilingual writers ample opportunity to encounter academic words and expressions; and 2) integrate academic language instruction that promote lexical variety and the acquisition of academic words and expressions.

To be able to compose analytically and persuasively and to write confidently and competently within an academic context, adolescent multilingual students need to know the lexical choices that are appropriate for academic genre conventions. Lexical constraints multilingual students face in composing analytical texts may thwart their ability to express complex ideas. Through exposure and explicit instruction, teachers can help developing multilingual writers not only to build lexical capacity for academic writing but also develop awareness that lexical choices are shaped by socially constructed genre-conventions. Classroom exposure and instruction for lexical development is particularly important for developing multilingual students in a secondary education context since school is the only place where many of these students are exposed to academic language.

Developing proficiency in academic vocabulary and language in general is necessary for adolescent multilingual students' admission to and completion of college that requires a mastery of academic language. Studies have shown that immigrant students in college who completed their secondary education in U.S. schools had lower writing performance compared to international students (di Gennaro, 2013) and exhibited limited knowledge of formal grammar and language because they are likely to have developed their proficiency in English by immersing themselves in casual, non-academic situations (Reid, 2006). For many language-minority students across different school districts and states, school is the only place where they are exposed to academic vocabulary and language. Oftentimes these students are restricted in their access to rigorous curricula, and as a result, become long-term English learners (Kanno & Cromley, 2015). Therefore, a pedagogical practice geared to meeting the lexical needs of these students within an academic writing context is a crucial step and effort to promote equal access and level the playing field for this student population.

Conclusion

It is crucial to identify lexical characteristics and understand lexical challenges middle and high school students, particularly diverse multilingual writers, encounter in composing source-based analytical essays within academic writing discourse. Previous studies that examined the lexical features of multilingual students' texts have provided insights and illuminated our understanding of students' productive vocabulary and its role in text quality. However, much of the research on corpus analysis and linguistic features of multilingual writers' texts has focused on independent writing tasks in the postsecondary context, exploring the language and writing development of undergraduate, graduate, ESL students in university-level intensive English programs (Yi et al., 2018).

This study adds to the extant literature by examining the lexical features of text-based analytical essays written by adolescent multilingual writers in the U.S. secondary education context. Furthermore, adolescent multilingual writers' text-based analytical essays are particularly important not only because of a dearth of research on this type of writing but also because it is an academic writing task that requires students to analyze, interpret, and integrate an external source. Unlike some independent writing tasks that only tap into students' experiences and opinions, "academic writing tasks commonly use external sources for content" (Plakans & Gebril, 2012, p. 19). In addition, Crossley (2020) stressed the importance of considering the writing task when assessing linguistic features of multilingual writers' texts because different writing tasks may require students to draw on different linguistic resources and skills. Since the focus of this study was on academic writing and academic vocabulary use, it was important to analyze the lexical features of multilingual students' texts that represent academic writing.

Vocabulary knowledge and lexical proficiency greatly contribute to the quality of writing for academic purposes. Therefore, understanding lexical challenges adolescent multilingual students encounter when composing academic texts is a crucial step toward addressing their linguistic needs in both language and content area classrooms. This is particularly important since many of these students are often marginalized in their classrooms and schools, offered "reductionist curriculum focusing primarily on skill and drill" (Olson & Land, 2007, p. 217), and are restricted in their access to rigorous curricula (Kanno & Cromley, 2015). Understanding adolescent multilingual students' lexical challenges and addressing their lexical needs for the demands of academic writing are crucial for supporting them in their effort to develop academic literacy skills that are necessary for access to and success in postsecondary education.

References

Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy,* 5(1), 1–43.

Biber, G., & Gray, B. (2013). *Discourse characteristics of writing and speaking task types on the TOEFL ibt® test: A lexico-grammatical analysis (TOEFL iBT® research series No. TOEFLiBT 19).* Educational Testing Service.

California Department of Education. (2020). English learners. Retrieved from https://www.cde.ca.gov/sp/el.

Cobb, T. (2018). *Compleat lexical tutor.* Retrieved from http://www.lextutor.ca.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly,* 34(2), 213–238.

Coxhead, A., & Nation, I. S. P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes*. Cambridge University Press.

Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research,* 11(3), 415–443.

Crossley, S. A., & Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing,* 38, 46–50.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35, 115–135.

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41, 965–981.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics,* 36, 570–590.

Crossley, S. A., Salsbury, T., McNamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing,* 29, 243–263.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods,* 1, 16–29.

di Gennaro, K. (2013). How different are they? A comparison of Generation 1.5 and international L2 learners' writing ability. *Assessing Writing,* 18, 154–172.

Durrant, P., Moxley, J., & McCallum, L. (2019). Vocabulary sophistication in first-year composition assignments. *International Journal of Corpus Linguistics*, 24, 33–66.

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

Gebril, A., & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for Academic Purposes,* 24, 78–88.

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing,* 18(3), 218–238.

Hair, J., Black, W. C., Babin, B. J. & Anderson, R. E. (2010) *Multivariate data analysis* (7th ed.). Pearson Educational International.

Halliday, M. A. K. (1993). Some grammatical problems in scientific English. In M.A.K. Halliday, & J.R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 69–85). University of Pittsburgh Press.

Hou, J., Verspoor, M., & Loerts, H. (2016). An exploratory study into the dynamics of Chinese L2 writing development. *Dutch Journal of Applied Linguistics,* 5, 65–96.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning,* 63, 87–106.

Johansson, V. (2009). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics,* 53, 61–79.

Kanno, Y., & Cromley, J. G. (2015). English language learners' pathways to four-year colleges. *Teachers College Record,* 117, 1–44.

Kim, M., & Crossley, S. A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing,* 37, 39–56.

Kyle, K. (2020). The relationship between features of source text use and integrated writing quality. *Assessing Writing,* 45, *100467.*

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings and application. *TESOL Quarterly*, 49, 757–786.

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing,* 34, 12–24.

Kyle, K., Crossley, S. A., & Berger, C. (2018). The Tool for the Automatic Analysis of Lexical Sophistication Version 2.0. *Behavior Research Methods,* 50, 1030–1046.

Laufer, N., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics,* 16, 307–322.

McCarthy, P. M. & Jarvis, S. (2007). *vocd:* A theoretical and empirical evaluation. *Language Testing,* 24, 459–488.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods,* 42, 381–392.

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

McNamara, D. S., Graesser, A. C. McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Morris, L., & Cobb, T (2004). Vocabulary profiles as predictors of the academic performance of teaching English as a second language trainees. *System*, 32, 75–78.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly,* 47, 91–108.

Olson, C. B., Scarcella, R., & Matuchniak, T. (2015). *Helping English learners to write: Meeting common core standard, grades 6-12.* Teachers College Press.

Olson, C., & Land, R. (2007). A cognitive strategies approach to reading and writing instruction for English language learners in secondary school. *Research in the Teaching English,* 41, 269–303.

Ortmeier-Hooper, C., & Enright, K. A. (2011). Mapping new territory: Toward an understanding of adolescent L2 writers and writing in US contexts. *Journal of Second Language Writing,* 20, 167–181.

Pajares, F., & Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary Educational Psychology,* 20, 366–381.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13, 111–129.

Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26, 561–587.

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, 17, 18–34.

Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22, 217–230.

Ranney, S. (2012). Defining and teaching academic language: Developments in K-12 ESL. *Language and Linguistics Compass*, 9, 560–574.

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

Reid, J. (2006). "Eye" learners and "ear" learners: Identifying the language needs of international students and U.S. resident writers. In: P. K. Matsuda, M. Cox, J. Jordan, & C. Ortmeier-Hooper (Eds.), *Second-language writing in the composition classroom: A critical sourcebook* (pp. 76–88). Bedford/St. Martin's.

Salsbury, T., Crossley, S. A, & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, 27, 343–360.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly*, 22, 69–90.

Scarcella, R. (2003). Academic English: A conceptual framework. *The University of California Linguistic Minority Research Institute Technical Report 2003-1*, Retrieved from https://escholarship.org/uc/item/6pd082d4.

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education,* 12, 431–459.

Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing.* Cambridge University Press.

Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sanchez, E. (2014). Core Academic Language Skills (CALS): An expanded operational construct and a novel instrument to chart school-relevant language proficiency in per-adolescent and adolescent learners. *Applied Psycholinguistics*, 36, 1077–1109.

Uccelli, P., Dobbs, C. L., & Scott, J. (2012). Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication,* 30, 36–62.

Vande Kopple, W.J. (1994). Some characteristics and functions of grammatical subjects in scientific discourse. *Written Communication,* 11, 534–564.

Vidakovic, I., & Barker, F. (2010). *Use of words and multi-word unites in Skills for Life writing examinations. IELTS research reports 41*. IELTS Australia/British Council 7–14.

Vögelin, C., Jansen, T., Keller, S. D., Machts, N., & Möller, J. (2019). The influence of lexical features on teacher judgments of ESL argumentative essays. *Assessing Writing,* 39, 50–63.

West, M. (1953). *A general service list of English words.* Longman.

Yi, Y., & Kao, C., & Kang, J. (2018). Middle and high school writing. In J. I. Liontas & M. DelliCarpini. (Eds.). *The TESOL Encyclopedia of English Language Teaching.* John Wiley & Sons, Inc.

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, 51, 275–301.

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics,* 31, 236–259.

CHAPTER 4

Study 2. Syntactic and Lexical Features of Adolescent Multilingual Students' Academic Writing

Abstract

With a premium placed on academic writing in the U.S. secondary education context, adolescent writers are expected to develop advanced skills to analyze, interpret, and produce complex texts in a variety of content areas, which require proficiency in academic language. To enhance the academic literacy of developing multilingual writers and to meet their linguistic needs, it is crucial to identify the unique language features of their academic writing. This study seeks to examine syntactic and lexical features of the text-based analytical essays written by Spanish-speaking multilingual students (7^{th} – 12^{th} grades) from a public school in a western state of the United States. Employing manual sentence coding and quantitative measures of selected linguistic variables from Coh-Metrix, I analyzed student texts (N=86) to identify common linguistic patterns and to examine how these linguistic features predict writing quality. The findings reveal that sentence boundary issues, lack of syntactic variety, the underuse of sophisticated subordination common in academic writing, and low use of advanced vocabulary are common linguistic features in adolescent multilingual students' academic writing. The results of regression analyses show that syntactic complexity and lexical sophistication predict human-judged holistic writing quality score. Implications for pedagogy to address the linguistic needs of developing multilingual writers are discussed.

Keywords: Syntactic and lexical features, Academic writing, Text-based analytical writing, Adolescent multilingual students

Introduction

In the U.S. secondary education context, the significance of academic writing is underscored in the Common Core State Standards (CCSS). The Common Core State Standards for English Language Arts & Literacy in History/Social Studies, and Technical Subjects (CCSS-ELA) have placed a premium on academic writing, requiring middle and high school students to be able to "comprehend and evaluate complex texts across a range of types and disciplines" and "construct effective arguments and convey intricate and multifaceted information" (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010, p. 7). As early as 6th grade, students are expected to develop a claim about a complex topic and substantiate it with relevant examples and evidence using credible sources while maintaining a formal style of writing. They must demonstrate "the ability to analyze and interpret challenging texts and to write about those texts using academic discourse in extended pieces of writing" (Olson et al., 2015, p. 5). In addition, students are expected to use precise language, domain-specific vocabulary, and nuanced transitional devices to show relations between ideas. In short, the CCSS-ELA's emphasis on academic writing demands high levels of analytical writing proficiency and academic language fluency.

While the standards, requirements, and expectations are high, the majority of the students in U.S. secondary schools have low performance on national writing assessments (NAEP, 2012, 2019). For example, the results of the first computer-based writing assessment conducted by the National Assessment of Educational Progress (NAEP) indicate that only about 27% of students in 8th and 12th grades and only 1% of English Learners (ELs) in both grades perform at or above the proficient level (U.S. Department of Education, NCES, 2012). The NAEP report shows that the majority of the students in middle and high school in the United States struggle as academic

writers and in meeting the expectations set forth by the CCSS. The alarming 1% proficiency level for ELs indicates that the challenge is much greater for adolescent multilingual students who are still in process of developing their English language proficiency. As of Fall 2017, ELs made up about 10% of students attending public schools in the United States (U.S. Department of Education, NCES, 2020), and the number of ELs has been continuously increasing. Despite their growing rate in the U.S. secondary education context, adolescent multilingual students are an underexplored population in L2 writing research. Much of the research in L2 writing, more specifically on the linguistic features of L2 texts, has focused on postsecondary students (Ortega & Carson, 2010).

While the gaps in research remain, the gaps in pedagogy to address adolescent multilingual students' academic writing development continue to be prevalent. Applebee and Langer's (2011) study, which surveyed teachers across the United States and examined how writing is taught in 260 middle and high school classes in five different states, found that writing practice and instruction were minimal, with only 7.7% of class time spent on composing extended texts. Similarly, Graham et al. (2014) reported that the amount of writing practice and the language instruction students receive are not sufficient in a typical middle school classroom. Graham et al.'s (2014) study highlighted the discrepancy between writing practices in secondary school and writing standards as defined by the CCSS (Graham et al., 2014). Both studies suggest multilingual students in secondary school are not receiving sufficient writing practice and are not exposed to high-quality writing and language instruction in general. These gaps in pedagogy and research warrant a need to investigate the linguistic features of adolescent multilingual students' academic writing. Bunch et al. (2012) emphasized the importance of understanding "the use of language and literacy relevant to the challenges facing ELs in light of the [Common Core]

Standards" (p. 3). Such an understanding is vital in writing pedagogy that is geared towards helping developing multilingual writers in secondary school to learn the language conventions of academic written discourse.

Literature Review

The study is framed within the domain of academic writing and focuses on the language features of adolescent multilingual students' text-based analytical essays with an intent to determine the extent to which these students are progressing towards acquiring the target language features of academic writing. Academic written register, which is characterized by its highly informational purpose, impersonal style and absence of narrative features, exhibits salient patterns of language use and grammatical complexity that need to be acquired by students when composing within an academic context (Biber, 2006; Biber et al., 2011). Thus, standards-aligned classroom pedagogy geared towards facilitating adolescent multilingual students' acquisition of language skills that are necessary to meet the demands of academic writing need to consider both the language features of academic writing and the linguistic needs of striving writers.

The present study reviews the language features of academic writing and analyzes multilingual students' text-based analytical essays to identify common linguistic patterns. While reviewing the language features of academic writing helps us understand the highly specialized grammatical structures and lexical features of academic written register, a systematic examination of linguistic features of adolescent multilingual writers' texts can help us identify common linguistic patterns observed in those texts. Understanding of both the demands of academic writing and the linguistic challenges developing multilingual writers face to meet the demands can inform us about these students' linguistic needs. In what follows, I provide an overview of the linguistic features of academic writing, with an intent to identify common

features shared within and across disciplines and to determine language demands necessary to build proficiency in academic writing.

Academic Writing and its Linguistic Features

Academic writing, loosely defined as specialized writing in an academic context that demonstrates knowledge and proficiency in "certain disciplinary skills of thinking, interpreting, and presenting," is a complex, multi-dimensional task and process (Irvin, 2010, p. 8). Writing in an academic context is guided by and aligned with communicative purposes and language conventions sanctioned by the members of discourse communities (Swales, 1990). Regardless of the discipline in which it is situated, academic writing is marked by a well-informed and focused topic/argument, evidence-based development, textual coherence based on logical order and relations between ideas, and formal tone. To accomplish these higher-order tasks, academic writers heavily rely on their linguistic resources; namely, syntax and vocabulary are two fundamental linguistic resources used to construct meaning and communicate ideas. Although academic writing styles vary across disciplines, there are common linguistic features of academic written discourse shared among different disciplines. The expression and the development of complex ideas in an academic context exhibits common register features because of the similarity in purposes of academic genres (Schleppegrell, 2001). In other words, academic register is marked by a constellation of linguistic features, including syntactic and lexical properties, that are unique to formal written discourse.

Syntactic Features of Academic Writing. The grammatical structures of academic writing tend to be highly specialized (Biber et al., 2011). One prominent syntactic and stylistic feature that is common across various academic written genres is the application of and dependence on hypotactic syntactic constructions that are more sophisticated and less common

than those found in casual conversations. For example, Biber et al. (2011) found that nonfinite relative clauses, WH relative clauses (post-modifiers using *who, that, which, where, when)*, noun complement clauses with *that* and *to*, and "dependent clauses functioning as constituents in a noun phrase" are much more prevalent in academic writing (p. 28). On the other hand, finite dependent clauses with *if* and *because* conjunctions and finite complement clauses (verb + *that* clause as in "I think that ...") are more common in conversations than academic writing. Sophisticated forms of hypotaxis or subordination of clauses are common in academic writing for the dual purpose of achieving efficiency/economy and establishing a hierarchical order of ideas. Syntactic efficiency is achieved through embedding a subordinate idea within the main sentence rather than expressing it in a separate, independent sentence. Hypotactic syntactic construction also makes the logical relations and interdependence of ideas explicit, showing arrangements of clauses and phrases that are subordinate to and dependent on main clauses and phrases.

Other specific syntactic features of academic writing include embedded phrases (participial and absolute phrases, and apposition embedded within sentences), complex phrasal structures (noun and prepositional phrases), and hierarchical structure (phrase and clause subordination). For instance, post-modifying prepositional phrases, pre-modifying nouns, and attributive adjectives in noun phrases are common in academic writing (Biber et al., 2011). In academic writing, the subordination of clauses makes explicit the logical connections between ideas and show the hierarchical structure of clauses. In addition, the common use of apposition and participial phrases embedded in sentences is a way to efficiently combine related ideas. In general, academic writing is characterized by complex grammatical structures, interdependent ideas expressed in hierarchical structures, embedding and nesting of clauses and phrases. These

clause-combining strategies are used by academic writers to achieve efficiency and to build logico-semantic relationships.

Lexical Features of Academic Writing. Academic writing across various genres and disciplines also shares common lexical features. A linguistic feature that is conspicuous in academic writing is "the vocabulary of academic language, often described as comparatively large, precise, and formal" (Ranney, 2012, p. 563). Comparing the lexical features in spoken interactions and school-based texts, Schleppegrell (2001) notes that words used in academic writing tend to be specific and technical, and academic texts have higher lexical density. The density, precision, and technicality of academic lexicons are associated with informational density, technical concepts, and content-specific words that are required to convey complex ideas. Corpus-based research investigating the lexical features of academic writing has led to the development of vocabulary typologies. For example, Scarcella (2003) identified three major categories of academic vocabulary: (1) general words used across academic disciplines and in everyday situations; (2) academic words that are common across different disciplines; and (3) technical words found in specific academic fields. The words in the latter two categories tend to be highly specialized. Taken together, understanding these common linguistic features of academic writing is crucial in writing instruction that is geared towards developing multilingual learners' academic language and writing proficiency.

Linguistic Demands of Academic Writing

Academic writing requires the ability to analyze, interpret, synthesize, and explain complex ideas. Performing these higher-order analytical tasks successfully depends on linguistic resources and requires a higher level of language proficiency. Kellogg (2008) asserts, "Without knowledge [of language] being accessible and creatively applied by the writer, it remains inert

during composition and unable to yield the desired fluency and quality of writing" (p. 3). This is especially true at an advanced level of writing that is beyond what Bereiter and Scardamalia (1987) referred to as "knowledge telling," in which students rely upon summarizing what they have read. Thus, writing at the level of "knowledge transformation" that is characterized by a complex rhetorical problem-solving process and interpretation rather than summary demands advanced language skills and mastery of comprehensive language knowledge (Bereiter & Scardamalia, 1987).

To communicate well in an academic context, students need to have comprehensive knowledge of sentence structures, types, and functions. This includes the ability to coordinate and subordinate sentences, as well as the ability to make a variety of syntactic choices appropriate for the genre and audience. Further, syntactic knowledge is necessary to avoid sentence boundary errors. For example, the Analytic Writing Continuum (a rubric for literary analysis) developed by the National Writing Project (NWP) that connects U.S. educators across different states requires proficient writers to: (1) "demonstrate effective phrasing so that each sentence flows easily into the next"; (2) "include sentences that vary in structure and length, creating an extremely effective text"; and (3) construct sentences "that are consistently logical and clear so that the relationships among ideas are firmly and smoothly established" (National Writing Project, 2010, p. 2). Because academic writing demands complexity and variety in sentence structure, as well as conventionalized clause organization, developing writers need to learn formalized syntactic structures (Schleppegrell, 2001).

Academic writing also demands a large lexical capacity and in-depth knowledge of words, including their morphological, semantic, and pragmatic properties. While knowing more words is important, knowledge about the multi-dimensionality of vocabulary is vital in learning

to write in an academic context (Kieffer & Lesaux, 2012). In addition, students need to learn to make appropriate lexical choices considering the academic audience and genre. Because developing multilingual writers are constrained by various factors affecting language development, writing pedagogy in a secondary context needs to address ways of expanding students' linguistic resources and repertoires.

Syntactic and Lexical Features in L2 Writing Research

For adolescent multilingual writers who are developing their academic language proficiency, expressing complex ideas in words and sentences to form coherent writing is a challenging task. Thus, understanding how linguistic features might relate to writing quality and proficiency has been an important step in L2 writing research. In several studies of L2 learner corpora, syntactic and lexical features have been used to predict writing quality and proficiency based on the notion that more complex, skilled syntax and sophisticated lexical production lead to better writing quality (Bulté & Housen, 2014; Crossley et al., 2014; Crossley & McNamara, 2012; Ferris, 1994; Grant & Ginther, 2000; Guo et al., 2013; MacArthur et al., 2019; McNamara et al., 2010).

L2 writing research indicates that "higher quality writing generally contains more complex syntactic features" (Crossley, 2020, p. 422). For example, Lu (2010; 2011) found that L2 texts that show higher writing proficiency exhibited longer and more varied sentences. Furthermore, high quality student texts have greater clausal subordination and complex syntactic structures (Friginal & Weigle, 2014; Grant & Ginther, 2000), greater production of dependent clauses (Crossley & McNamara, 2014), greater phrasal complexity (Kyle & Crossley, 2018; Taguchi et al., 2013). Kyle & Crossley (2017) found that both large-grained indices of syntactic complexity and fine-grained syntactic complexity features were predictors of writing quality.

Like syntactic complexity, lexical features are predictive of writing quality (Crossley, 2020). A variety of lexical features, including lexical diversity (Crossley et al., 2014; Engber, 1995) and lexical sophistication (Crossley et al., 2013; Crossley & McNamara, 2012; Kyle et al., 2018; Kyle & Crossley, 2016; Laufer & Nation, 1995) have been found to have a significant correlation with writing quality and proficiency.

Most of the extant studies, however, focus on independent writing tasks (as opposed to source- or text-based writing) performed by students in a postsecondary context. In research examining linguistic features of student writing, the type of writing task and genre is an important consideration since different tasks and genres may demand different linguistic skills (Crossley, 2020). In addition, linguistic features of adolescent multilingual students' writing in a secondary education context continues to be an underexplored area in corpus-based research. This study focuses on the linguistic features of text-based analytical writing (a task and genre different from both independent and typical source-based writing tasks) of adolescent multilingual students in mainstream classrooms in a U.S. public education context. These students constitute a unique group with differing exposure to and experiences of navigating the complex landscape of US secondary education.

The Study

The goal of this study is to systematically examine syntactic and lexical features and their relations to writing quality in adolescent multilingual students' academic essays. This systematic examination is crucial in identifying the linguistic needs of this population within an academic context. This study is guided by the following research questions:

1. What syntactic and lexical characteristics do students' text-based analytical essays exhibit?

2. To what extent do syntactic and lexical features predict writing quality?

As this study focuses on adolescent multilingual students in the U.S. public secondary education context, it has important implications for writing research that deals with multilingual students' texts and writing pedagogy that is geared towards developing multilingual writers' academic literacy. The findings of this study can inform us of the adolescent multilingual students' linguistic needs, as well as the language skills that are necessary to meet the demands of academic writing.

The study's intent to inform the writing research and pedagogy in a secondary education context is particularly important since there is a dearth of research that focuses on the linguistic features of adolescent multilingual students' text-based analytical writing. Crossley (2020) pointed out that "the majority of information available to the field about the interactions between writing quality/development and linguistic features is derived from a single task (independent writing), shedding some confidence on the generalizability of the findings to other tasks" (p. 432). To my knowledge, no study has examined linguistic features of text-based analytical writing that asked students to analyze a theme of a nonfiction article, make a claim about the author's message, use evidence to support the claim, analyze the author's craft, and discuss the author's purpose. It is a writing task and genre that is distinct from typical source-based, argumentative writing. Unlike typical source-based writing that often integrates multiple sources, this type of writing task draws on a single source and represents an argument of interpretive analysis or literary explication. It is aligned with the CCSS-ELA's emphasis on developing the ability to analyze, argue, interpret, and use evidence to support ideas about the meaning and message of a literary work. It is also similar to the Advanced Placement (AP) English Literature and Composition Assessment (an examination offered by the College Board's Advanced

Placement Program) that many students striving for college entrance may take. The present study contributes to the extant literature by examining the linguistic features in lesser studied texts produced by underexplored multilingual student population.

Methodology

Sample and Setting

The unit of analysis in this study is 86 on-demand, text-based, analytical essays written by Spanish-speaking multilingual students from 7th to 12th grade in a large urban school district in California. Close to 80% of the students in this school district are Latinx, and about 33% are classified either as English Learner (EL) or Reclassified Fluent English Proficient (RFEP) based on 2016-2017 enrollment data. The essays were selected from students who participated in a reading and writing intervention and professional development program called the Pathway Project developed by the University of California, Irvine (UCI) Writing Project. The essays were written at the beginning of the 2016-2017 academic year before the commencement of the intervention. All 86 essays are written by Spanish-speaking multilingual students who are classified either as an EL (19%) or RFEP (81%). While 42% of the essays are written by students in middle school, the remaining 58% of the essays are written by high school students.

Writing Task and Holistic Scoring

At the beginning of the school year before the intervention, the participating students completed reading and analytical writing tasks over two class periods. During the first period, the students read a short newspaper article and completed a post-reading packet focusing on summarizing, visualizing, analyzing a theme and author's craft, and evaluating and reflecting on the purpose and significance of the article. During the second period, they wrote analytical essays responding to a writing prompt that asked them to select a theme in the article, make a

claim about the theme, and support their claim with evidence from the text. The writing task was designed to prompt students to write interpretive responses. The essays represent academic writing that requires students to analyze, interpret, argue, and integrate a source text.

All essays were scored holistically using a 6-point scale analytical writing assessment rubric developed by UCI Writing Project to assess the writing quality. The rubric for scoring the essays was based on those used to evaluate the essay portion of the CAHSEE (California Department of Education, 2008a), the California STAR 7 Direct Writing Assessment (California Department of Education, 2008b), and the NAEP (ACT, Inc., 2007). As indicated in the analytical writing assessment rubric, the writing quality was assessed based on the quality of interpretation, the clarity of thesis/claim, the organization of ideas, the integration of evidence, syntactic variety and sentence fluency, diction and word choice, and language conventions. Each essay was scored by two trained expert raters, scoring independently, and the two scores were added together for a final score between 2 and 12. If there was a discrepancy of two or more points between the raters, the essay was scored by a third rater. Readers for all but one paper were in exact agreement or differed by one point. The first- and second-reader scores were highly reliable with an alpha of 0.87. Raters agreed within a single score point or better for 99% of the papers; 56% of the papers had exact agreement between the two readers.

Corpus Description and Analysis

The essays in the corpus vary in length. The average word count of essays in the full sample is 198.7 (SD = 96.95), and the shortest essay is 42 words long while the longest essay has 443 words. Table 1 provides descriptive details, including average paragraph, sentence, and word counts and lengths by students' grade levels. The average length of essays increases as the grade levels advance, suggesting that students in higher grades write relatively longer essays compared

to those in lower grades, with an exception that the average length of 10th grade students' essays is lower than that of 9th grade essays. Similar trends are observed for paragraph and sentence counts. In contrast, the average length of paragraphs, sentences, and words do not follow the similar upward trend. The average length of paragraphs written by 7th grade students is higher than that of 12th grade essays, suggesting that students in lower grades have longer paragraphs with fewer paragraph divisions, compared to essays in higher grades. Essays in lower grades tend to consist of one paragraph. With respect to sentence length, the average number of words per sentence in the full sample is 18.7 (SD = 6.7); the essays in 9th grade have the highest mean score (M = 20.4; SD = 7.2) while essay in 10th grade has the lowest mean score (M = 15.9; SD = 5.0) in number of words per sentence.

Table 4.1

Descriptive Statistics: Paragraph, Sentence, and Word Counts

	Full sample		7th Grade		8th Grade		9th Grade		10th Grade		11th Grade		12th Grade	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Paragraph count	2.4	1.4	1.4	0.8	2.3	1.6	2.9	1.6	2.9	1.3	2.3	1.1	3.6	1.4
Sentence count	11.4	6.0	8.6	4.9	11.1	6.3	10.9	6.1	12.3	5.5	13.0	5.4	14.8	7.7
Word count	198.7	96.9	144.8	71.1	191.2	90.9	212.3	115.6	194.2	111.9	225.9	80.6	255.4	91.8
Paragraph length (sent)	5.8	3.9	7.1	4.0	5.9	2.8	4.3	1.8	5.6	5.2	7.0	5.3	3.9	0.9
Sentence length (word)	18.7	6.7	18.4	5.7	19.3	8.5	20.4	7.2	15.9	5.0	18.7	6.1	19.4	6.2
Word Length (letter)	4.3	0.3	4.3	0.3	4.3	0.3	4.2	0.3	4.3	0.2	4.2	0.3	4.2	0.3
	N=86		n=16		n=19		n=15		n=13		n=15		n=8	

While the full sample of 86 essays were used for the manual sentence coding to identify syntactic patterns, I excluded essays that are too short and used a selected sample for Coh-Metrix (McNamara et al., 2014) analysis. From the full sample of 86 essays, 16 essays were excluded

for Coh-Metrix analysis for two reasons. First, essays that are too short (less than 100 words) were not included, considering the fact that the suggested text length for Coh-Metrix analysis is 100 to 1000 words. Second, essays that are predominantly comprised of large chunks of quotes and sentences that are copied verbatim from the reading text were not included in Coh-Metrix analysis. The quoted and copied sentences have a confounding effect since the purpose of the study concerns how students construct sentences and use words in their own writing. Thus, the selected sample for quantitative analysis and Coh-Metrix measures consisted of 70 essays.

Procedures

To analyze the syntactic and lexical features of these essays, a combination of manual sentence coding and a computational tool Coh-Metrix were used. All sentences in uncorrected essays (N=86) were coded by two human raters both of whom have extensive experience teaching university-level writing and composition courses. Table 4.2 exhibits the sentence coding procedure that specifies the sentence categories and classifications with examples representing each classification or subcategory. The raters coded sentences based on 1) sentence structures — simple, compound/coordinated, complex/subordinated, and compound-complex, 2) types of clauses (i.e., finite complement clause, finite adjective clause, finite adverbial clause, nominal clause) and phrases (i.e., participial phrase, absolute phrase, appositive), and 3) sentence boundary issues — fragment, run-on, and faulty sentence (incomprehensible and structurally-flawed sentences that do not conform to the rules of English grammar). Fragment, run-on, and faulty sentences were classified as unconventional sentences as these syntactic structures are not conventional in academic written discourse. Finite complement clauses (*I think...*, *I believe...*) were classified as a simple structure (Table 2). Inter-rater reliability (IRR) analysis was performed to assess the degree the coders consistently classified sentences into the suitable

categories. The coders had substantial agreement (93.2%) in categorizing sentences, with a modest amount of variance mainly with discrepancy between faulty and run-on sentences.

Table 4.2

Sentence Coding Procedure

Sentence Categories	Subcategories/Classifications	Examples
Simple structure	Simple sentence (1 clause)	The man in the Water has his own natural powers.
	Finite complement clause (verb + *that* clause)	I think that the theme is courage and bravery.
Compound sentence	Coordinated clauses (2 or more clauses)	This man lost the battle but he was able to save 5 other men in the process.
Complex sentence	Subordinate clause (2 or more clauses): Finite adverbial clauses	When something tragic happened, they just move along with there business and forget the past.
	Finite noun modifier clauses	We just sit and watch everything that is happening.
Unconventional sentence	Fragment (incomplete sentence)	To just ignore the cruelties of the earth's nature.
	Run-on (comma splices are classified as run-on)	Haiti people were poor they had to do what they need everyone will help people if they were injured or had been stuck.
	Faulty sentences (sentences that have semantic and structural problems)	While sinking a man risking his life to save other knowing his death will come soon.

In addition to the manual coding for measures of syntactic features, the computational tool Coh-Metrix was used to obtain quantitative syntactic complexity measures. Coh-Metrix provides a wide variety of language and discourse measures (McNamara et al., 2014). I selected the indices that are pertinent to syntactic and lexical features (Table 3). The selection of Coh-Metrix measures was informed by the findings of previous studies that examined linguistic

features of students' writing. For syntactic features, I included sentence length, syntax similarity, left embeddedness (number of words before main verb), and number of modifiers per noun clause. Sentence length and variety were found to be important indicators of writing quality based on the findings that proficient writers produced longer and more varied syntactic structures (Lu, 2011; Ortega, 2003). Phrase-level complexity was also found to be an important indicator of writing quality (Guo et al., 2013; Kyle & Crossley, 2018; Taguchi et al., 2013). Kyle & Crossley (2017) reported both large-grained and fine-grained syntactic measures were significant predictors of writing quality. Thus, I included sentence length and syntax similarity as large-grained, clause-level measures and left embeddedness and number of modifiers per noun clause as fine-grained, phrase-level measures.

For lexical features, I included a lexical diversity measure, specifically the Measure of Textual Lexical Diversity (MTLD), and lexical sophistication measures, including word length, word frequency, and five psycholinguistic measures: age of acquisition, word familiarity, concreteness, imageability, and meaningfulness. Although Coh-Metrix provides several lexical diversity measures, such as type-token ratio and *vocd*, MTLD was chosen for two reasons. First, type-token ratio is confounded by text length as longer texts have more words that are repeated (Jarvis, 2013; McCarthy & Jarvis, 2007). MTLD attempts to overcome this issue by using estimation algorithms (McNamara et al., 2014). Second, McCarthy & Jarvis (2010), in their validation study, found that "MTLD performs well with respect to all four types of validity and is, in fact, the only index not found to vary as a function of text length" (p. 381). Similarly, Zenker & Kyle (2021) reported MTLD was stable in short student texts. When it comes to lexical sophistication, there is no agreed-upon method of operationalizing it; however, previous studies indicate that higher quality academic writing contains longer and less familiar words

(Crossley et al., 2011), greater use of infrequent, advanced, and academic words with higher scores of age of acquisition (Maamuujav, 2021; McNamara et al., 2010); less concrete, imageable, and meaningful words (Crossley et al., 2014; Crossley & McNamara, 2012; Salisbury et al., 2011). Table 3 shows the Coh-Metrix indices selected for this study.

Table 4.3

Coh-Metrix Indices Used in the Study

Linguistics Features	Measures/Indices
Syntactic Features	
Large-grained length feature	Sentence length (number of words)
Syntactic variety	Sentence syntax similarity (across paragraphs)
Phrase-level complexity	Left embeddedness
Phrase-level complexity	Number of modifiers per noun phrase
Lexical Features	
Length feature	Word length (letters)
Lexical Diversity	Measure of Textual Lexical Diversity (MTLD)
Lexical Frequency	CELEX word frequency (content words)
Psycholinguistic properties	Age of acquisition (content words)
	Familiarity (content words)
	Concreteness (content words)
	Imageability (content words)
	Meaningfulness (content words)

The initial phase of the study involved preparing the learner corpus for syntactic and lexical analysis. This included typing the essays since all of them were handwritten and coding sentences. For Coh-Metrix analysis, I corrected spelling errors and deleted sentences and large chunks of phrases (three or more consecutive words) that were either copied verbatim or quoted from the text to which students responded. The spelling errors were corrected in order to obtain accurate measures of syntactic and lexical features for Coh-Metrix output. Sentences and large phrases that were quoted and copied from the literary nonfiction article were deleted from the essays prior to the Coh-Metrix analysis because they reflected the syntactic and lexical features

of the source material rather than students' language use. The data cleaning process underwent three rounds of reviewing each essay to make sure the texts were clean and satisfied Coh-Metrix criteria. This was important since "dirty" texts with spelling errors and odd symbols undermine Coh-Metrix's validity (Dowell et al., 2016). I used the version of Coh-Metrix designed for researchers to process a batch of texts, but I also randomly selected 10 essays to run through the publicly available version that processes individual texts to check if the measures of the individual essays were consistent with scores obtained using the researcher version. The scores of the 10 essays were consistent.

Analytic Strategies

To analyze the syntactic and lexical features and their relation to writing quality, descriptive statistics, Pearson correlations, and multiple regression analyses were conducted. The descriptive statistics was used to answer the first research question to observe and identify the common syntactic and lexical features using measures derived from manual sentence coding and Coh-Metrix. The manual sentence coding, which was conducted with uncorrected texts in the full sample (n=86), was informed by the Systemic Functional Linguistic (SFL) approach (Halliday, 1994) that posits certain patterns of language use within a specific discursive context can be detected through a systemic approach and functional analysis of text construction. Since the intent of the study was to examine the language use and detect language patterns in students' written texts, Halliday's (1994) systemic approach and functional grammar that deals with structural features related to English phrases, clauses, and sentences served as an analytic strategy that guides this study.

The second research question investigates the relation between the linguistic features and writing quality. To address this, Pearson Correlations were first conducted between all syntactic

and lexical variables and the human-judged holistic score of writing quality. All predictor/independent variables and outcome variable of writing score were checked for normality and multicollinearity. The variance inflation factor (VIF) analysis was used to check multicollinearity. Following these procedures, multiple regression analyses were conducted in steps to examine the relationship between the syntactic and lexical features and writing quality. Since multiple regression shows combined and independent contributions of predictors (Kim, 2017), this analytical strategy was used to examine the extent to which the syntactic and lexical features independently and together predict the holistic writing quality score given by two human raters. The units of analysis, measures used in the study, and the analytical approaches that are described above are summarized and presented by research questions in the following Table 4.4 of data matrix.

Table 4.4

Data Matrix

Research Questions	Measures	Analytical Approach
What syntactic and lexical characteristics do students' text-based analytical essays exhibit?	Syntactic measures derived from manual coding • Simple sentences (percentage) • Compound sentences (percentage) • Complex sentences (percentage) • Faulty sentences (percentage) Coh-Metrix syntactic indices • Sentence length • Sentence syntax similarity • Left embeddedness • Number of modifiers per noun phrase Coh-Metrix lexical indices • MTLD • Word frequency • 5 psycholinguistic measures	• Human coding of sentences informed by Systemic Functional Approach (Halliday, 1994) • Descriptive statistics
To what extent do syntactic and lexical features predict writing quality?	• Holistic score of writing quality (dependent variable) • Syntactic features (predictors) • Lexical features (predictors)	• Correlation analysis • Multiple regression • Multicollinearity diagnostics (VIF)

Results

The goal of the study was (1) to investigate syntactic and lexical features in the adolescent multilingual students' text-based analytical essays, and (2) to analyze the extent to which the identified syntactic and lexical features predict holistic writing quality score.

1. Syntactic and Lexical Features in Students' Text-based Analytical Essays

1.1. Syntactic Features

Several common syntactic features at a clause level were identified based on manual coding of sentences. These include patterns of unconventional sentences, repeated use of basic coordination and subordination (finite complement, finite adverbial, and finite noun modifier clauses) that are common in colloquial discourse, and underuse of non-finite clauses (participials, absolutes, and appositives) that are typical in academic written discourse. The results of the systematic analysis based on manual coding of sentences revealed that 33% of all sentences in the full sample (N = 86) were categorized as unconventional sentences with boundary issues and other structural problems, making up the most frequent pattern. Unconventional sentences include run-ons and comma splices (14.76%), fragments or incomplete sentences (6.27%), and faulty sentences that do not conform to the rules of standard English grammar (11.93%). Simple structures that include both one-clause simple sentences and finite complement clauses (verb + *that* clause and verb + Ø complementizer) were the next prominent pattern, accounting for 23.86% of the total number of sentences. Among subordinated sentences, which constituted 19.21% of all sentences, finite adverbial clauses (11.02%) and finite noun modifiers clauses (7.18%) were common patterns. Nonfinite clauses were rare in student writing, with nonfinite noun modifier clauses accounting for only 1.01% of the sentences. Coordinated clauses made up about 11% of all sentences (Table 4.5).

Table 4.5

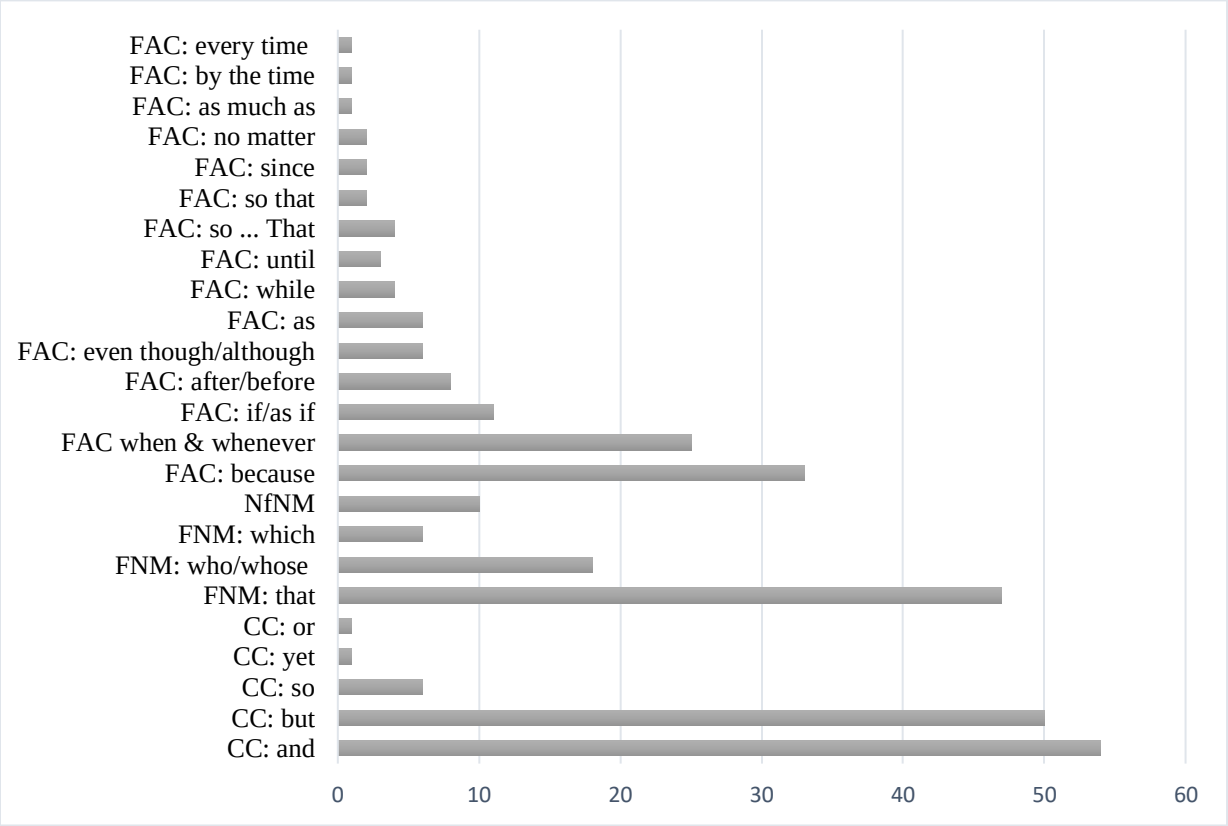
Syntactic Patterns Based on Manual Sentence Coding (Full Sample)

Sentence categories	Subcategories	Percentage Breakdown
Unconventional sentences with boundary issues (33%)	Run-on sentences and comma splices	14.76%
	Faulty sentences	11.93%
	Fragments or incomplete sentence	6.27%
Simple structures (24%)	One-clause simple sentences	18.1%
	Finite complement clauses	5.76%
Complex sentences with subordinated clauses (19%)	Finite adverbial clauses	11.02%
	Finite noun modifier (*that, which, who* relative clauses)	7.18%
	Nonfinite noun modifier clauses (-*ing* and -*ed* clauses)	1.01%
Compound sentences with coordinated clauses (11%)	Coordinated sentences with *and* conjunction	5.46%
	Coordinated sentences with *but* conjunction	5.06%
	Coordinated sentences with *so, yet, or* conjunctions	0.81%

Clause-level complexity in syntax was further investigated by examining the types of clauses and the connectives that were used to combine clauses (Figure 1). The functional classification of clauses indicated the use of basic coordinating conjunctions and a limited variety of subordination. Figure 4.1 shows the incidences of types of clauses and connectives used in the full sample. As exhibited in Figure 4.1, coordinated sentences with *and* and *but* conjunctions were the most frequent patterns with 10.52% of all coordinated and subordinated sentences. In addition, the three most frequently used subordinating clauses were finite noun modifier (FNM) clauses with a relative pronoun *that,* finite adverbial clauses (FAC) with a causal connective *because,* and FAC with a temporal connective *when.* These three structures alone constituted more than 50% of the subordinated clauses or complex sentences. This indicates a lack of syntactic variety in the text-based analytical writing, as well as a paucity of sophisticated clausal embedding with nonfinite clauses that are common in academic writing.

Figure 4.1

Types of Clauses and Connectives Used (full sample)



Note. FAC=Finite Adverbial Clause, NfNM=Nonfinite Noun Modifier, FNM=Finite Noun Modifier, CC=Coordinated Clauses.

For the quantitative analysis, I used four Coh-Metrix measures of syntactic features: sentence length, sentence similarity across paragraphs, left embeddedness, number of modifiers per noun clause. Table 4.6 shows the mean score and the standard deviation of the four syntactic measures obtained by Coh-Metrix by grade level. The mean scores of left embeddedness (M = 3.86; SD = 1.64) and number of modifiers per noun clause (M = .69; SD = .16) were substantially lower than the grade-level estimates based on the Degrees of Reading Power (DRP) adjusted according to grade bands used within the CCSS (see McNamara et al., 2014 for Coh-Metrix norms by grade level). On these two syntactic complexity measures, the mean score of the full sample was at around the fourth and fifth grade level. For example, the mean score of left

embeddedness based on the norms of grade level estimate is 4.078 (SD = 1.7) for grades 6-8, and 4.644 (SD = 2.335) for grades 9-10, and 5.512 (SD = 2.430) for grades 11 and up (McNamara et al., 2014). In the sample of the study, however, the mean scores of left embeddedness were substantially lower in all grades.

In addition, sentences in the sample of text-based analytical essays had fewer words before the main verb, showing less syntactic complexity. A similar trend is observed in the number of modifiers per noun clause, suggesting the essays generally have simpler noun clause structures with fewer modifiers. On the other hand, sentence similarity across paragraphs were relatively stable across grade levels, as well as in comparison to grade-level estimates. This indicates that the essays have structurally diverse sentences. Sentence length varied by grade level, with 8th (M = 21.28; SD = 7.69) and 9th (M = 21.26; SD = 6.24) grade essays having larger mean scores of sentence length and 10th grade essays with the smallest mean score (M = 15.53; SD = 4.53). It is important to note that a lot of the long sentences that the students produced were run-on sentences or faulty sentences (common patterns based on the manual coding). For example, the sentence below is 46-words long, but because it is a run-on sentence, it may not be considered a sophisticated or complex sentence construction.

A sample sentence from a student text:

> [In this essay I will talk about the Man in the water] [he is not just an ordinary person] [he was a hero] [why you may ask] [because he risk his life for other people's life] [not many people do that] [he saved 5 people's life.]

The sample sentence has 7 clauses (indicated by brackets) put together without any conjunctions, forming a long run-on sentence. This indicates that the number of clauses within the sentence and the length of the sentence may not necessarily indicate syntactic complexity as the clauses

were not combined with sentence embedding techniques strategically used to show both logical connectedness and hierarchy.

Table 4.6

Syntactic Features by Grade Level: Selected Sample

	Selected sample	7th Grade	8th Grade	9th Grade	10th Grade	11th Grade	12th Grade
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Syntactic Complexity							
Number of words before the main verb	3.86 (1.64)	2.59 (0.97)	3.51 (1.69)	4.50 (2.17)	3.29 (0.94)	4.66 (1.48)	4.50 (0.97)
Number of Modifiers per noun clause	0.69 (0.16)	0.65 (0.19)	0.66 (0.18)	0.65 (0.08)	0.66 (0.10)	0.82 (0.17)	0.69 (0.13)
Sentence similarity across paragraphs	0.09 (0.03)	0.09 (0.04)	0.09 (0.04)	0.08 (0.03)	0.09 (0.02)	0.09 (0.02)	0.08 (0.03)
Sentence Length	19.15 (6.24)	18.20 (6.63)	21.28 (7.69)	21.26 (6.24)	15.53 (4.53)	19.20 (5.98)	18.47 (4.20)
Observations	70	11	13	13	11	14	8

Putting it all together, I now summarize the key findings of syntactic features below. In general, the text-based analytical essays of adolescent multilingual students in the sample exhibit the following:

- Unconventional sentences (run-ons, fragments, faulty sentences) and simple structures make up the most frequent patterns.
- Repeated use of basic and commonly used coordinating (*but, and)* and subordinating (*that, because, when)* conjunctions.
- Underuse of varied and sophisticated subordination that are common in academic writing.
- Fewer words before the main verb showing less syntactic complexity (approximately two or more grades below the grade level norm).
- Simpler noun clause structures (two or more grades below the grade level norm).

These patterns indicate syntactic features common in colloquial discourse and point to the challenges of making syntactic choices based on the language conventions of academic written discourse. To communicate well in an academic context, adolescent multilingual writers need to have comprehensive knowledge of sentence structures, types, and functions. This includes the ability to coordinate and subordinate sentences, as well as the ability to choose among a variety of structures and make syntactic choices appropriate for the genre and audience.

1.2. Lexical Features

Lexical characteristics are based on Coh-Metrix indices of lexical diversity (LD), lexical frequency, and five psycholinguistic measures of words — age of acquisition, content word familiarity, concreteness, imageability, and meaningfulness. On the LD measure, the essays typically have low scores. This indicates that the text-based analytical essays in the sample have less lexical diversity and more repetition of content and function words. The LD score was lower than estimated grade-level norms. For example, the following excerpt from student writing shows a repetitive use of the phrase "*the man in the water*," leading to a repetition of both content words (*man, water*) and function words (*the, in*) within a paragraph.

> I think the theme of "**The man in the water**" was about cruel and charity. **The man in the water** was more kind than nature. **The man in the water** was willing to give up his life for others and save any survivors from the plane crash. I think what the author was trying to show us two different acts of two different people like the act of the **man in the water** and nature. **Man in the water** put his life at risk to try to do something real good.

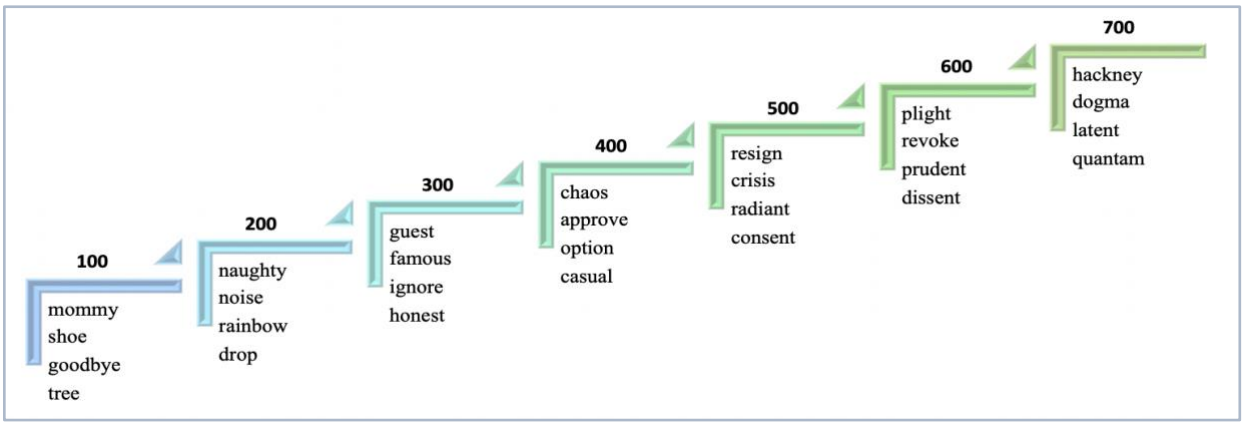
Repetitions of the bolded words such as this, which were pervasive in the sample of multilingual student texts, not only lead to less lexical diversity but signal the writers' lack of lexico-grammatical awareness.

The word frequency indices indicate that the essays containing more frequent words show less lexical sophistication. Word frequency measures have been used to analyze lexical sophistication based on the notion that less proficient learners use more frequent words and, thus, the use of less frequent words show more sophistication. In the essays analyzed in this study, word frequency for content words (M = 2.57; SD = .15) and all words (M = 3.20; SD = .10) had substantially higher frequency scores compared to grade level estimates, suggesting the use of more frequent words. This is consistent with the findings of previous studies that demonstrated the use of more frequent words in low-proficient students' essays (McNamara et al., 2010; Crossley et al., 2011).

In addition to LD and word frequency measures, five psycholinguistic measures of words were examined. Age of acquisition for content words is a psycholinguistic measure of words that is based on the notion that "some words appear in children's language earlier than others" (McNamara, 2014, p. 74). Higher scores (600-700) of age of acquisition indicate academic and technical words that are not common in everyday speech. In the sample of essays, the mean age-of-acquisition score was 308.71 (SD = 27.68), indicating that the students' vocabulary level, on average, is in the lower tiers of vocabulary (Beck et al., 2002), progressing from tier 1, which includes 100-200 level words based on Bristol norms, to tier 2, which includes 300-400 level words (see Figure 4.2 for comparison based on Bristol norms). This is consistent with the mean score of content word familiarity, which is based on "a rating of how familiar a word seems to an adult" (McNamara, 2014, p. 74). The scores of the essays on content word familiarity indicated a frequent use of familiar and basic words, on average. The three other psycholinguistic measures — word concreteness (M=376.40; SD=26.76), word imageability (M=410.19; SD=25.61), and word meaningfulness (M=446.07; SD=20.19) — were comparable to grade level estimates.

Figure 4.2

Sample Words Representing AoA Scores Based on Bristol Norms



Taken together, the analysis of lexical features based on quantitative measures of lexical diversity, lexical frequency, and psycholinguistic properties of words suggest the following:

- Unnecessary repetitions of content and function words, indicating less lexical diversity.
- Dependence on more frequent, familiar, and basic words that are common in spoken interactions and early stages of vocabulary acquisition.
- Low use of academic and sophisticated vocabulary.

These lexical features common in the text-based analytical essays in the sample of the study point to students' lexical challenges when engaging in academic writing, The findings further suggest the need for adolescent multilingual writers to build lexical capacity for academic writing, to develop knowledge and awareness of multidimensionality of words, and to learn to make lexical choices considering the genre conventions and audience expectations.

2. Relations between Syntactic and Lexical Features and Writing Quality

This section addresses the second research question that focuses on the relation between the identified linguistic features and holistic writing quality. Correlations between the writing quality and six of the linguistic features were significant in the selected sample (n=70): percentage of complex and compound sentences (r = .61, p <. 001), percentage of faulty

sentences (r = -.49, p <. 001), age of acquisition for content words (r = .34, p = . 005), word concreteness (r = -.30, p = . 011), imageability (r = -.37, p = .002), and meaningfulness (r = -.32, p = .006). All other syntactic and lexical features were not significantly correlated with writing quality. Table 4.7 exhibits a correlation matrix, both significant and nonsignificant correlations of all identified linguistic features with the holistic writing quality score.

Table 4.7

Correlation Matrix: Syntactic and Lexical Features and Holistic Score

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Holistic writing quality score	5.74	1.80	—														
2. Simple sentences	20.61	16.92	-.09	—													
3. Complex & compound	34.64	24.41	.61***	-.14	—												
4. Unconventional sentences	44.74	27.65	-.49***	-.49***	**-.80*** **	—											
5. Sentence length	19.15	6.23	.02	-.59***	-.16	.50***	—										
6. Sentence similarity	0.92	0.03	.18	.39***	.23	-.44***	-.51***	—									
7. Left embeddedness	3.88	1.62	-.13	-.17	-.06	.15	.25*	-.06	—								
8. Number of modifiers	0.69	0.16	-.10	.15	.07	-.15	.03	.15	.44***	—							
9. Word length (letters)	4.19	0.24	.12	.06	.15	-.16	-.16	.10	.04	.27*	—						
10. Diversity MTLD	62.55	18.36	.22	.01	.20	-.18	-.01	-.09	.10	-.02	.21	—					
11. Word Frequency	2.57	0.15	.01	-.12	.04	.04	-.03	-.04	-.32**	-.51***	-.51***	-.30*	—				
12. Age of acquisition	308.57	26.92	.34**	.14	.35**	-.40***	-.15	.11	-.03	.09	.18	.10	-.20	—			
13. Word Familiarity	583.42	6.02	-.13	-.08	-.03	.08	-.03	.04	-.27*	-.29*	-.30*	-.30*	.65***	-.45***	—		
14. Word Concreteness	376.40	26.76	-.30*	.16	-.29*	.16	.02	.03	.28*	.48***	.09	-.17	-.49***	-.28*	-.17	—	
15. Word Imageability	410.19	25.61	-.37**	.20	-.37**	.20	.00	.01	.30*	.45***	.12	-.12	-.53***	-.31**	-.24*	**.96*** **	**— **
16. Word Meaningfulness	446.07	20.19	-.32**	.22	-.34**	.17	-.03	.08	.21	.34**	.13	-.27*	-.41***	-.28*	-.16	**.82*** **	**.87*** **

Note. n=70. * p<0.05. ** p<0.01. *** p<0.001. simple, complex & compound sentences, and faulty sentences are in percentages.

All syntactic and lexical variables were checked for multicollinearity to see whether any of the variables were highly correlated. This was important since highly correlated variables prevent the regression model from accurately and precisely estimating the relationship between the dependent variable and the independent variables (Wooldridge, 2006). The results of Pearson correlations showed that complex and compound sentences were highly correlated with unconventional sentences (r = .796); word concreteness was highly correlated with imageability (r= .957*)* and meaningfulness (r= .824); and meaningfulness was highly correlated with imageability (r= .869). Thus, I excluded unconventional sentences, word concreteness, and meaningfulness from the regression analyses. When I computed the variance inflation factors (VIF) on the remaining variables, the VIF values for all remaining variables were between 1.47 and 3.61 and all tolerance levels were beyond the .2 threshold, indicating that the multicollinearity was not an issue (Gordon, 2010).

In regression analyses, I included six measures of syntactic features in model 1, six measures of lexical features in model 2, and three syntactic and lexical measures that were significantly correlated with writing score in model 3 (Table 4.8). In regression model 1, only syntactic features were regressed against the writing quality score. The six syntactic features used in this regression model significantly predicted writing quality, $F(6, 63) = 8.23$, $p < .001$, $r^2 = .44$. The six syntactic features together accounted for 44% of the variance in the human-rated evaluation of writing quality. In regression model 2, six measures of lexical features were regressed against the writing quality score. These six lexical features together accounted for 23% of the variance in the human-rated evaluation of writing quality, $F(6, 63) = 3.07$, $p = .011$. Model 3 combined only the syntactic and lexical variables that had significant correlation with writing quality, excluding the three variables that were dropped due to multicollinearity. The

three combined variables (percentage of complex and compound sentences, age of acquisition, and word imageability) accounted for 41% of the variance in the holistic score of writing quality, $F(3, 66) = 15.13$, $p < .001$. To sum up, both syntactic and lexical features were found to be crucial indicators of writing quality. These results based on the three regression models are exhibited in Table 4.8 below.

Table 4.8

Regression Models Examining Linguistic Features and Writing Quality

	B	*SE*	*t*	*p*	R^2	$F(6, 63)$
Model 1: Syntactic Features					.44	8.23***
Constant	2.99*	1.29	2.32	.024		
Simple sentences	0.01	0.01	1.04	.304		
Complex & compound sentences	0.05***	0.01	6.25	.000		
Sentence length	0.08*	0.04	2.18	.033		
Sentence similarity	8.13	6.43	1.26	.211		
Left embeddedness	-0.06	0.12	-.47	.641		
Number of Modifiers	-1.92	1.24	-1.55	.126		
	B	*SE*	*t*	*p*	R^2	$F(6, 63)$
Model 2: Lexical Features					.23	3.07*
Constant	18.92	30.75	0.62	.541		
Word length	0.52	0.98	0.53	.599		
MTLD	0.01	0.01	0.93	.355		
Word frequency	-0.42	2.45	-0.17	.864		
Age of acquisition	0.01	0.01	1.20	.235		
Word familiarity	-0.01	0.05	0.76	.761		
Word imageability	-0.02*	0.01	0.04	.044		
	B	*SE*	*t*	*p*	R^2	$F(3, 66)$
Model 3: Combined					.41	15.13***
Constant	6.48	4.06	1.60	.115		
Complex & compound sentences	0.04	0.01	4.90	.000		
Age of acquisition	0.01	0.01	1.03	.305		
Word imageability	-0.01	0.01	-1.42	.160		

Note. n=70. * p<0.05. ** p<0.01. *** p<0.001.

Discussion

This study systematically analyzed the linguistic features of adolescent multilingual students' text-based analytical essays in order to identify the unique language features of their academic writing and to understand how these linguistic features predict their writing quality. The findings of the study suggest that linguistic features substantially contribute to writing quality and that syntactic complexity and lexical sophistication play an important role in composing academic texts that require higher levels of analytical skills. Sentence boundary issues, lack of syntactic variety, the underuse of sophisticated subordination to show connections between ideas, and low use of advanced vocabulary are common features observed in adolescent multilingual students' academic writing. These common patterns exhibited in students' writing are not in accord with the highly specialized linguistic features and language conventions of academic prose (see Biber 2006; Biber et al., 2011 for grammatical features that are characteristic of academic writing). Rather, these patterns reflect language features that are common in informal conversations.

These results are consistent with and correspond to what Shaughnessy (1977) contends in *Errors and Expectations;* developing writers' "unfamiliarity with certain features of the code that governs formal written English" may interfere with and impinge on their writing quality (p. 45). This, particularly, may be the case for developing multilingual writers who have not fully acquired and internalized the language features of the formal code that govern academic writing. Syntactic and lexical constraints can inhibit fluency and thwart a writer's ability to express complex thoughts when engaging in academic writing that heavily depends on linguistic resources. Thus, developing multilingual writers' ability to produce highly analytical academic texts may be undermined if they have not yet learned the specialized language features of

academic writing. In this vein, Kellogg (2008) noted that students' linguistic choices are limited when their knowledge of syntax, discourse, and language in general is not available and easily accessible. Due to their unfamiliarity with the language conventions of academic writing, developing multilingual writers are likely to draw on the linguistic resources that are readily available to them, especially in on-demand writing situations. The findings of this study demonstrate the need for adolescent multilingual students to develop their language resources and expand their linguistic repertoires as they are required to engage in complex academic writing tasks.

Although the results of the present study contribute to our understanding of the linguistic needs of adolescent multilingual students, there are several limitations that warrant discussion. First, the sample in this study consists of a small sample of students who represent only Spanish-speaking multilingual students from one school district in California. This is an important limitation because adolescent multilingual students constitute a diverse group with unique backgrounds and differing linguistic needs. In order to see whether the linguistic needs identified in this study pertain to other language-minority students, a much larger sample that represents diverse L1 language backgrounds is needed. Future research should address this limitation and use a more inclusive sample that represents a more linguistically heterogeneous mixture of students.

Second, the analysis based on manual coding of syntactic features focused on sentence boundary issues and clause-level features. Biber et al. (2011) identified complex noun phrase structures in academic texts and suggested that phrasal complexity is an important feature of academic writing. For phrase-level features and lexical analysis, I used Coh-Metrix variables that are based on incidences, occurrences, and frequencies. While the quantitative analysis of phrasal

complexity and lexical features provides us with an insight into general patterns of language use, qualitative measures of phrasal complexity and vocabulary use will provide a more comprehensive picture of syntactic and lexical features in students' academic writing.

Finally, the unit of analysis in this study is student texts that were produced in an on-demand writing situation. This, in fact, is a common trend among the previous studies that examine linguistic features of student writing. While it is important to examine what students can do during an on-demand situation because standardized, timed writing assessment is a large part of the public education system and oftentimes the basis for college admission, it is important to recognize that writing at the level of knowledge transformation involves a complex rhetorical problem-solving process that requires a significant amount of time and multiple opportunities for revision. I acknowledge that many of these adolescent multilingual students have the potential to perform better when given sufficient time to write and multiple opportunities to revise their work.

Pedagogical Implications

The complexity and academic nature of writing tasks that are emphasized in the Common Core State Standards for English Language Arts (CCSS-ELA) require students to analyze, interpret, and produce a variety of complex texts. While the standards have placed a premium on academic writing and set a high bar for students, how these ambitious goals may be achieved are not often clear. Importantly, the common core standards explicitly state that defining the range of institutional and instructional supports that are available and appropriate for English language learners is "beyond the scope of the Standards," yet all students are held to and must meet the same high standards (California Common Core State Standards, 2013, p. 5). Thus, the

responsibility is on the teachers to find ways to meet the differing and complex linguistic needs of adolescent multilingual students for academic writing.

Since the findings of the study demonstrate the need for adolescent multilingual students' development of linguistic resources and repertoires, an important pedagogical question is how teachers in a secondary education context can support adolescent multilingual students in their effort to become proficient academic writers well versed in the specialized language of academic discourse. While there is no agreed upon instructional approach, language and literacy scholars have suggested ways to meet the linguistic needs of multilingual students and, thereby, the challenging demands of the CCSS-ELA. Seeing "transformational opportunities" afforded by the CCSS-ELA, Kibler et al. (2015), propose an instructional approach that is aimed "to socialize students into the language and literacy practices emphasized in the Common Core State Standards," that provides guidance, support, and scaffolding tailored to individual students and texts, and that engages students in materials and activities designed to offer multiple opportunities to learn the content and language (p. 29). Similarly, Walqui and Bunch (2019) argue for the amplification of the curriculum, rather than reduction and simplification, and emphasize the importance of enacting "stimulating, demanding, well-supported lessons to transform what is currently offered to many English learners" (p. 21). In general, recent research on L2 writing instruction suggests that a literacy-rich curriculum that promotes meaningful and socially contextualized engagement in rigorous materials, as well as disciplinary practices, and that provides students multiple opportunities to productively engage with complex, challenging, and academic texts is essential in supporting diverse multilingual students to develop their academic language and literacy skills (Bunch et al., 2019; Kibler et al., 2015; Olson et al., 2017; Walqui & Bunch, 2019).

To be able to produce convincing and coherent academic texts, adolescent multilingual students need to be conversant with the language conventions and the formal code of academic written discourse. Students need to know not just what linguistic choices they can make, but why they make these choices and when these choices are fitting. In order to expand multilingual students' linguistic resources and repertoires, teachers may need a multifaceted pedagogical approach which provides: 1) exposure to rich materials and activities that represent and align with the content and conventions of academic writing; 2) explicit instruction that draws learners' attention to syntactic structures and lexical use; 3) strategy instruction that shows how language is used to construct meaning; and 4) guided practice that promotes the application of their acquired linguistic knowledge in composing, revising, and editing processes (Maamuujav & Olson, 2019). This way, teachers can help students determine how writers create meaning from and with texts and how linguistic choices are shaped by socially constructed genre-conventions.

Conclusion

Previous studies have provided insights into linguistic features of multilingual writers' texts; however, much of the research on the language features of student texts has focused on the postsecondary context, exploring the language and writing development of undergraduate, graduate, English as a second language (ESL), English for academic purposes (EAP), and English as a foreign language (EFL) students at collegiate and pre-collegiate levels. This study, thus, adds to the extant literature by examining the linguistic features of secondary-level adolescent multilingual writers' text-based analytical essays in an effort to determine language attributes that these texts exhibit.

The study informs us of the linguistic needs of adolescent multilingual students who are often marginalized in their classrooms and schools. Kanno and Cromley (2015) assert that many

language-minority students in secondary school are restricted in their access to rigorous curricula and become "long-term" English learners due to instructional and institutional barriers. For many language-minority students across different school districts and states, school is the only place where they are exposed to academic language. Schleppegrell's (2004) remark is noteworthy here; she asserts, "In the absence of an explicit focus on language, students from certain social class backgrounds continue to be privileged and others to be disadvantaged in learning, assessment, and promotion, perpetuating the obvious inequalities that exist today" (p. 3). Thus, meeting the linguistic needs of adolescent multilingual students at instructional and institutional levels is a way to promote equal access and level the playing field for this student population.

Developing proficiency in academic writing is a requisite for college and career success, and students who strive to pursue a path of higher education must learn to use academic language effectively. To become successful academic writers, adolescent multilingual students need to develop their linguistic competence, syntactic fluency, and lexical capacity; they need knowledge of language, text, and discourse to be able to write confidently and competently, to be able to compose analytically and persuasively, to meet the standards set forth by CCSS, and to be ready for college and career success.

References

ACT, Inc. (2007). *Writing framework for the 2011 National Assessment of Educational Progress*. National Assessment Governing Board.

Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal*, 100, 14–27.

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). The Guilford Press.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* John Benjamins.

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

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42–65.

Bunch, G. C., Kibler, A. K., & Pimentel, S. (2012). *Realizing opportunities for English learners in the common core English language arts and disciplinary literacy standards*. Paper presented at the Understanding Language Conference, Stanford, CA. Retrieved from http://ell.stanford.edu/papers.

Bunch, G. C., Schlaman, H., & Rutherford-Quach, S. (2019). Where the rubber meets the road: What teachers and students report when implementing high-quality learning designs for English learners. In A. Walqui & G. C. Bunch (Eds.). *Amplifying the curriculum:*

Designing quality learning opportunities for English learners. (pp. 21-41). Teachers College Press.

California Common Core State Standards. (2013). *English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Retrieved from https://www.cde.ca.gov/be/st/ss/documents/finalelaccssstandards.pdf.

California Department of Education. (2008a). *English language arts teacher guide for California High School Exit Examination*. Sacramento, CA: Author.

California Department of Education. (2008b). *Teacher guide for the 2008 California Writing Standards Test in Grade 7*. Sacramento, CA: Author.

Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research,* 11(3), 415–443.

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41, 965–981.

Crossley, S. A., Kyle, K., Varner, L., Gou, L., & McNamara, D. S. (2014). Linguistic microfeatures to predict L2 writing proficiency: A case study in automated writing evaluation. *Journal of Writing Assessment,* 7(1).

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading,* 35(2), 115–135.

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing,* 26 (4), 66–79.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29, 243–263.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics,* 36, 570–590.

Dowell, N. M., Graesser, A. C., & Cai, Z. (2016). Language and Discourse Analysis with Coh-Metrix: Applications from Educational Material to Learning Environments at Scale. *Journal of Learning Analytics*, 3, 72–95.

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

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly,* 28, 414–420.

Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multidimensional analysis. *Journal of Second Language Writing*, 26, 80–95.

Gordon, R. (2010). *Regression Analysis for the Social Sciences.* Routledge.

Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., . . . Smither Wulsin, C. (2016). *Teaching secondary students to write effectively* (NCEE 2017-4002). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Graham, S., Capizzi, A., Harris, K. R., Heber, M., & Morphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading and Writing,* 27, 1015–1042.

Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-based writing practices and the common core. *The Elementary School Journal,* 115, 498–522.

Grant, L. & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing,* 9, 123–145.

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing,* 18, 218–238.

Halliday, M. A. K. (1994). *Introduction to functional grammar* (2nd ed.). Edward Arnold.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning,* 63, 87–106.

Kanno, Y., & Cromley, J. G. (2015). English language learners' pathways to four-year colleges. *Teachers College Record,* 117, 1–44.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research,* 1, 1–26.

Kibler, A. K., Walqui, A., & Bunch, G. C. (2015). Transformational opportunities: Language and literacy instruction for English language learners in the Common Core ear in the United States. *TESOL Journal,* 6, 9–35

Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing,* 25, 347–373.

Kyle, K., & Crossley, S. A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing,* 34, 12–24.

Kyle, K., & Crossley, S. A. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing,* 34, 513–535.

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal,* 102, 333–349.

Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication version 2.0. *Behavior Research Methods,* 50, 1030–1046.

Laufer, N., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics,* 16, 307–322.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15, 474–496.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45, 36–62.

Maamuujav, U., & Olson, C. B. (2019). Meeting the linguistic needs of adolescent multilingual writers for academic writing. In J. I. Liontas (Ed.). *The TESOL Encyclopedia of English Language Teaching.* John Wiley & Sons, Inc.

MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing*, 32, 1553–1574.

McCarthy, P. M., & Jarvis, S. (2007). *vocd*: A theoretical and empirical evaluation. *Language Testing*, 24, 459–488.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381–392.

McNamara, D. S., Graesser, A. C. McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

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

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Author.

National Writing Project. (2010). *Analytical Writing Continuum*. Available at https://archive.nwp.org/cs/public/download/nwp_file/21466/NWP_Analytic_Writing_Continuum.pdf?x-r=pcfile_d.

Olson, C. B., Matuchniak, T., Chung, H. Q., Stumpf, R., & Farkas, G. (2017). Reducing achievement gaps in academic writing for Latinos and English learners in Grades 7–12. *Journal of Educational Psychology,* 109, 1–21.

Olson, C. B., Scarcella, R., & Matuchniak, T. (2015). *Helping English learners to write: Meeting common core standard, grades 6-12.* Teachers College Press.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics,* 24, 492–518.

Ortega, L., & Carson, J. (2010). Multicompetence, social context, and L2 writing research praxis. In T. Silva & P.K. Matsuda (Eds.). *Practicing theory in second language writing*. (pp. 48–71). Parlor Press.

Ranney, S. (2012). Defining and teaching academic language: Developments in K-12 ESL. *Language and Linguistics Compass,* 9, 560–574.

Scarcella, R. (2003). Academic English: A conceptual framework. *The University of California Linguistic Minority Research Institute Technical Report 2003-1,* Retrieved from https://escholarship.org/uc/item/6pd082d4.

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education,* 12, 431–459.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective.* Lawrence Erlbaum Associates, Inc.

Shaughnessy, M. P. (1977). *Errors and Expectations: A guide for the teacher of basic writing*. Oxford University Press.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

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

U.S. Department of Education. National Center for Education Statistics. (2012). *The Nation's Report Card: Writing 2011* (NCES 2012-470). Institute of Education Sciences, U.S. Department of Education, Washington, DC.

U.S. Department of Education. National Center for Education Statistics. (2020). English language learners in public schools. Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp.

Walqui, A., & Bunch, G. C. (2019). What is quality learning for English learners? In A. Walqui & G. C. Bunch (Eds.). *Amplifying the curriculum: Designing quality learning opportunities for English learners.* (pp. 21–41). Teachers College Press.

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, 47. 100505.

CHAPTER 5

Study 3. An Investigation of Linguistic and Rhetorical Features in the Academic Writing of Culturally Diverse Adolescent Students

Abstract

Understanding what features characterize proficient writing and how linguistic and rhetorical features contribute to human-judged writing quality is crucial in addressing the academic literacy needs of diverse students. To this end, this study examines lexical, syntactic, and rhetorical features of academic texts written by linguistically diverse students (7^{th}-12^{th} grades), using data collected from public schools in six U.S. states. The primary goals of the study were: 1) to identify linguistic and rhetorical features exhibited in students' academic writing; 2) to examine whether the identified linguistic and rhetorical features vary among students with different first/home languages (L1); and 3) to investigate how the identified linguistic and rhetorical features relate to each other and to human-judged writing quality. The results showed that the academic texts of adolescent writers (n=340), in general, exhibit low use of academic vocabulary, less syntactic variety, and reliance on summary structures. Results also indicated no significant difference in rhetorical features among L1 groups, but there was variability among L1 groups in linguistic features. The latent construct of rhetorical features predicts holistic writing score. Lexical and syntactic features, while not significant predictors of holistic writing score, were strongly correlated to one another and to rhetorical features. The study contributes to our understanding of the linguistic challenges students face in their academic writing, the variability among multilingual students, and the relations of linguistic features to rhetorical elements and writing quality.

Keywords: Lexical sophistication, Syntactic complexity, Rhetorical features, Academic writing, Text-based analytical essays, adolescent writers

Introduction

Proficiency in academic writing and mastery of academic language play a critical role in students' admission to and success in colleges and universities. However, developing proficiency in academic writing is a challenging endeavor and complex process that requires in-depth understanding of rhetorical situations, genre conventions, domain knowledge, and linguistic principles (Maamuujav et al., 2019). The challenge adolescent writers face in meeting the demands of academic writing is evident from the results of national writing assessments administered by National Assessment of Educational Progress (NAEP), with only 18% of 12th graders scoring at or above proficient in 2011 writing assessment (U.S. Department of Education, Institute of Educational Sciences, National Center for Education Statistics, 2012). A large disparity exists between the performance of English learners (EL) and non-ELs, with 80% of ELs in 12th grade performing below basic as compared with 19% of non-ELs performing below basic. Only 1% of 12th-grade ELs performed at or above proficient level in the 2011 writing assessment. The writing results of the most recent 2017 administration of NAEP are not available, but the technical summary of preliminary analyses reported a pattern of lower performance. The NAEP reports show that a majority of students in middle and high school in the United States are challenged by the demands of academic writing. The challenge is much greater for adolescent multilingual students who are still in process of developing their English language proficiency.

For academic writing development, students' language skills, knowledge about language conventions, and understanding of rhetorical elements are essential. It is well documented in the literature that language, as a resource for meaning making, plays an important role in the interpretation and construction of texts (Grabe & Kaplan, 1996; Halliday, 1994; Halliday &

Matthiessen, 2014; Schleppegrell, 2004). Academic writing requires the ability to analyze, interpret, synthesize, and explain complex ideas. Performing these higher-order analytical tasks depends on linguistic resources and demands a higher level of language proficiency (Kellogg, 2008). Without mastery of language skills and knowledge about language functions in different contexts, students are constrained as they engage in academic writing tasks that involve an interpretation of texts, analysis and construction of an argument, and the process of complex rhetorical problem solving. In addition, an in-depth understanding of the specialized lexical, syntactic, and discursive features of academic texts is necessary to write well in an academic context. The lack of command of the specialized language of academic written discourse may affect students' ability to analyze texts and express complex ideas. Thus, to better support adolescent students in their effort to become proficient academic writers, it is crucial to understand how they use language in their academic writing and thereby identify what linguistic needs they have in order to be able to write successfully in an academic context.

In addition to language skills, students need to develop a complex rhetorical problem-solving skill to accomplish the higher order tasks demanded by academic writing. For example, students often need to use sources strategically, integrate source material in their writing purposefully, construct and develop an argument, and provide interpretation and analysis. Accomplishing these higher order tasks requires students to make informed choices with an understanding of communicative goals and genre conventions. They need to move beyond "knowledge telling" to compose at the advanced level of "knowledge transformation" (Bereiter & Scardamalia, 1987) as they analyze, argue, interpret, and use evidence to support their argument. These higher-order tasks are particularly challenging for multilingual students who are still in process of developing their proficiency in English. To better understand the challenges

students must address when writing in an academic context, this study systematically analyzes students' text-based analytical essays and identifies common linguistic and rhetorical features of their writing.

Although the academic literacy development of adolescent multilingual students has gained attention among scholars at the intersection of language and education, there is a paucity of studies that systematically examine language and rhetorical features of academic writing of linguistically diverse students in a secondary education context. Most of the extant studies investigating features of students' texts have examined linguistic features without attending to rhetorical elements of the text. In addition, the majority of the studies that examine linguistic features in learner corpora have focused on 1) independent writing that does not require source use, and 2) academic writing of student populations in a postsecondary context. The present study addresses these gaps in the research by examining text-based analytical essays, which is a lesser studied yet an important academic writing genre, written by culturally and linguistically diverse adolescent students in a U.S. public secondary education setting. Examining student texts that represent different populations, genres, and discursive practices is crucial for a more comprehensive understanding of the complexity of writing and written texts.

As academic writing is a gatekeeper for access to college, a requisite skill for college persistence, and a threshold requirement for entry to and promotion of many careers, addressing the academic literacy needs of developing multilingual writers is of paramount importance in a secondary education context (Olson et al., 2020). To address the academic literacy needs of developing writers, it is crucial to identify the linguistic and rhetorical features of their academic writing to better understand how they use language and what rhetorical choices they make in

producing academic texts. Such understanding is critical for writing pedagogy geared toward meeting the academic literacy needs of adolescent writers for college and career readiness.

Overview of the study

The present study examines linguistic and rhetorical features of text-based analytical essays written by linguistically and culturally diverse students attending public middle and high schools in the United States. The principal goal of the study was to identify linguistic needs of these students in meeting the demands of academic writing by systematically analyzing lexical, syntactic, and rhetorical features of the texts they produced as part of the school-based, academic writing assignment. Specifically, the study aims to identify language use and linguistic patterns exhibited in adolescent students' academic writing, examine whether the identified features vary among different groups with different first/home (L1) languages, and finally to investigate the relations between the linguistic features, rhetorical elements, and writing quality by developing a model of writing dimensions and quality. The study is guided by the following research questions:

1. What lexical, syntactic, and rhetorical characteristics do students' text-based analytical essays exhibit?
2. To what extent do the identified lexical, syntactic, and rhetorical features vary among three L1 groups?
3. How do the latent variables of lexical, syntactic, and rhetorical features relate to one another and predict human-judged holistic writing quality score?

Investigating these questions not only paints a descriptive picture of diverse students' academic writing but also helps us understand the complex relations among linguistic features, rhetorical elements, and writing quality.

The linguistic and rhetorical measures were obtained by a combination of methods including: 1) manual coding of sentences based on functional grammar and rhetorical structure theory (RST); 2) quantitative measures of syntactic, lexical, and cohesive features provided by a Natural Language Processing (NLP) tool; and 3) analytic scoring of language use and rhetorical features. In analyzing student texts, a text analytic approach (Sanders & Schilperoord, 2006) was used. It is a useful analytic tool used to systematically analyze the texts writers have produced to gain insight into the writers' representations and to discover the rules and resources they use when composing ((Sanders & Schilperoord, 2006). The research questions were addressed using several statistical methods, including descriptive statistics, multivariate analyses of variance (MANOVA) with accompanying post hoc tests, and structural equation modeling (SEM) techniques that included confirmatory factor analysis (CFA) and structural regression.

The study has important implications for writing pedagogy and research. Studying the language features of diverse adolescent writers' texts that are produced as part of a school-based, academic genre and their relation to rhetorical features and writing quality will: 1) provide insight into understanding their academic language proficiency and the literacy needs to meet the demands of academic writing; and 2) contribute to our understanding of complex relations among linguistic, rhetorical, and writing quality. Bazerman et al. (2017) reminds us that writing is a complex, multidimensional act that "involves multiple processes, each of which draws on different domains of development, which may interact among themselves" (p. 354). Having a better understanding of the characteristics of student' academic writing can inform interventions and curricular approaches that aim to address the academic literacy of multilingual students in a secondary school context. The study's intent to inform writing pedagogy in a secondary education context is particularly important since there is a dearth of research that focuses on both

the linguistic and rhetorical features of diverse adolescent students' text-based analytical writing. In addition, our knowledge about the linguistic features of student writing and their relations to writing quality is based largely on independent writing tasks that do not require students to analyze and integrate sources (Crossley, 2020). To my knowledge, no study has examined linguistic and rhetorical features together in text-based analytical writing. This study, thus, contributes to the extant literature by examining the lexical, syntactic, and rhetorical features together in a lesser studied writing genre.

Literature review

For developing writers, expressing complex ideas in words and sentences to form coherent writing is a challenging task. Text-based analytical writing, in particular, is a complex task as it requires additional rhetorical features of selecting, integrating, and commenting on relevant textual evidence to support an argument (Olson et al., 2020). Thus, examining linguistic and rhetorical features of student texts is crucial not only in determining how these features may affect writing quality (Hinkel, 2002) but also in identifying challenges and needs of students who are striving to develop their academic writing proficiency.

1. Linguistic features and writing quality

Writing quality has traditionally been determined by holistic scoring, a method of direct evaluation in which human raters judge the quality of a written text based on their holistic impression of multiple different components such as focus and depth of response, overall structure and organization, idea development, and language use. Holistic scoring is a parsimonious and flexible method that is accepted as an appropriate approach to evaluating writing quality as it reflects the idea that writing quality is unidimensional as the interaction of multiple elements forms the whole (Steiss et al., 2022). Witte & Faigley (1981) defined writing

quality as “the ‘fit’ of a particular text to its context, which includes such factors as the writer’s purpose, the discourse medium, and the audience’ knowledge of an interest in the subject” (p. 199). The fit of the text to its rhetorical context is often stipulated in the holistic scoring rubric. For example, the National Writing Project’s (NWP) analytical writing assessment rubric and many similar scoring rubrics are used to holistically determine writing quality based on clarity and cogency of argument, quality of interpretation, organization of ideas, integration of evidence, syntactic variety and fluency, diction and word choice, and language conventions. Although multiple components determine the overall quality of a written text in the context of academic discourse, linguistic features are an important component in the construction and interpretation of a text as the macro-level rhetorical features are contingent on micro-level linguistic expressions (Engber, 1995; Maamuujav, 2021).

The extent to which linguistic features predict writing quality has been extensively studied in writing, specifically in second language (L2) writing research. Choice of words and the way the chosen words are arranged in local syntactic structure and broader coherence patterns are main determinants of alphabetic text production. As a result, language and writing researchers have been interested in understanding what linguistic features characterize proficient writing and how linguistic features contribute to human-judged writing quality (McNamara et al., 2010). The linguistic features of student writing have revolved around two larger constructs related to language production — lexical sophistication and syntactic complexity (Crossley, 2020; MacArthur et al., 2019). In several studies of learner corpora, these two linguistic features have been examined independently and together to determine their contributions to writing quality. This is based on the notion that more complex skilled syntax and sophisticated lexical production lead to better writing quality (Crossley et al., 2014; Crossley & McNamara, 2012;

Guo et al., 2013; Kim & Crossley, 2018; MacArthur et al., 2019; McNamara et al., 2010). Most of the studies focus on one or both of these linguistic constructs and their relationships with writing quality. In what follows, I provide an overview of research on lexical sophistication, syntactic complexity, and rhetorical features that include text structure and cohesion, source use and integration, and balanced use of summary, evidence, and commentary.

1.1 Lexical Sophistication

Research on student writing, in general, indicates that lexical features are predictive of writing quality. In studies of learner corpora, lexical usage is operationalized in a variety of ways. While there is no agreed-upon method of operationalizing lexical features in writing, lexical sophistication has been a commonly used construct in relation to writing quality. Although lexical sophistication, in some studies, is loosely referred to as the percentage of 'advanced' words and less frequent words in a text (Laufer & Nation, 1995), there is no current consensus on its definition. While word frequency is "the prototypical measure of lexical sophistication" (Crossley & Kyle, 2018, p. 48), recent studies have proposed and used various different measures and indices for lexical sophistication such as word sequences, semantic relationships, and psycholinguistic properties (Crossley et al., 2013; Crossley & McNamara, 2012; Guo et al., 2013; Kyle & Crossley, 2015; 2016). For example, Kyle and Crossley (2016) expanded lexical sophistication by using word range, bigram and trigram frequency (commonly co-occurring two- or three-word sequences), hypernymy, and psycholinguistic measures (word familiarity, imageability, and age of acquisition) of words in a text. The variations of measures observed in extant studies attest to the complexity of operationalizing lexical sophistication, indicating that the construct of lexical sophistication is not clear-cut. Despite the variation in its measures, lexical sophistication is an important component of language and a strong predictor of

writing quality based on the notion that proficient writers with greater vocabulary skills use sophisticated words and expressions (Crossley & Kyle, 2018).

Several studies that analyze lexical sophistication use a word frequency count based on the idea that proficient writers use less frequent and more advanced words (Crossley et al., 2013; Laufer & Nation, 1995; Morris & Cobb, 2004). For example, Laufer and Nation's (1995) study analyzed university-level multilingual students' texts using the Lexical Frequency Profile. They found that texts written by more proficient writers used less frequent words in their texts. In addition, Crossley and McNamara (2012) found a moderate association between word frequency (r = -.34) and writing proficiency in their study that analyzed essays written by high school seniors. Their findings suggest that students with higher writing proficiency use less frequent, more advanced words. Recent studies have examined lexical sophistication in source-based academic writing in comparison to independent writing tasks using frequency count and other lexical sophistication measures, including word response norms and psycholinguistic measures (Guo et al., 2013; Kim & Crossley, 2018). Guo et al. (2013) found that integrated and independent tasks exhibited different lexical features. For example, word frequency was a significant predictor of writing quality in independent writing tasks, but not in integrated tasks. On the other hand, word familiarity predicted writing quality in integrated writing, but not in independent writing. Similarly, Kim and Crossley's (2018) study found that "word frequency measures were correlated with independent writing scores, but not with source-based writing scores" (p. 51). In general, research suggests that as learners develop higher language proficiency, they use less familiar words (Crossley et al., 2011), infrequent and advanced words (McNamara et al., 2010), and less concrete and more abstract words (Crossley et al., 2014; Crossley & McNamara, 2012; Salsbury et al., 2011).

1.2 Syntactic Complexity

Writing research indicates that syntactic properties of text contribute greatly to writing quality (Crossley, 2020). A variety of syntactic forms that are exhibited in student texts and the complexity of these forms have been common syntactic properties examined in writing studies based on the notion that more complex and varied syntactic structures indicate more advanced writing skills (Lu, 2011). Traditionally, syntactic complexity has been measured by large-grained or global syntactic measures, including sentence and clause length measures and count of a T-unit (main clause and the associated dependent clause/s) and clausal subordination and coordination. For example, based on their review of L2 writing research, Wolfe-Quintero et al., (1998) concluded that clauses per T-unit and subordinate clauses per independent clause were important syntactic complexity measures. Similarly, Ortega's (2003) meta-analysis of syntactic complexity measures of college-level writing confirms that these large-grained, global measures are crucial indicators of writing proficiency and development. Although earlier research established that "mean length of T-unit, mean length of clause, clauses per T-unit, and dependent clauses per clause were the most satisfactory measures" of syntactic complexity (Ortega, 2003, p. 493), recent studies point to the importance of using fine-grained, phrase-level complexity for measures of syntactic complexity. For example, Biber et al. (2011) proposed a new approach that included phrasal complexity measures in L2 student writing based on their findings that phrasal complexity, rather than clausal subordination, was characteristic of academic writing.

Syntactic complexity at a phrase level has been measured by number of dependents per nominal, nominal subjects, and direct objects, and prepositional objects as noun phrases with more dependents indicate more complex phrasal structure (Kyle, 2016; Kyle & Crossley, 2018). For example, in the sentence "*The recurring theme of the article is courage,"* the head noun

theme of the complex nominal subject *the recurring theme of the article* has three direct dependents (*the, recurring, of the article*) while the subject complement *courage* does not have any dependents. Biber et al. (2011) found that complex noun phrase structure is a hallmark of academic writing and an important syntactic characteristic of texts produced by more proficient and developed writers. Several writing studies that measured fine-grained phrasal complexity using the average number of dependents per noun and prepositional phrases, as well as the standard deviation of the number of dependents for noun phrase variety (Kyle 2016; Kyle & Crossley, 2018) found that phrasal complexity predicted writing quality, providing empirical support to the proposition made by Biber et al. (2011).

In general, syntactic complexity is a crucial indicator of writing quality as "higher quality writing generally contains more complex syntactic features" (Crossley, 2020, p. 422). Previous studies have found that both large-grained, global syntactic complexity measures and fine-grained phrasal complexity measures are predictors of writing quality. For example, Lu (2010; 2011) found texts that show higher writing proficiency exhibited longer and more varied sentences. Furthermore, high-quality student texts have greater clausal subordination and complex syntactic structures (Friginal & Weigle, 2014; Grant & Ginther, 2000), greater production of dependent clauses (Crossley & McNamara, 2014), and greater phrasal complexity (Kyle & Crossley, 2018; Taguchi et al., 2013). Kyle & Crossley's (2017) study compared the large-grained measures of syntactic complexity with fine-grained, usage-based complexity measures in relation to writing quality and found that both large-grained measures and fine-grained indices of syntactic complexity were predictors of writing quality. Given the findings of previous studies, syntactic complexity is examined both at clausal and phrasal levels in this study research.

2. Rhetorical Features of Text-based Analytical Writing

Rhetorical features of writing are broad and vary greatly by the writing task, genre, and purpose. Bouwer (2015) pointed out, "Genre differences, shaped by the rhetorical situation in the writing task, are likely to lead to differences in the text that has to be produced" (p. 86). Accordingly, texts that represent the same genre and communicative purpose share similarities in their global structure, rhetorical moves, style and conventions (Swales, 1990). While distinct from other genres, including narrative, expository, and other non-source-based writing, text-based analytical writing, as a school-based genre, is shaped by a distinctive rhetorical situation in which students are asked to present an argument of interpretive analysis, integrate examples and details from the source text for evidence, and provide commentary on their chosen evidence. Although there are multiple different dimensions pertaining to rhetorical and textual features (i.e., global organization, development of ideas), text cohesion, source use and integration, and the extent to which students blend and balance summary, evidence, and commentary are rhetorical features that are particularly important in understanding the quality of text-based analytical writing.

2.1 Text cohesion

Text cohesion, which refers to in-text linking that holds the text together, is an important dimension of writing that signals the relationship between parts of texts (Connor 1996; Halliday & Hasan, 1976). Cohesive ties within a text can indicate an argumentative and semantic relationship between and among sentences and other textual elements, signaling the interconnectivity of a text (Crossley, 2020; Witte & Faigley, 1981). Cohesion has been measured by "the presence or absence of explicit cues in the text that afford connecting segments of texts together" (Crossley, 2020). Halliday and Hasan (1976) identified 5 categories of cohesive

devices that include ellipsis, substitution, reference, lexical cohesion, and conjunction. Witte and Faigley (1981) pointed out that substitution and ellipsis are more frequent in conversation while the latter three are common in written discourse. Thus, referential cohesion, connectives, and lexical and semantic overlaps have been used as measures of text cohesion in students' writing. Recently, based on the natural language processing (NLP) tool TAACO (Tool for Automatic Analysis of Cohesion), Crossley et al., (2016) proposed local (sentence-level), global (paragraph-level), and text (entire text) cohesion as important cohesive features to determine textual interconnectivity.

In writing research, the relation between cohesion and writing quality has not been determined as studies have reported mixed results. An early study conducted by Witte and Faigley (1981) found that high-rated college students' essays had greater density and variety of cohesive ties. A recent study by MacArthur et al. (2019) that examined college basic writers' persuasive essays found a positive relationship between writing quality and referential cohesion that was measured by stem overlap across all and adjacent sentences and semantic overlap across adjacent sentences. However, Perin and Lauterbach's (2018) study on college students' text-based writing reported a negative relationship between referential cohesion and writing quality. Studies that compare different writing genres and task types suggest that the variability in how cohesive features are related to writing quality can be attributed to genre-based differences of texts (Crossley & McNamara, 2012; Crossley et al., 2016; Guo et al., 2013; Kim & Crossley, 2018). For example, Guo et al. (2013) found that the association between cohesion and writing quality differs between source-based and independent writing tasks. Kim and Crossley (2018) also reported that local cohesion at the sentence level was correlated with writing quality in source-based writing, but not in independent writing, while global cohesion at the paragraph

level predicted writing quality in both independent and source-based writing tasks. In sum, there is a growing consensus in L2 writing research that cohesive features of texts vary depending on writing tasks, genres, and purposes.

2.2 Source Integration.

Source integration is an important dimension of text-based analytical writing as the task requires students to analyze the source text and extract relevant details to support and develop their argument. Interpreting, summarizing, and commenting on source material, in fact, are challenging tasks that require critical and close reading of the source text and analytical thinking in selection and integration of source material (Plakans, 2009; Zhao & Hirvela, 2015). As many academic writing tasks "require comprehension and integration of source material" (Plakans & Gebril, 2013, p. 218), examining how students integrate source material in their writing and whether their integration follows the genre conventions and discursive norms is vital in understanding students' academic writing development and quality. In general, research suggests that "holistic writing quality has a strong relationship to more or less effective source use" (Doolan, 2021, p. 129).

Source use and ways of integrating source materials (i.e., quoting, paraphrasing, verbatim copying) have been examined in writing research (Cumming et al., 2006; Doolan, 2021; Gebril & Plakans, 2009; Plakans & Gebril, 2013; Shi, 2004; Watanabe, 2001). Studies have found that writers' source use and integration style are associated with writing quality (Doolan, 2021). While high-quality student texts exhibit more sophisticated use of source material with coherent, purposeful summary and paraphrasing and embedded quotes, texts that show low writing proficiency tend to have verbatim copying of sentences and phrases and low use of paraphrasing and purposeful summary (Cumming et al., 2006; Gebril & Plakans, 2009; 2013; Keck, 2014).

Source use, which is a common practice in academic writing, can be a challenging task for developing writers because integration of source material involves not only analytical decision making but also strategic use of source materials and nuanced way of textual borrowing for a variety of rhetorical purposes, including evidence for argument development.

2.2.1 **Balance of summary, evidence, and commentary.**

An important task of text-based analytical writing is the use of a purposeful summary, integration of relevant evidence, and inclusion of reasoning and commentary. Thus, the extent to which students' texts exhibit a balanced use of summary, evidence, and commentary is a crucial feature worth exploring. This feature of writing is unique to a specific writing task and genre, in this case, the text-based analytical essay. For this reason, it is an area that has not been much explored in writing research since most of the studies have focused on independent writing tasks and typical source-based writing that draws on and integrates multiple sources. While a typical source-based argumentative writing task often requires synthesis and integration of multiple sources for different rhetorical purposes, in writing text-based analytical essays, students often need to analyze a single source or text, provide purposeful summary of or from the text, integrate specific details and examples from the text as evidence to support their ideas, and interpret and comment on the evidence. The paucity of studies that analyze text-based analytical writing reveals that this area is nearly untapped. As most L2 writing studies have focused on linguistic features without attending to rhetorical features, how this rhetorical feature relates to writing quality and linguistic features hasn't been much explored.

To my knowledge, only a handful of studies that analyzed text-based analytical writing of students at the secondary level examined the students' use of summary, evidence, and commentary (Chen et al., 2020; and Nazzal et al., 2020). These studies focused only on

rhetorical features of text-based analytical writing. For example, Chen et al. (2020) analyzed non-passing, adequate-passing and strong-passing text-based analytical essays of middle and high school students. They found that high-scored texts exhibited purposeful summary, relevant quotes as evidence, and explicit commentary that explains how the evidence supports the writers' ideas. In addition, Nazzal et al.'s (2020) study that examined text-based analytical essays written by community college students reported that the relevant evidence "substantiated with commentary that discusses the significance of the evidence" was one of the distinctive features of high-quality writing (p. 282). Overreliance on summarizing and retelling, rather than providing commentary and interpretation of the evidence, is a shortcoming observed in the text-based analytical writing of developing L2 writers (Olson et al., 2017, 2020). This is supported by an earlier study by Bereiter and Scardamalia (1987) that introduced the notions of knowledge telling and knowledge transformation. As Bereiter and Scardamalia (1987) pointed out, inexperienced writers often rely on knowledge telling or a simplified idea generation and summarization process while experienced writers go beyond mere retelling and engage in knowledge transformation which involves complex problem solving and analytical skills.

Taken together, linguistic and rhetorical features of students' texts that have been examined in previous studies illuminate our understanding of the relation between these features and writing quality. However, there are still gaps and inconsistencies in extant literature. One important gap in research is tied to the scarcity of studies that explore lexical, syntactic, and rhetorical features in tandem and how these features independently and together contribute to writing quality in text-based analytical writing. To my knowledge, no study has examined linguistic and rhetorical features of student texts in tandem and the complex relationship between linguistic elements, rhetorical components, and writing quality. In addition, text-based analytical

writing is a genre that has been underexplored, yet it is an important academic writing genre that is aligned with the Advanced Placement (AP) English Literature and Composition Assessment that students take for college entrance. Further, adolescent multilingual students' academic writing remains an underexplored area (Fitzgerald, 2017). The present study addresses the gaps in writing research by examining the linguistic and rhetorical features together in a lesser studied writing genre using a large, representative sample of linguistically diverse adolescent students attending public secondary schools in multiple states in the United States.

Methodology

Data Source and Sample

For this study, 384 on-demand, text-based, analytical essays were collected from 23 English Language Arts (ELA) and English Language Development (ELD) teachers' classes (grades 7-12) in 15 public schools serving a large number of language-minority students in five states—Arizona, Minnesota, Texas, Utah, and Wisconsin. These teachers are participants in the Pathway to Academic Success Project's professional development program funded by a U.S. Department of Education's Education Innovation and Research (EIR) grant. The participating teachers administered reading and text-based analytical writing tasks over two class periods. During the first period, students read a literary nonfiction text (a short newspaper article) and completed a post-reading worksheet focusing on summarizing, visualizing, analyzing a theme and author's craft, and reflecting on the purpose and significance of the article. During a second period, the students were asked to write an interpretive response of the text, which is an argument writing of literary analysis (see Appendix B and C). The task requires students to identify an important theme in the article, make a claim about the theme they selected, and support their claim with evidence from the text. The writing task is designed to prompt students

to formulate and advance an argument of literary analysis and use details and examples from the source text to support their argument. The administration of these tasks was completed prior to the Pathway to Academic Success professional development intervention.

The participating teachers completed rosters for their class sections, providing students' demographic information, including gender, language status, first and home language, and English language proficiency score (if available). Each participating teacher completed a comprehensive survey that consists of four sections: teacher's background, the types of writing students engage in during the school year, instructional practice for content development (ideas and organization), and instructional practice for sentence fluency and language conventions. The survey was given to gain an insight into how much exposure students have to various writing genres and language instruction and how frequently they engage in extended writing practice.

The data were collected in two phases prior to the Pathway Project Professional Development intervention; the first phase of data collection took place in Spring and Fall of 2019, and the second phase took place in Fall 2020. During phase 1, I collected 338 essays from 20 classrooms with 413 students (72 students didn't complete the writing task either because they were absent, chose to opt out, or did not produce any text during the allotted writing time). Out of these 338 essays, 29% of the essays were written by students whose L1 is English, 50% were written by students whose L1 is Spanish, and only 21% percent of the essays represented students with other L1s. During phase 2, additional essays and student demographic information were collected to diversify the sample and add more essays written by students whose first language is not English or Spanish. During this phase, I collected 46 additional essays. The total number of essays collected was 384 text-based analytical essays. Out of these essays, 44 were excluded because they were too short (fewer than 50 words). Thus, the number of essays that

constitute the selected sample is 340 essays written by culturally and linguistically diverse adolescent students with L1s representing 27 different languages. Of the essays, 52% were written by female students. The largest language groups are students with Spanish L1 (45%) and students with English L1 (27%). The remaining 29% represent multilingual students with 24 different L1s, including Somali, Hmong, Arabic, Burmese, Amharic, Karen and so on (see Appendix G for Summary of Data).

Corpus Description

The corpus comprises the selected sample of 340 text-based analytical essays written by middle school (30%) and high school students (70%) for the first two research questions. For the third research question, I added 70 essays written by Spanish-speaking multilinguals from the previous two studies of the dissertation. Thus, the full sample used for SEM analyses is 410 student texts. This was to increase the sample size for modeling as SEM techniques require a large sample and some estimates in SEM, including standard errors for effects of latent variables, may not be accurate when sample size is small (Kline, 2016). Since SEM techniques were used for modeling writing dimensions and examining relationships among the latent constructs of linguistic and rhetorical features, it was also important to use a more representative sample. The terms selected sample (n=340) and full sample (N=410) are used in the following sections.

The essays in the corpus vary in length. The average word count of the essays in the selected sample is 186.86 (SD 107.56), with grade 12 having the highest score (M=212; SD=112.28) and grade 11 with the lowest score (M=150.12; SD=70.57). The average number of paragraphs is 2.81 (SD=1.74), and the average number of sentences is 10.94 (SD=6.76). The mean length of sentences (MLS) measured by the number of words per sentence is 19.23 (SD=10.93), with students writing longer sentences in 7^{th} grade (M=21.56; SD=9.46) and 8^{th} grade (21.13;

SD=18.25), compared to those in higher grades (M=15.83; SD=5.43 in 11th grade and M=18.60; SD=6.04 in 12th grade). The mean length of clauses (MLC) is 8.10 (SD=1.59) and is relatively stable across grades. The average writing quality score is 4.68 (SD=1.57). As the maximum score one can receive is 12, the average score is relatively low across grades. Length and quality of the essays are variable across grades and do not necessarily follow an upward trend as the grade levels advance. These descriptive details are summarized in Table 5.1 below.

Table 5.1

Corpus Description by Grade Level

	Full sample		7th Grade		8th Grade		9th Grade		10th Grade		11th Grade		12th Grade	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Paragraph count	2.81	1.74	2.22	1.50	2.58	1.64	2.58	1.87	3.00	1.97	3.12	1.89	3.13	1.48
Sentence count	10.94	6.76	10.19	7.92	9.72	5.79	11.66	8.23	10.89	6.93	9.88	5.49	11.98	5.93
Word count	186.9	107.6	189.9	133.8	162.2	68.9	190.5	108.8	186.2	122.8	150.1	70.6	212.2	112.3
MLS	19.23	10.93	21.56	9.46	21.13	18.25	18.22	7.54	18.80	9.09	15.83	5.43	18.60	6.04
MLC	8.10	1.59	8.21	1.56	7.48	1.27	7.85	1.87	8.27	1.57	8.27	1.45	8.59	1.49
Quality score	4.68	1.57	4.37	1.52	4.12	1.49	4.32	1.30	4.71	1.42	4.69	0.95	5.51	1.72
Observations	340		27		74		62		76		16		85	

Procedures

The initial phase of the study involved preparing the corpus for analyses and typing the handwritten essays. The steps after typing and reviewing all essays for typographical errors included holistic scoring, analytic scoring of all essays, manual coding of all sentences in each essay, and running the essays through Natural Language Processing (NLP) tools. The final step included conducting statistical analyses with the following steps: 1) descriptive statistics to test normality and to identify common linguistic and rhetorical patterns; 2) correlation analyses and multicollinearity testing using Stata 15.1; and 3) conducting MANOVA and multiple ANOVAs with post hoc Bonferroni using Stata 15.1; and 4) using SEM techniques —CFA to test three

alternative models and structural regression to examine the relations between the latent constructs of lexical, syntactic, and rhetorical features and writing quality. For SEM analysis, I used MPlus Version 8.6. These steps are further elaborated below.

Holistic Scoring of Writing Quality

The essays were scored holistically using a 6-point scale analytical writing assessment rubric developed by the UCI Writing Project to assess the writing quality. The rubric for scoring the essays is based on those used to evaluate the essay portion of the CAHSEE (California Department of Education, 2008a), the California STAR 7 Direct Writing Assessment (California Department of Education, 2008b), and the NAEP (ACT, Inc., 2007). As indicated in the analytical writing assessment rubric, the writing quality was assessed based on the quality of interpretation, the clarity of thesis, the organization of ideas, the integration of evidence, and language conventions. (See Appendix D for rubric). Essays were independently scored by two expert raters who had extensive experiences teaching, coaching ELA teachers, and holistically scoring text-based analytical essays. The scores were then added together for a final score between 2 and 12. If there was a discrepancy of two or more points between the raters, the essay was scored by a third rater whose score and the closest score given by either first or the second raters were added together for a final score. If the third reader's score was in the middle of the first and second raters' scores, the average score of the first and second raters was added to the third rater's score for a final score. 97% of the essays had either exact or within 1 point score agreement between two raters, and the remaining 3% needed a third reader.

Analytic Scoring for Linguistic and Rhetorical features.

Analytic scoring focused on specific linguistic and rhetorical features of the essays. I adapted the analytic scoring framework (Steiss et al., 2022) that was developed by the UCI's

Pathway Project research team based on the National Writing Project's (NWP) Analytic Writing Continuum Rubric and in consultation with research, literature, and common rubrics used to assess writing quality. The adapted analytic scoring framework used for this study includes five items—lexical sophistication, syntactic complexity, text cohesion, source use, and balance of summary, evidence, and commentary (see Appendix E for the analytic scoring framework). These five items are continuous measures scored on a scale of 1-7, with 7 indicating "effective" and 1 indicating "ineffective" or "not evident." To ensure reliability, consensus coding and double coding were undertaken. Two coders with extensive experience of teaching university-level writing courses coded 10% of the essays (n=35) together (consensus coding). Once the consensus coding was complete, 26% of the essays (n=90) were double coded by the same two coders. For all items, an agreement between a score point (on a 7-point scale) was considered acceptable (Bang, 2013). The agreement within 1 point for lexical sophistication was 95%, syntactic complexity 93%, cohesion 86%, source use 97%, and balance of summary, evidence, and commentary 91%.

Manual Coding of Sentences

All sentences in each essay were coded manually by two human raters, both of whom have expertise in applied linguistics and experience in teaching university-level academic writing courses (see Appendix F for manual coding procedures). The sentences were coded based on three syntactic features: 1) sentence structure — simple, compound, coordinated, and complex-subordinated; 2) sentence types — types of clauses (relative, adverbial, complement, noun modifier) and phrases (participial, absolute, appositive); and 3) sentence boundary issues — fragment, run-on, and faulty sentence. Sentence coding also included a rhetorical feature related to source integration (summary/paraphrased sentence, floating quotation, verbatim copied, patch

written sentences, embedded quote, interpretive sentence). The interrater agreement between the coders was 94% for identifying sentence boundary issue, 92% for sentence structure and types, and 88% for source integration.

Natural Language Processing Tools for Quantitative Measures

The NPL tools TAALES 2.5.5 (Kyle & Crossley, 2015), TAASSC 1.3.8 (Kyle 2016), TAACO 2.0.4 (Crossley et al., 2019), and VocabProfilers Web VP Classic Version 4 (Cobb, 2018) were used to obtain quantitative measures of lexical, syntactic, and cohesive features of student texts. Essays with a length of 50 words and more were used for NLP measures as 50 words is the recommended minimum length for lexical features (Crossley, 2018; Zenker & Kyle, 2021). Spelling errors were corrected to obtain accurate quantitative measures. Sentences and large phrases (3 and more consecutive words) that were copied or quoted from the source text were deleted because they reflect the syntactic and lexical features of the source material rather than students' language use. The data cleaning process underwent a robust review process to ensure that the texts were clean and satisfy NLP criteria. This was important since "dirty" texts with spelling errors and odd symbols undermine the validity of NLP measures. Indices related to syntactic, lexical, and cohesive features of the texts were selected based on theoretical considerations and the findings of the previous studies.

Analytic Strategies

Addressing the research questions involved several different analytic strategies, including a text analytic approach, descriptive statistics, multivariate analysis of variance (MANOVA) and Structural Equation Modeling (SEM) techniques. The first research question was exploratory and aimed to identify common linguistic and rhetorical patterns and features in student texts using the text analytic approach and descriptive statistics. The manual coding of the sentences for their

rhetorical and syntactic functions were informed by the text-analytic approach, which draws on rhetorical structure theory (RST; Mann & Thompson, 1988) and the systemic functional linguistic (SFL) approach (Halliday 1994). While RST centers on the macro level features of the text including text structure, organization, and rhetorical relations such as restatement and summary, interpretation and evaluation, and integration of evidence, SFL deals with micro level linguistic features based on the idea that certain patterns of language use within a specific discursive context can be detected through a systematic approach and functional analysis of text construction. Text analysis is a useful analytic tool used to systematically analyze the texts writers have produced to gain insight into the writers' representations and to discover the rules and resources they use when composing ((Sanders & Schilperoord, 2006).

To answer the second research question, MANOVA and the accompanying post hoc tests were employed. For MANOVA analyses, the texts were divided into three different groups based on students' first and home languages (L1s) and the linguistic distance of their L1s to English. Table 5.2 shows the language grouping based on students' L1s and the language families they belong to. For example, texts in group 1 were written by students who reported English as their L1. Texts in group 2 were written by students with Spanish and French L1s, both of which belong to the Romance language family. Romance languages have a higher percentage of cognate words with English, compared to the languages in group 3 and, therefore, their linguistic distance to English is lower. The group 3 represents a rich variety of languages that belong to various language families such as Afroasiatic, Sino-Tibetan, Austronesian, and Niger-Congo. All the languages other than English, Spanish, and French were put into one group for two reasons. First, the languages in group 3 have higher linguistic distance to English, compared to Romance languages. Second, comparable group sizes were needed for robustness of the equal variance

assumption of MANOVA. The post hoc tests included multiple univariate analyses of variance (ANOVAs) with Bonferroni correction and post hoc analyses.

Table 5.2

Language Grouping Based on First and Home Languages

	Language Grouping	Language Family	First/Home Language	#
Group 1	L1—English	Germanic	English	92
	(Total 94, 28%)	Germanic/Romance	English/Spanish	2
Group 2	L1—Romance Languages	Romance	Spanish	152
	(Total 154, 45%)	Romance	French	2
Group 3	L1— All Other Languages	Afroasiatic, Cushitic	Somali	16
	(Total 92, 27%)	Hmong-Mien	Hmong	12
		Afroasiatic, Semitic	Arabic	11
		Sino-Tibetan	Burmese	9
		Afroasiatic, Semitic	Amharic	8
		Sino-Tibetan	Karen	5
		Austroasiatic	Vietnamese	5
		Slavic	Russian	3
		Niger-Congo	Swahili	3
		Afroasiatic, Semitic	Assyrian	2
		Austronesian	Malay	2
		Eastern Indo-Aryan	Rohingya	2
		Niger-Congo	Yoruba	2
		Afroasiatic, Cushitic	Oromo	2
		Afroasiatic	Oromo/Amharic	1
		Sino-Tibetan	Cantonese/Taiwanese	1
		Sino-Tibetan	Falam	1
		Niger-Congo	Ibibio	1
		Kra-Dai, Tai	Laotian	1
		Austronesian	Micronesian	1
		Afroasiatic, Chadic	Mina	1
		Austronesian	Tagalog	1
		Afroasiatic, Semitic	Tigrinya	1
		Niger-Congo	Twi	1

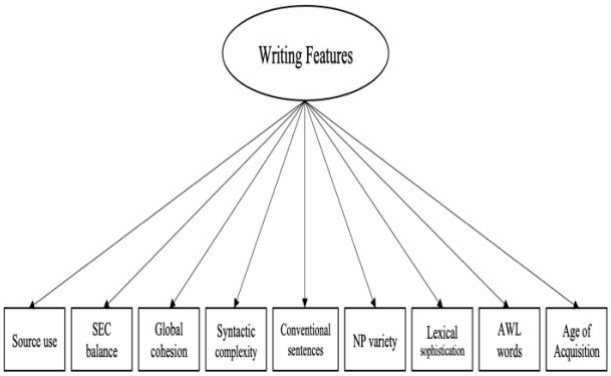
The third research question was addressed using structural equation modeling (SEM) techniques which included confirmatory factor analysis (CFA) and structural regression. First, the dimensions of writing using lexical, syntactic, and rhetorical indices were examined by fitting 3 alternative models using CFA. The first model was a baseline, single-factor model that hypothesized that all the observed lexical, syntactic, and rhetorical variables indicate a unidimensional writing feature. The second model was an alternative, two-factor model in which

linguistic (lexical and syntactic combined) and rhetorical features were distinct but related factors of writing that covary. The third model is a three-factor model in which lexical sophistication, syntactic complexity, and rhetorical features were distinct but related factors that covary. Four measures of model fit, including Chi-square (χ^2) difference, root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residual (SRMR), were used to evaluate the model fits. The best fitting model was chosen for structural regression to examine the relations of these variables and writing quality.

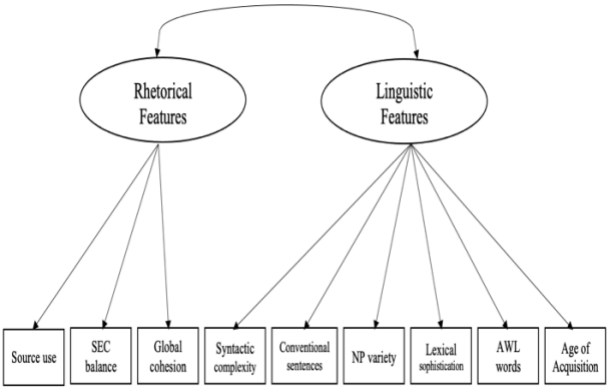
Figure 5.1

Modeling the Features of Writing: Three Models Tested

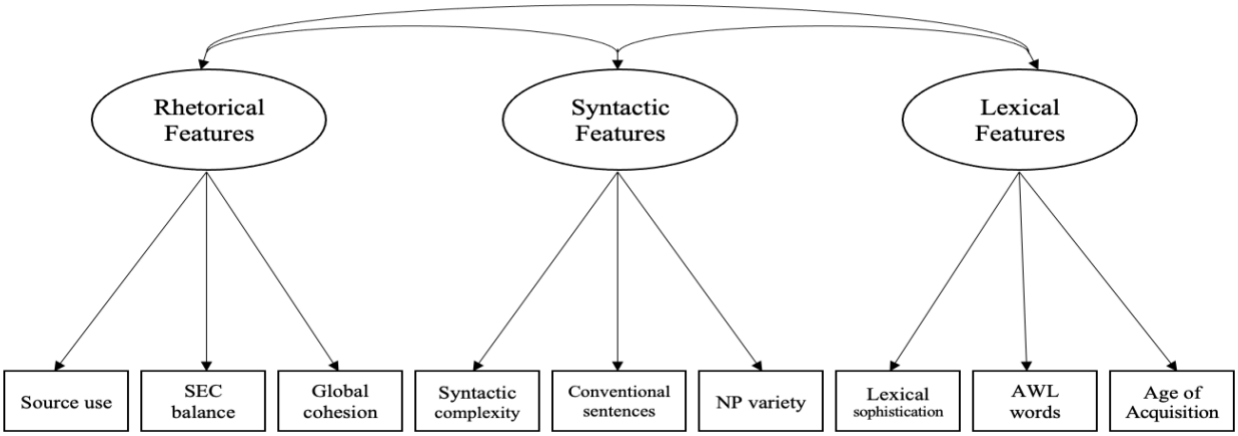
a. One-factor, baseline model



b. Model 2: Two-factor model



c. Model 3: Three-factor model



The measures of lexical sophistication, syntactic complexity, and rhetorical features that were used in analyzing the text-based analytical essays are summarized by research questions in Table 5.3 below.

Table 5.3

Data Matrix: Measures Used in the Study by Research Questions

RQs	Measures
What lexical, syntactic, and rhetorical characteristics do students' text-based analytical essays exhibit?	Lexical sophistication: • Lexical sophistication (LS) analytic score • NLP measures: lexical diversity (LD), lexical frequency, AWL words, age of acquisition (AoA), content word (CW) familiarity. Syntactic complexity: • Syntactic complexity (SC) analytic score • Manual coding measures: Conventional/unconventional sentences, Simple, coordinated, and subordinated sentences, clause types • NLP measures: Noun phrase complexity and variety indices: dependents per nominal and its SD, dependents per nominal subject and its SD, dependents per direct object and its SD, dependent per object of prepositions and its SD. Rhetorical features: • Analytic scores of text cohesion, source use and integration, and balance of summary, evidence, commentary (SEC) • Manual coding measures: text structures, types of source integration • NLP Measures: global cohesion measured by lexical overlap across paragraphs, all connectives, word2vec similarity score.
To what extent do the identified lexical, syntactic, and rhetorical features vary among three L1 groups.	Lexical sophistication: • LS analytic score • NLP indices: LD, K1 words, AWL words, CW familiarity, AoA. Syntactic complexity: • SC analytic score • Manual coding measures: Unconventional sentence, complex/compound sentences • NLP measures: noun phrase complexity factor score, noun phrase variety factor score Rhetorical features: • Analytic scores: source use, SEC balance, cohesion • NLP measures: global cohesion measured by lexical overlap, all connectives
How do the latent variables of lexical, syntactic, and rhetorical features relate to one another and predict human-judged holistic writing quality score?	• Lexical sophistication indicated by: 1) LS analytic score, 2) AWL words, 3) AoA • Syntactic complexity indicated by: 1) syntactic complexity analytic score, 2) conventional sentences, 3) NLP index of noun phrase variety • Rhetorical features indicated by: 1) source use analytic score, 2) SEC balance analytic score, 3) NLP index of lexical overlap across paragraphs (global cohesion) • Holistic Writing Quality Score

Results

The goal of the study was: 1) to identify common linguistic and rhetorical features in text-based analytical writing of adolescent students; 2) investigate whether the identified linguistic and rhetorical features vary among three L1 groups; and 3) examine the relations among the identified features and the extent to which they predict writing quality.

Linguistic Features in Students' Text-based Analytical Writing (RQ1)

For linguistic features, I examined lexical diversity and sophistication and syntactic complexity at both clausal and phrasal levels. Syntax and vocabulary are two fundamental linguistic resources synergistically used to construct meaning and communicate ideas. As writing text-based analytical essays requires students to draw on their linguistic resources, systematically analyzing students' syntactic construction and productive vocabulary use can help us understand the linguistic resources students rely on, as well as linguistic challenges they may encounter when composing.

Lexical Features

Lexical characteristics of the corpus were examined using quantitative measures of lexical diversity (LD), frequency measures based on General Service List (GSL), AWL words, and two psycholinguistic measures of words — age of acquisition and content word familiarity. On the LD measure, the average score of type-token ratio (TTR) for content words in the full sample was 0.67 (SD = 0.10), lower than the estimated mean TTR score of grades 6 through 8 (M=0.813; SD=0.069) based on grade-level norms of language arts writing (see McNamara et al., 2014 for grade-level norms). This low LD score indicates less lexical variety and more repetition of content words. For example, the following introductory paragraph from an essay

exhibits the use of several words and phrases multiple times, leading to a repetition of content words and indicating less lexical diversity.

> The **theme** of the story "Sometimes **the earth is cruel**" is actually that the **Earth is cruel**. My **reason** for this **theme** is because of the **natural disaster**. Like how **natural disasters** kill lots of people. In the next paragraph, I am going to show you evidence that **natural disasters** are one of the **reasons earth is cruel**.

While some repetitions are necessary, especially when strategically used to create more cohesion, clarity, and cadence, excessive and unnecessary repetition often signals the writer's lack of lexico-grammatical awareness. Such needless repetition is a common lexical feature observed in the sample of the text-based analytic essays.

A lexical frequency measure was used to examine the frequency level of the words based on the notion that the use of less frequent words indicates more lexical sophistication. The essays exhibited a higher percentage of more frequent words and less sophisticated words. On average, the words from GSL 1-1000 list (k1 words), which indicate more frequent and basic words that are common in everyday communication, covered 87.29% (SD = 4.78) while the average percentage of the words from GSL 1001-2000 list (k2 words) is 4.50 (SD = 2.20). The overall GSL coverage of the productive vocabulary in the study corpus was 91.79% whereas words from the Academic Word List (AWL), on average, covered 2.37% (SD = 1.94) of the text. The average score for AWL coverage is higher for essays in the upper grades (11^{th} and 12 grades). Words covered by AWL include more advanced words common in academic texts. In the corpus of academic texts compiled by Coxhead (2000), words from GSL account for 76.1% of the words while AWL coverage accounts for 10% of the productive vocabulary. In general, the AWL coverage in the sample was far below the estimated percentage (approximately 10%) of

AWL words used in academic writing. The mean scores and the standard deviations of the lexical features by grade levels are provided in Table 5.4 below.

Table 5.4

Lexical Features by Grade Level: Mean Scores and Standard Deviations

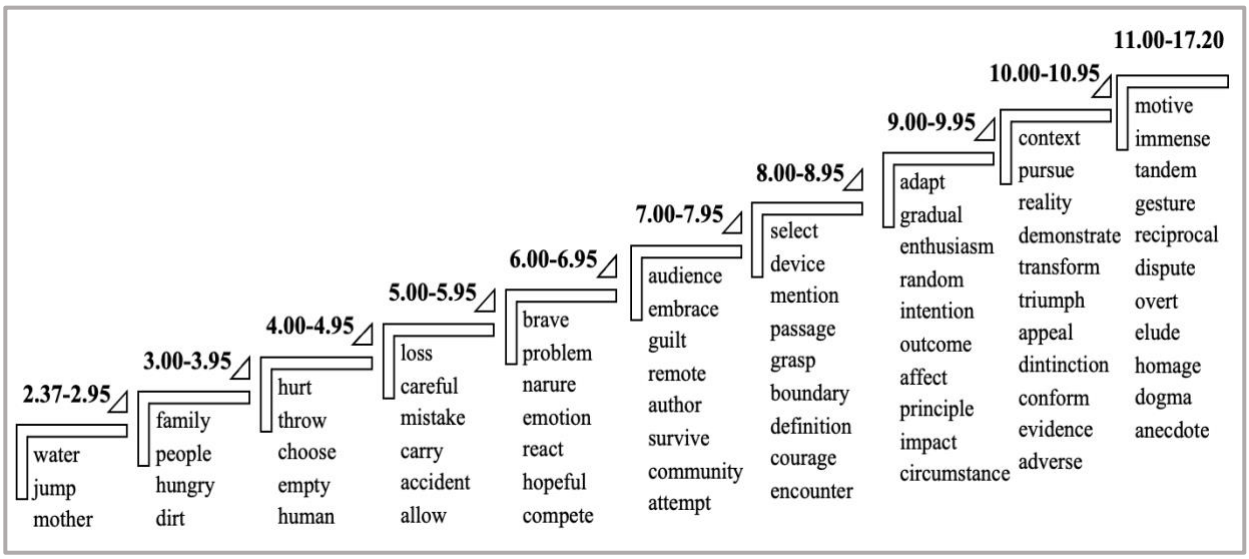
	Full sample	7th Grade	8th Grade	9th Grade	10th Grade	11th Grade	12th Grade
	Mean *(SD)*	Mean *(SD)*	Mean *(SD)*	Mean *(SD)*	Mean *(SD)*	Mean *(SD)*	Mean *(SD)*
LD: TTR for content words	0.67 (0.10)	0.63 (0.1-)	0.68 (0.08)	0.64 (0.11)	0.67 (0.12)	0.67 (0.11)	0.68 (0.10)
K1 Words: GSL 1-1000	87.29 (4.78)	87.40 (4.53)	87.30 (4.70)	88.56 (5.54)	87.51 (4.85)	87.80 (5.42)	86.01 (3.90)
K2 Words: GSL 1001-2000	4.50 (2.20)	5.89 (2.99)	4.38 (2.31)	4.68 (2.04)	4.41 (2.20)	4.26 (2.00)	4.16 (1.80)
AWL Coverage	2.37 (1.94)	1.94 (1.24)	2.19 (1.86)	1.58 (1.43)	2.16 (1.79)	2.44 (2.36)	3.43 (2.15)
Age of Acquisition	5.02 (0.30)	4.92 (0.30)	4.99 (0.26)	4.87 (0.26)	5.05 (0.30)	5.02 (0.39)	5.17 (0.26)
Content Word Familiarity	582.94 (7.26)	582.89 (7.51)	582.42 (6.79)	585.41 (7.25)	583.19 (7.55)	585.65 (12.05)	580.89 (5.43)
Observations	340	27	74	62	76	16	85

The psycholinguistic indices of the words — age of acquisition (AoA) and content word familiarity — showed a similar trend, indicating students' reliance on more frequent and familiar words that are in the lower tiers of vocabulary level. The mean score of AoA in the sample was 5.02 (SD = 0.30), indicating a lower level of vocabulary in AoA ratings based on the study conducted by Kuperman et al. (2012) that used 30,000 English words. Similar to the AWL coverage, the grade-level mean scores of AoA showed an upward trend, with texts from upper grades (grades 10-12) scoring higher than those from lower grades. As indicated in Figure 2, which was compiled from Kuperman et al.'s (2012) study, higher AoA scores (9.00 and above) represent words that are commonly used in academic written discourse while lower AoA scores

(5 and below) denote more frequent and familiar words that are common in everyday speech. In the sample, even the highest mean score observed in grade 12 (M=5.17; SD=0.26) fell within the lower range of vocabulary level. This was consistent with the mean scores of content word familiarity (M=582.94; SD = 7.26) based on MRC ratings (100-700). A low score on content word familiarity indicates less familiar and frequent words which are, therefore, more sophisticated. On the other hand, words with high familiarity scores are more familiar words that are used frequently in everyday communication. For example, words such as *fallacy* (379), *equity* (319), *impetus* (353), *mastery* (371), *resolve* (387) are within 300 range while words such as *earth (*580), *education* (585), *truth* (581), *keep* (584), *mistake* (584) are within 500 range on MRC rating of word familiarity (Coltheart, 1981).

Figure 5.2

Sample Words Representing AoA Scores Based on Kuperman Ratings



In general, the analysis of lexical features based on quantitative measures obtained from NLP tools indicated that the texts in the sample had average low scores of lexical diversity, AWL word coverage, and age of acquisition score and average high scores of word frequency and content word familiarity. This suggests that adolescent students' academic writing shows a

higher percentage of more frequent, familiar words and low use of academic and advanced words. These characteristics are in accordance with the results of the human-judged analytic scoring of lexical sophistication, which showed that the majority of the essays (66.18%) received a score of 3 (developing) and below while only 17.65% of the essays received a score of 5 (competent) and above.

Syntactic Features

Syntactic features of student texts were examined using measures derived from manual coding and the computational tool TAASSC (Kyle, 2016). The results of the systematic analysis based on manual coding of sentences revealed that 34% of all sentences in the sample were unconventional sentences with boundary issues and other structural problems, making up the most frequent pattern. Unconventional sentences include run-ons and comma splices (16.83%), faulty sentences that have semantic and structural problems (10.95%), and fragments or incomplete sentences (6.10%). One-clause simple sentences constituted another common pattern, accounting for 20.66% of the total number of sentences in the sample. Subordinated sentences with finite dependent clauses, which accounted for 22.03% of the sentences, include finite complement clauses (6.24%), finite adverbial clauses (5.91%), and finite noun modifier clauses (5.43), and sentences with three or more finite dependent clauses (4.45%). Subordinated sentences with nonfinite dependent clauses, including nonfinite complements (2.58%), nonfinite adverbials (2.04%), nonfinite noun modifiers (0.73%), and sentences with multiple nonfinite combinations (0.31%) made up only 5.66% of all sentences in the sample. Subordinated sentences with a combination of finite and nonfinite dependent clauses accounted for 6.52% of the sentences. Sentences with both subordinated and coordinated clauses (three or more clauses) made up about 7.59% of all sentences in the sample. These results are summarized in Table 5.5.

Table 5.5

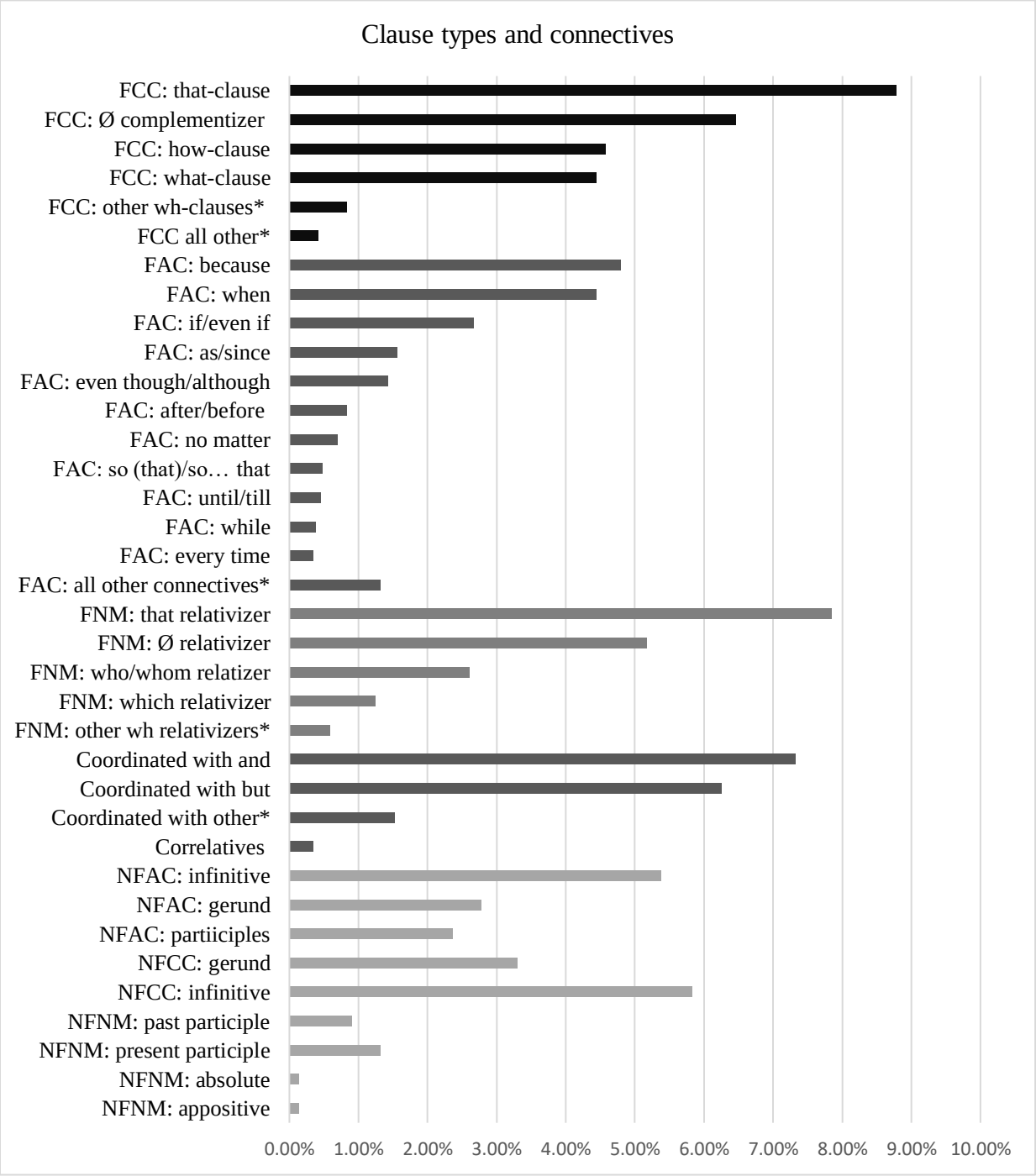
Syntactic Patterns Based on Manual Sentence Coding

Categories	Subcategories	%	Examples
Unconventional sentences (33.87%)	Run-on	16.83%	In "Sometimes the Earth is cruel" the author explains the situation of the country of Haiti one of the themes in this article is human competition.
	Faulty	10.95%	Bad things happen to people but people always overcame for explain people go back to their regular life.
	Fragment	6.10%	Which then ties back to the theme of an eye for an eye.
Simple sentences (one clause)		20.66%	His descriptions of disasters are so powerful.
Subordinated sentences with finite dependent clause/s (22.03%)	Finite complement	6.24%	Additionally, the author says that there has been a history of political instability.
	Finite adverbial	5.91%	The Earth is cruel because we don't take care of it.
	Finite noun modifier	5.43%	However, human cruelty is something that can be avoided, justified, punished, or ignored completely.
	Multiple finite dependent clauses	4.45%	In conclusion, the power of nature is kindness and caring because the man in the water helped people that were drowning even though he was also fighting for survival.
Subordinated sentences with nonfinite dependent clause/s (5.66%)	Nonfinite complement	2.58%	Using repetition gives his article a good flow (gerund). Pitt's purpose in this article is to express human stubbornness and determination (infinitive)
	Nonfinite adverbial	2.04%	The author uses figurative language to describe the destructive ways of the water and nature (infinitive). The article informs us about the plane crash while delivering the message about an ordinary man and his selfless actions (gerund).
	Nonfinite noun modifier	0.73%	These are disasters authored by human hands and by human greed and corruption (past participial). This man was just an ordinary middle-aged man planning to relax on his flight like everybody else (present participial).
	Multiple nonfinite clauses	0.31%	The purpose of the article was to inform about the event while delivering the message about an ordinary man and his selfless actions (nonfinite complement + nonfinite adverbial).
Subordinated sentences with finite and nonfinite clauses (three or more clauses)		6.52%	Rosenblatt carefully constructed the image of man standing up against nature in order to show that, [even though a civilian lost the physical battle,] the emotional battle had been won in the minds of those who saw him.
Coordinated clauses (3.67%)	*and* conjunction	1.85%	[He chose to "challenge death"] **and** [no one really knows why].
	but conjunction	1.46%	[He didn't have super strength or laser eyes], **but** [he had a kind heart and a brave soul].
	Other conjunctions	0.36%	[It was not his duty] **nor** [was he obligated to do so]. [Like everyone else he wanted to live] **so** [his final act was so stunning]. [**Not only** is this story a real life attention grabber], **but** [it **also** tells a story about the good in people].
Sentence with subordinated and coordinated clauses (three or more clauses)		7.59%	[He knew he would die and suffer] **but** [that didn't matter to him because he knew (that he had to sacrifice his life for others)] **and** [that's a brave thing to do].

The syntactic complexity was further examined by analyzing the types of clauses and the connectives that were used to combine clauses. Figure 5.3 exhibits the percentages of clauses and connectives used in the learner corpus. As indicated in Figure 5.3, finite complement clauses, finite adverbial clauses, and finite noun modifiers were common patterns among the subordinated and coordinated clauses, while nonfinite dependent clauses were less frequent structures in the sample. Finite complement clauses with the complementizer *that* and Ø complementizer were the most common structure within the finite complement clause category. In addition, the two most frequent finite adverbial clauses were those with the causal connective *because* and temporal connective *when,* together constituting 48% of all finite adverbial clauses. Within finite noun modifier clauses, those with the relative pronoun *that* and Ø relative pronoun were common structures. Among sentence coordination, clauses with *"and"* and "*but"* conjunctions were the most common, constituting 88% of all coordinated sentences. Nonfinite noun modifiers, including present and past participials, absolutes, and appositives, were the least frequent structures while nonfinite infinitive structure was the most common among the nonfinite dependent clauses. These results based on manual coding suggest that sentence boundary issues, lack of syntactic variety, and underuse of nonfinite dependent clauses that are common in academic written discourse (Biber et al., 2011) are prevalent syntactic issues observed in adolescent writers' academic writing. The sentence boundary issues include run-on sentences, comma splices, and fragments. As Ferris (2014) noted, these syntactic issues also occur most often in college writers' texts for school and professional purposes. The patterns of and the dependence on the basic coordinating (*and, but*) and subordinating (*that, because, when*) conjunctions indicate lack of syntactic variety. And finally, nonfinite dependent clauses, including participials, absolutes, and appositives, are rare in the sample of the essays analyzed.

Figure 5.3

Types of Clauses and Connectives Used in the Learner Corpus



Note: FCC—Finite Complement Clause; FAC—Finite Adverbial Clause; FNM—Finite Noun Modifier; NFAC—Nonfinite Adverbial Clause; NFCC—Nonfinite Complement Clause; NFNM—Nonfinite Noun Modifier; FCC other wh-clauses include *why who, when, where*; FCC all other include *like, if, as though, whether*; FAC all other connectives include *as if, like, whether, what/whatever, as long as, by the time, once, where, as though, as … as, unless*; FNM other wh relativizers include *when, where, why*; Coordinated with other conjunctions include *so, or, yet, nor, for*.

As researchers have argued for the need to use more fine-grained measures at the phrasal level to assess syntactic complexity in student writing (Crossley & Kyle, 2018; Deng et al., 2021; Kyle, 2016), I used four fine-grained phrasal complexity indices and four fine-grained phrasal variety indices obtained from TAASSC (Kyle, 2016). Table 5.6 shows the mean scores of these fine-grained syntactic measures by grade level and their correlations with the holistic writing quality score and the analytic score that focuses on syntactic complexity and fluency. The mean score of the average number of dependents per nominal subject (M=.85; SD=0.33) was lower compared to the average number of dependents per nominal, direct object, and object of the preposition. This suggests that the essays, on average, exhibit less complex noun phrases structure per nominal subject. The mean scores and the standard deviations of the average number of dependents per nominal, nominal subjects, direct object, and the object of the preposition were relatively stable across grade levels, indicating that phrasal complexity and variety mean scores did not increase as grades advance. Based on the results of the Pearson Correlation, the average number of dependents per nominal, nominal subject, direct object, and object of the preposition were not significantly correlated with either the holistic score writing quality or the analytic score of sentence complexity and structure. However, the standard deviations of the number of dependents, which measure syntactic variety, were significantly correlated with both holistic and analytic scores (see Table 5.6). This suggests that high-scored essays on the human-judged measures of both holistic writing quality and analytic sentence complexity tend to exhibit a wider range of dependents per noun and prepositional phrases. Importantly, variety in noun phrase structures, rather than number of incidences, may be an important syntactic complexity feature of student writing. In other words, more may not mean better if more of the same or repeated syntactic structures are used.

Table 5.6

Phrasal Complexity and Variety Indices and their Correlations

	7th Grade Mean *(SD)*	8th Grade Mean *(SD)*	9th Grade Mean *(SD)*	10th Grade Mean *(SD)*	11th Grade Mean *(SD)*	12th Grade Mean *(SD)*	Full Sample Mean *(SD)*	Correlation with Holistic Score	Correlation with Analytic Score
Noun phrase complexity indices									
Dependents per nominal	1.06 (0.19)	0.99 (0.22)	1.04 (0.20)	0.97 (0.18)	1.04 (0.21)	1.05 (0.19)	1.02 (0.20)	0.026	0.007
Dependents per nominal subject	0.95 (0.36)	0.80 (0.35)	0.85 (0.27)	0.80 (0.32)	0.95 (0.34)	0.88 (0.35)	0.85 (0.33)	0.026	-0.071
Dependents per direct object	1.16 (0.39)	0.98 (0.47)	1.09 (0.46)	1.04 (0.49)	1.05 (0.50)	1.10 (0.44)	1.06 (0.46)	0.040	0.034
Dependents per object of the preposition	1.06 (0.35)	1.01 (0.33)	1.05 (0.35)	0.97 (0.26)	1.03 (0.32)	1.02 (0.24)	1.02 (0.30)	0.026	0.037
Noun phrase variety indices									
Dependents per nominal SD	1.00 (0.19)	1.00 (0.23)	1.00 (0.17)	1.03 (0.21)	1.03 (0.15)	1.06 (0.17)	1.02 (0.20)	0.125*	0.222***
Dependents per nominal subject SD	0.80 (0.25)	0.76 (0.28)	0.82 (0.25)	0.79 (0.28)	0.87 (0.26)	0.89 (0.23)	0.82 (0.26)	0.173**	0.176**
Dependents per direct object SD	0.92 (0.33)	0.78 (0.36)	0.79 (0.35)	0.83 (0.43)	0.75 (0.35)	0.90 (0.40)	0.83 (0.38)	0.254***	0.286***
Dependents per object of the preposition SD	0.88 (0.21)	0.88 (0.27)	0.94 (0.29)	0.94 (0.26)	0.93 (0.25)	0.95 (0.20)	0.92 (0.25)	0.216***	0.237***
Observations	27	74	62	76	16	85	340		

To illustrate the syntactic features within authentic texts, a paragraph from an essay that received a holistic score of 11 (out of 12) and a paragraph from an essay that is scored 4 are juxtaposed in Table 5.7. The independent and embedded clauses in the paragraphs are in brackets and numbered, complex noun phrases are underlined, and unconventional sentences that include run-ons and fragments are in italics. The syntactic features of the paragraph from the high-scored essay are characterized by complex noun phrase structures, as well as various types of clausal subordination and embedding, exhibiting more syntactic variety at both clause and phrase levels.

Table 5.7

Comparison of Syntactic Features Exhibited in High- and Low-scored Essays

A paragraph from an essay scored 11 (out of 12)	A paragraph from an essay scored 4 (out of 12)
The 7.0 magnitude earthquake (1) [that shook Port-au-Prince in 2008] is one of many (2) [that have claimed thousands of lives]. The colonialism, exploitation and constant devastation Haiti encountered has haunted its citizens, (3) [causing pain and suffering]. The problems (4) [facing this small nation (5) [bullied by the earth]] only further proves one point of many: the vulnerable are put through the worst. This is the recurring theme of the article (6) [written by Leonard Pitts (7) [after the 2008 tragedy in Haiti]]. Pitts' purpose is (8) [to send actual help to these people] (9) [because they are in great need]. We must help them rebuild and move on (10) [instead of watching them do it on their own]. The message is significant (11) [as we have the capability (12) [to help the vulnerable (13) [who are struggling through devastation and suffering]]].	This earth is cruel, with all of its huge earthquakes (1) [that shred cities and kill hundreds]. *Also all the hurricanes (2) [that sweep and blow trees of the ground] and all the floods (3) [that drown many]*. The world is cruel (4) [but in the story "Sometimes, the Earth is Cruel" it talks about the Haitians and (5) [how hard life they have]]. The story also shows us (6) [how they lived through the storms (7) [but had a tons of obstacles*]]. Like being one of the most poor countries in the world*. It also has a bad history of political instability. *Most of all they are the most prone to earthquakes they were hit by a 7.0-magnitude earthquake (8) [which is HUGE]*! But in the end they have to recover some how right? The man (9) [who wrote the story "Leonard Pitts"] is trying to show us (10) [that everyone in this world has to deal with something terrible (11) [but in the end we have to recover]].
Syntactic features at a clause level: • Total number of clauses: 13 • Number of distinct clauses: 10 • Number of distinct finite clauses: 5 • Number of distinct nonfinite clauses: 5	• Total number of clauses: 11 • Number of distinct clauses: 3 • Number of distinct finite dependent clauses: 2 • Number of distinct coordinated clauses: 1 • Number of unconventional sentences: 3

Syntactic features at a phrase level:
• Nominal subjects, subject complements, and direct object with complex noun phrase structures with the following constituents:
  • pre-modifying noun phrase (*the 7.0 magnitude*), adjectives (*constant, recurring*), and quantifier (*one of many*)
  • post-modifying finite relative clauses (*that shook Port-au-Prince in 2008; that have claimed thousands of live; who are struggling through devastation and suffering; Haiti encountered*)
  • post-modifying nonfinite dependent clauses — present participial (*facing this small nation*) and past participial (*bullied by the earth; written by Leonard Pitts after the 2008 tragedy in Haiti)*
  • post-modifying nonfinite complement to-clause controlled by noun (*capability to help the vulnerable who are struggling through devastation)*
  • post-modifying prepositional phrase (*of the article; in Haiti*)
  • coordinated noun phrases in parallel structure
	• Prepositional phrases, nominal subject, and direct object with complex noun phrase structures with the following constituents:
	  • pre-modifying adjectives (*huge, poor, bad*), quantifier (*one of*), noun phrase (*a 7.0-magnitude*)
	  • post-modifying finite relative clauses (*that shred the cities and kill hundreds; that sweep and blow trees of the ground; that drown many; which is huge; who wrote the story)*
	  • post-modifying prepositional phrases *(in the world; of political instability*)
	  • Appositive noun phrase (*Leonard Pitts)*

In the high-scored paragraph, out of the total 13 clauses that include both finite and nonfinite dependent clauses, 10 are unique clause types with distinct grammatical functions (noun modifier, adverbial, and complement). In addition, important syntactic features that distinguish the high-scored paragraph from the low-scored one are the absence of unconventional sentences and the presence of various types of nonfinite clauses including present participial (*causing pain and suffering; facing this small nation*), past participial (*bullied by the earth; written by Leonard Pitts after the 2008 tragedy in Haiti)*, infinitive or to-clause controlled by verb (is *to send actual help to these people*) and noun (capability *to help the vulnerable who are struggling through devastation and suffering),* and gerund clause following preposition (instead of *watching them do it on their own*). At a phrase level, the noun phrases that function as nominal subjects, direct objects, and subject complement include more and various types of pre- and post-modifiers, leading to more complex and varied phrasal structures. These syntactic features are characteristics of academic writing.

In contrast, the paragraph from the low-scored essay has repetitive use of the same structures —5 incidences of finite dependent clauses as noun modifiers (i.e., relative clauses) and 3 incidences of finite dependent clauses as complements (*that*-clause and *wh*-clause complement clauses controlled by verbs), and 3 incidences of coordinated/independent clause with the same conjunction *but*. These three structures alone constitute the unique clause types used in the paragraph. The absence of nonfinite dependent clauses and the presence of unconventional sentences including both run-on sentences and fragments indicate informal stylistic features that are common in colloquial discourse. At the phrase level, fewer nominal subjects and direct objects with complex noun phrase structures are observed throughout the paragraph. Two

distinct types of post-modifiers — finite relative clauses and prepositional phrases — are used several times throughout the paragraph (see Table 5.7).

Putting it all together, I now summarize the key findings of the linguistic features observed in text-based analytical writing of adolescent students. In general, texts in the sample of the study exhibit common linguistic patterns summarized below:

- Unconventional sentences (run-ons, fragments, and faulty sentences) and one-clause simple sentences were the most common syntactic patterns, together accounting for 55% of all sentences.
- Among subordinated and coordinated clauses, finite clauses with *that, because, when* connectives and coordination with *and* and *but* conjunctions were the most frequent structures, together constituting about 40% of clausal structures. This indicates overuse of common clause structures and underuse of varied and nonfinite clausal structures that are prevalent in academic written discourse.
- Simpler and less varied noun phrase structures were common characteristics of texts that scored low on holistic writing quality and analytic scoring for syntactic complexity.
- Repetitions of content and function words were a common lexical feature, indicating less lexical diversity.
- Higher percentage of frequent and familiar words and lower percentage of advanced and academic words were a prominent lexical feature observed in the learner corpus.

While these common features observed in adolescent writers' text-based analytical essays demonstrate successful linguistic performance for clear communication of ideas, they also point to challenges in linguistic choices based on genre conventions and audience expectations of academic written discourse. I now turn to rhetorical features of students' texts at a macro level,

focusing on cohesion, source use, and balanced use of summary, evidence, and commentary which are crucial rhetorical components of text-based analytical writing.

Rhetorical Features in Students' Text-based Analytical Writing (RQ1)

For rhetorical features in students' text-based analytical writing that requires students to develop an argument of literary analysis, I examined three areas: 1) text structure and cohesion, 2) source use and integration, and 3) balanced use of summary, evidence, and commentary for argument development. These three global discourse features are particularly important in text-based analytical writing in which students are expected to present their ideas cohesively following argument structure, draw on the source text for background and evidence, and provide interpretation and commentary to advance their argument.

Text Structure and Cohesion

The text structure was analyzed using a text-analytic approach (Sanders & Schilperoord, 2006) that involves analyzing how the rhetorical functions of syntactic segments establish the overall text organization and structure. The results, which are summarized in Table 5.8, indicated that summary or summary-dominant structures were the most prevalent, accounting for 48.06% of the texts. From the four structural patterns identified within the summary text type, the most common pattern was general summary or retelling of textual details followed by a short claim or commentary (19.38%). The next common pattern was mere summary without claim or commentary (12.40%). While 10.85% of the texts started with a claim or commentary expressed in a sentence or two and moved on to summarize or retell what the text is about, 5.43% followed similar structures but with reiteration of claim at the end of the summary. The summary-dominant patterns with a short statement of claim either at the beginning, end, or both showed an effort to address the prompt which explicitly asked students to make a claim about the theme of

the source text. However, it was common that students often resorted to summarizing the text rather than developing a claim and weaving in evidence and interpretations to support the claim.

Table 5.8

Text Structure: Types, Patterns, and Explanations of Patterns

Text Types	Structural Patterns	%	Explanation of Patterns
Summary/ Summary-dominant structures (48.06%)	Summary Only	12.40%	Only summarizes the text or retells what the text is about in chronological or additive orders.
	Summary + Claim/Commentary	19.38%	Summary or retelling dominates text: Mainly summarizes the text or retells what the text is about followed by a short (1-2 sentence) commentary or claim.
	Claim/Commentary + Summary	10.85%	Summary or retelling dominates text: Starts with a short commentary/claim (usually a single sentence) and summarizes the text or retells what the text is about.
	Claim/Commentary + Summary + Claim Reiteration	5.43%	Summary or retelling dominates text: Starts with a short claim/commentary (usually a single sentence), summarizes the text or retells what the text is about, and ends with short reiteration of claim (1-2 sentence).
Interpretive (29.97%)	Interpretation only	4.91%	Only provides interpretations based on an idea picked up from the text.
	Summary + Interpretation	5.17%	Interpretation dominates the text: Starts with a short summary or retell and provides lengthy interpretation
	Interpretation + Summary	1.55%	Interpretation dominates the text: Starts with lengthy interpretation and ends with short summary or retell.
	Point-to-point interpretations	18.35%	Moves from one point to another: Selects or summarizes several ideas/details from the text and comments on each.
Argumentative (20.67%)	Single point based on textual evidence	6.20%	The argument is supported by a single point based on textual evidence.
	Single point based on non-textual evidence	0.52%	The argument is supported by a single point based on non-textual evidence (general, anecdotal, and hypothetical examples).
	Multiple points based on textual evidence	12.14%	The argument is supported by multiple points based on textual evidence only.
	Multiple points based on textual and non-textual evidence	1.81%	The argument is supported by multiple points based on both textual and non-textual evidence (general, anecdotal, and hypothetical examples).
Other (1.29%)	Hypothetical	0.52%	No reference to text: Provides hypothetical situations and discusses personal responses to those situations.
	Expository	0.78%	No reference to text: Discusses general topics.

The second prevalent organization beyond summary structure was interpretive text type (29.97%), which centers around the interpretations of the main idea of the source text or of specific details from the text (or both in some cases). The most common interpretive structure was point-to-point interpretations (18.35%), in which the text moved from one point to another, selecting and summarizing several key points or details from the texts and providing an interpretation on each. Other interpretive structures included an interpretation only pattern (4.91%), summary followed by interpretations pattern (5.17%), and interpretations followed by summary pattern (1.55%). In the latter two structures, lengthy interpretations were provided following or followed by a short summary. Although text-based analytical writing requires developing an argument of literary analysis following logical modes of organization, the texts following argumentative structure accounted for only 20.67% of the sample, and only 13.95% of the texts had an argument structure with a claim supported by multiple points with evidence. Within the argumentative type, 6.72% of the texts exhibited an argument structure with a claim supported by a single point with evidence (Table 5.8).

Text cohesion was examined based on (1) an analytic scoring of essays for cohesive ties, and (2) two quantitative measures—lexical overlap across paragraphs and connectives—obtained from TAACO (Crossley et al., 2016). The results of analytic scoring based on a scale of 1 (ineffective) to 7 (effective) measuring the extent to which the ideas and sentences are connected to form a cohesive text showed that only 17% of the essays were given a score of 5 (competent) and above while 63% of the essays were scored 3 (developing) and below. The cohesion measure based on the analytic scoring was strongly correlated with writing quality score ($r = .65$, $p < 0.001$), indicating that sentences and ideas in texts scored high on quality were well connected using effective cohesive devices. The result of the analytic scoring was

consistent with the results based on the two quantitative measures of cohesion. The lexical overlap across paragraphs (M = .21; SD = .24), which indicates global cohesion, was significantly correlated with both the writing quality score (r = .44, p < 0.001) and the analytic score of cohesion (r = .38, p < 0.001). Connectives (all connective incidences), the second quantitative measure that was used to examine text cohesion, was also correlated (but weakly) with writing quality score (r = *.12*, p < 0.05) and analytic score of cohesion (r = *.16*, p < 0.01), indicating high scored texts used more explicit cues to establish cohesive links between ideas. Among all connectives (M = .069; SD = .023), additive connectives (*and, again, also, another)* were the most common (M = .047; SD = .019) while incidences of causal connectives (*therefore, because, as a result*) were much lower (M = .014; SD = .019). As cohesion is an important textual feature related to the interconnectivity of text segments, the results based on human-judged analytic scoring and NLP measures of local and global cohesion suggest that the ideas and sentences in high-quality essays, compared to low-scored essays, are well connected to form a cohesive text.

Source Use and Integration

Source use and integration were examined based on 1) analytic scoring for source use, 2) quantitative measure of source similarity (word2vec) obtained by TAACO, and 3) manual scoring of sentences for their rhetorical functions and for the types of source integration. Based on the results of the analytic scoring for source use (7-effective, 1-not used), the majority of the essays (73%) were scored 3 (developing) and below while only 12% of the essays received a score of 5 (competent) and above. The analytic scoring for source use specifically assesses the extent to which students integrated varied and relevant source material in a purposeful manner. The score of 1 (9.71% of the essays) indicated no or limited reference to the source, the score of

3 (39.71% of the essays) indicated the use of some source material but not varied and not often in a purposeful manner that contributes to the broader rhetorical goals. The qualitative measure of source use based on the analytic scoring was strongly correlated with writing quality score ($r = .67$, $p < 0.001$), indicating that texts scored high on quality exhibited effective use of source material by skillfully integrating varied and relevant source materials in a purposeful manner. The quantitative measure of source similarity (M = .76; SD = .08) based on word2vec similarity score between student text and the source text was also significantly correlated with both writing quality score ($r = .21$, $p < 0.001$) and analytic scoring for source use ($r = .22$, $p < 0.001$). As the task of literary analysis requires students to analyze the source text and extract relevant details from the text to develop their argument, a high source similarity score may indicate the use and integration of source to a greater extent.

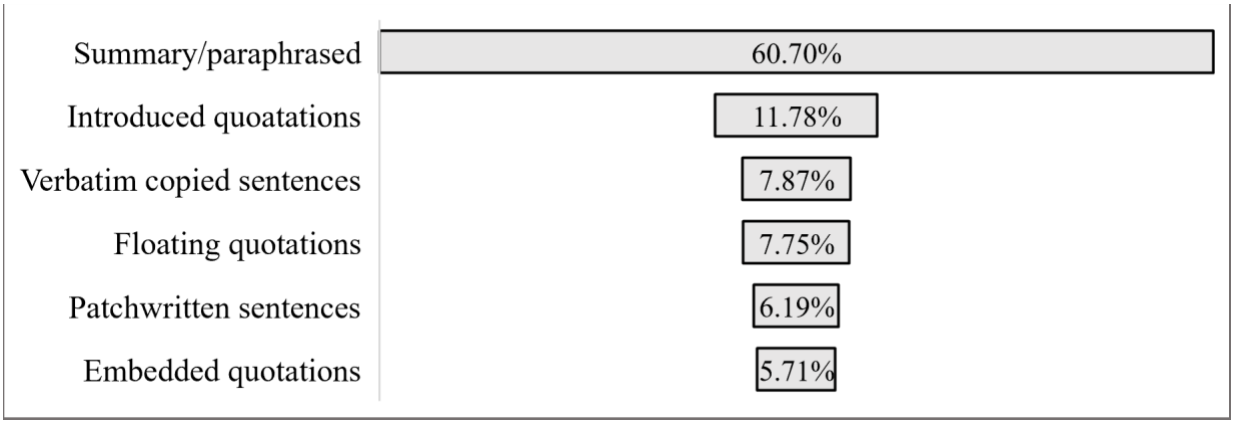
Using manual coding, I further examined the source use focusing specifically on source attribution (TAG – title, author, genre) and types of source integration (quotations, paraphrases, verbatim copied sentences). In the sample, only 14% of the essays included all three elements of source attribution (Title, Author, and Genre — TAG) while 40% of the essays had none of these three elements and summarized and discussed the source text without providing clear attribution. Among the essays that used one element of source attribution (28%), attributing the source with the genre (14%) using generic words such as *this article* and *this story* was more frequent than attributing sources using author (9%) and title (5%). Among the essays that used two elements of source attribution (19%), using title and genre was more common than using title and author (7.18%) or author and genre (2.93) to attribute the source text.

In terms of the types of source integration, the results based on manual coding of sentences indicated that the majority of the sentences integrated from the source text were

summary or paraphrased sentences (61%). Verbatim copied sentences (word-for-word copied sentences without quotation marks) and patch-written sentences (copied sentences with minor modifications and without quotation marks) together constituted 14% of the sentences, indicating a common pattern after the summary and paraphrased sentences. Introduced quotations, which refer to quoted sentences integrated with introductory phrases, accounted for about 12% of the borrowed sentences, with a high frequency of an introductory phrase *according to …* and the reporting verbs *say* and *explain.* Embedded quotations (6%), which is a more sophisticated way of integrating quoted phrases into sentences, were the least common among the types of source integration. These results are presented in Figure 5.4. In general, high-quality texts exhibited a more sophisticated use of source material with purposeful summary and paraphrasing and quotations that were embedded into students' own sentences and integrated with introductory phrases. However, essays that received low scores on both holistic scoring and analytic scoring for source use had verbatim copied sentences, floating or stand-alone quotations that did not serve as evidence, and low use of paraphrasing and purposeful summary. In essays that followed the format of general summary and retelling, the source integration was often general summary rather than strategic use of textual borrowing to support a claim and develop an argument.

Figure 5.4

Types of Source Integration Based on Manual Coding of Sentence



Balancing Summary, Evidence, and Commentary

An important rhetorical feature of text-based analytic writing is the extent to which the writing exhibits a balanced use of summary, evidence, and commentary. The results of analytic scoring for this rhetorical feature showed that the majority of the essays (78%) were scored 3 (developing) and below while only 9% of the essays received a score of 5 (competent) and above, indicating a challenge of skillfully weaving together summary, evidence, and commentary for a higher rhetorical purpose. The analytic score for the balanced use of summary, evidence, and commentary was strongly correlated with the writing quality score ($r = .71$, $p < 0.001$), indicating that high-scored texts exhibited more skilled integration and the balanced use of the three elements (summary, evidence, and commentary). For example, the sample paragraph from a high-scored essay presented below demonstrates a balanced use of summary, evidence, and commentary (summary and evidence in italics and commentary underlined), integrating varied types of source material (summary, paraphrasing and quotation) and weaving interpretive comments into the paragraph for the higher rhetorical goal of advancing an argument.

A body paragraph from a high-scored essay:

> *Like everyone else, the man in the water didn't know the plane was going to crash, like everyone else he was an ordinary man. Then the plane crashes and he decides to put the other survivors before himself.* Rosenblatt uses powerful descriptions of the crash and the man's action to help us, the readers, understand how an ordinary man's character stands against the overwhelming power of nature. He writes, *"Here were two forms of nature in collision: the elements and human character … the elements, indifferent as ever, brought down Flight 90, but human nature rose to the occasion"* (page 1, paragraph 2). While we see the powerful effect brought out from the crash, we also see the beauty behind it.

While the nature decided on destruction and death, the man in the water decided on sacrifice and selfless heroism. It is a testament to the relationship between nature and humankind, two opposing forces that together create something utterly impactful.

In the paragraph above, the sentences in italics are summary/paraphrased sentences and quotations integrated from the source text. Underlined sentences are commentary reflecting interpretations, explanations, and connections. The paragraph starts with a paraphrased sentence about the man and then moves on to a summary sentence that retells the key event. The third sentence which is underlined is commentary on the author's craft, which is the key point of the paragraph. The following sentence is a quotation integrated with an introductory phrase. This quoted sentence is used as textual evidence to support the key point. The paragraph ends with sentences commenting on and interpreting the textual evidence and making connections to the key point.

While the example of the high-scored text indicates purposeful and strategic rhetorical moves marked by a skillful use of textual evidence substantiated with interpretation, reasoning, and commentary, the majority of the texts in the learner corpus showed either overreliance on summary and retelling without commentary and interpretations (48.06%) or predominance of interpretations without textual evidence (20.41%). In Table 5.9 below, sample body paragraphs from two different essays are juxtaposed to show the two patterns identified in the sample of text-based analytical essays. The sample paragraph 1 presented in Table 5.9 shows overreliance on summary and retelling without offering much commentary and interpretations. In fact, this is a common shortcoming observed in student writing in the sample of this study. On the other hand, sample paragraph 2 was developed by predominantly interpretive comments without textual evidence.

Table 5.9

Sample Paragraphs: Summary and Interpretive Structures

Sample 1: Predominantly summary/retelling	**Sample 2: Predominantly interpretive**
Furthermore, the terrible situation of the crash in the Man in the Water was described to show the man risked his life to save others. It was one man risking his life to save others. The Man in the Water cared for other people. First, when the helicopter came back to pick up another survivor, he kept on handing the rope to save others, but not himself. The article states, "... he must have realized he would not live if he continued to hand over the rope and ring to others." Second, no matter how cold the river was he didn't stop trying to rescue others. In the article, it sates, "... no matter how gradual the effect of the cold." This shows the Man in the Water cared for other people more then himself.	*After a tragic plane crash, an unidentified man helps save the survivor*. In the situation of disaster like the plane crash, everything seems chaotic, overwhelming, and loud. It is easy for people to transform into something they are not. They will do anything in their power to survive, to persevere and to surpass those around them. The man in the water is the exception. His selfless persona placed the need of others before his own. The ability to become that in a time of implacable chaos is a power. Not everyone places in the same situation would react the same. It takes a special person, a person with power, power to overcome obstacles.

Note. Sentences in italics are summary/paraphrased sentences and quotations integrated from the source text. Underlined sentences are commentary reflecting interpretations, explanations, and connections.

The paragraph characterized by predominantly summary and retelling is from an essay scored 6 (out of 12) on a holistic writing quality and 3 (developing) on analytic scoring that focuses on the extent to which the essay demonstrates a balanced use of summary, evidence, and commentary. While there is some element of commentary as indicated in the last sentence of the paragraph (underlined), the paragraph largely summarized the details stated in the source text without offering much commentary and interpretations. On the other hand, the predominantly interpretive paragraph from a text, which was scored 5 on holistic writing quality and 3 (developing) on analytic scoring, provides mainly interpretive comments after the summary sentence (italicized) without providing much textual evidence. In general, the results indicate that this area of weaving summary, evidence, and commentary in a balanced way is indeed a challenging task for many students as it involves complex rhetorical problem-solving.

Taken together, the results of the analysis of rhetorical features in the text-based analytical writing of adolescent students indicated considerable challenges with higher-order tasks involved in writing arguments of literary analysis. Text-based analytical writing demands argument structure and organization that are different from narrative and summary writing, strategic use of source materials for argument development, and skillful weaving of summary, evidence, and commentary. While these higher order tasks involving complex rhetorical problem solving are demanded by the writing task, the following common rhetorical features were observed in the learner corpus:

- Summary and summary-dominant structures were the most prevalent text structures, accounting for 48% of the texts.
- Low-scored essays exhibited low global cohesion indicated by fewer lexical overlaps between paragraphs.
- Additive connectives (e.g., *and, also, another*) were the most used connectives for local cohesion.
- General summary and paraphrasing were the most common types of source integration, accounting for 61% of the source material integrated in writing.
- Not attributing the source and attributing the source with generic words (*this story, this article*) were common (54% of the texts).
- Verbatim copied and patchwritten sentences were more common than embedded quotes and quotations integrated with introductory phrases.
- Overreliance on summary without commentary and predominance of interpretative comments without textual evidence were common rhetorical patterns.

These rhetorical features observed in the sample of text-based analytical writing demonstrate several strengths, including comprehension of the source text and proficiency in summary writing. However, they also indicate both the need and the challenge for students to move beyond what Bereiter and Scardamalia (1987) referred to as "knowledge telling" or summary and retelling to an advanced level of "knowledge transformation" that involves analyzing, arguing, interpreting, and using textual evidence strategically for the broader rhetorical goals associated with argument construction.

Examining the Linguistic and Rhetorical Features by L1 Groups (RQ2)

The second research question of the study was to examine whether the identified lexical, syntactic, and rhetorical features vary among three language groups. The language grouping was based on students' L1s and their linguistic distances to English. To examine whether the identified lexical, syntactic, and rhetorical features vary among the groups, MANOVA was first conducted for each feature. For the composite features (lexical, syntactic, rhetorical) that showed significant differences based on MANOVA results, post hoc tests were run. The follow-up tests included multiple univariate analysis of variance (ANOVA) on each specific component of the composite features that showed significant differences among groups. For each component that showed a significant difference on ANOVA, Bonferroni post hoc analyses were employed for pairwise comparisons between the three groups. Table 5.10 shows descriptive statistics of the specific components of the lexical, syntactic, and rhetorical features that were used in the analyses of variance. As indicated in Table 5.10, the skewness and kurtosis for each variable were within the suggested normality thresholds of 2.0 (for skewness) and 7.0 (for kurtosis) when assessing multivariate normality which is assumed in factor analysis and MANOVA (Byrne, 2010; Curran et al., 1996).

Table 5.10

Descriptive Statistics: Rhetorical and Linguistic Features by L1 Groups

	L1-English		L1-Romance Languages		L1-All Other Languages		Full Sample			
	M	SD	M	SD	M	SD	M	SD	Skew-ness	Kurt-osis
Rhetorical Features										
Source Use$_{(as)}$	2.98	1.32	3.13	1.21	2.79	0.98	3.00	1.19	0.47	3.21
SEC Balance$_{(as)}$	2.64	1.30	2.75	1.37	2.30	1.00	2.60	1.27	0.59	2.99
Cohesion$_{(as)}$	3.28	1.31	3.23	1.36	2.91	1.33	3.16	1.35	0.43	2.83
Global Cohesion$_{(nlp)}$	0.24	0.25	0.19	0.23	0.21	0.25	0.21	0.24	0.52	1.84
Connectives$_{(nlp)}$	0.07	0.02	0.07	0.02	0.07	0.03	0.07	0.02	0.16	2.93
Syntactic Features										
Syntactic Complexity$_{(as)}$	3.13	0.99	2.85	1.15	2.64	0.97	2.87	1.07	0.32	3.28
Mean Length of Clause$_{(nlp)}$	8.00	1.33	8.06	1.47	8.26	1.96	8.10	1.59	1.12	5.74
Unconventional sent-s$_{(mc)}$	0.35	0.21	0.36	0.25	0.39	0.22	0.36	0.23	0.53	2.89
Complex/Compound$_{(mc)}$	0.38	0.19	0.40	0.24	0.34	0.22	0.37	0.22	0.29	2.92
Noun Phrase Complexity*	0.94	0.20	0.99	0.23	1.02	0.20	0.99	0.21	0.15	5.14
Noun Phrase Variety*	0.92	0.17	0.90	0.18	0.87	0.17	0.90	0.18	-0.37	4.79
Lexical Features										
Lexical Sophistication$_{(as)}$	3.52	1.14	3.34	1.39	2.84	1.03	3.26	1.26	0.39	3.25
Lexical Diversity$_{(nlp)}$	0.66	0.10	0.69	0.11	0.64	0.10	0.67	0.10	0.05	3.26
K1 Words$_{(nlp)}$	87.62	4.65	86.44	4.74	88.35	4.76	87.29	4.78	-0.29	3.09
AWL Words$_{(nlp)}$	2.25	1.61	2.49	1.77	1.85	1.40	2.25	1.65	0.83	3.80
CW Familiarity$_{(nlp)}$	582.7	6.41	581.3	6.48	585.9	8.39	582.9	7.26	0.34	3.55
Age of Acquisition$_{(nlp)}$	5.03	0.29	5.08	0.29	4.91	0.29	5.02	0.30	0.19	2.89
Observations	94		154		92		340			

Note. as—analytic scoring variable; nlp—variable obtained from Natural Language Processing tool; mc—manual coding variable; *a latent variable with a factor score

All the variables, except for noun phrase complexity and noun phrase variety, listed in Table 5.10 are observed/measured variables obtained from analytic scoring, manual coding, and NLP tools. The two variables—noun phrase complexity and noun phrase variety—are latent variables with factor scores. To reduce the number of items for noun phrase complexity and variety, I conducted factor analysis and generated factor scores for noun phrase complexity and

noun phrase variety. The four noun phrase complexity variables obtained from TAASSC, which are dependents per nominal, dependent per nominal subject, dependents per direct object, and dependent per object of the preposition, were used to create a noun phrase complexity factor score. The factor loadings of all four variables were between 0.51 and 0.91, all above the threshold of 0.4 that indicates the item is considered a good indicator of the factor (Acock, 2008). The four noun phrase variety indices from TAASC, which are the standard deviations of the same four noun phrase complexity variables, were used to create a noun phrase variety factor score. The factor loadings of all four noun phrase variety indices were above 0.4 threshold (all between 0.43 and 0.91). Instead of simply averaging the items, a factor score was generated for each latent variable since "it weighs each item based on how related it is to the factor" (Acock, 2008).

The results of MANOVA analyses showed no significant differences among the L1 groups in the rhetorical features λ =0.96, $F(10, 666) = 1.42$, $p = 0.165$. However, significant differences were found in syntactic (λ=0.93, $F(12, 664) = 2.04$, $p = 0.019$) and lexical (λ =0.86, $F(12, 664) = 4.25$, $p < 0.001$) features among L1 groups (Table 5.11). Following the MANOVA analyses, multiple univariate ANOVAs were conducted to examine which specific components of the syntactic and lexical features varied between the groups. In the area of syntactic features, there were significant differences among language groups in two syntactic components (out of six listed in Table 5.10)—syntactic complexity based on analytic scoring ($F(2, 337) = 4.96$, $p = 0.008$) and noun phrase complexity ($F(2, 337) = 3.09$, $p = 0.046$). No significant differences among groups were found in the other four syntactic components: 1) unconventional sentences, 2) complex and compound sentences, 3) mean length of clause, and 4) noun phrase variety. However, with adjusted a level of 0.017 for multiple comparisons after Bonferroni correction for

potential Type 1 error inflation that may result from multiple comparisons, only syntactic complexity based on analytic scoring was significant. Pairwise comparisons based on the post hoc Bonferroni test indicated a significance difference only between L1-English and L1-All Other Languages groups in syntactic complexity measure based on analytic scoring (*r = -0.49, p* = 0.006), suggesting that texts in L1-All Other Languages group had lower syntactic complexity scores compared to those in L1-English group. However, there were no significant differences between L1-English and L1-Romance Languages groups and L1-Romance Languages and L1-All Other Languages groups.

Table 5.11

MANOVA Results for Rhetorical, Syntactic, and Lexical Features

Writing Features	Wilks' λ	$d.f._{num}$	$d.f._{den}$	*F*	*p*
Rhetorical Features	0.959	10	666	1.42	0.165
Syntactic Features	0.930	12	664	2.04	0.019
Lexical Features	0.862	12	664	4.25	0.000

In the area of lexical features, significant differences among language groups were found in all six lexical components. On the lexical sophistication measure based on analytic scoring ($F(2, 337) = 7.85$, $p = 0.001$), the pairwise comparisons based on the post hoc Bonferroni test showed no significant difference between L1-English and L1-Romance Languages groups; however, there were significant differences between L1-English and L1-All Other Languages groups (*r = -0.68, p* = 0.001), as well as L1-Romance Languages and L1-All other languages groups (*r = -0.51, p* = 0.006), indicating that texts in L1-All Other Languages had lower scores of lexical sophistication compared to those in the other two groups. On the measures of lexical diversity ($F(2, 337) = 6.51$, $p = 0.002$), percentage of K1 words ($F(2, 337) = 5.06$, $p = 0.007$), and percentage of AWL words ($F(2, 337) = 4.36$, $p = 0.014$), significant differences were found

only between L1-Romance Languages and L1-All Other Language Groups. On the measures of content word familiarity ($F(2, 337) = 12.06$, $p < 0.001$) and age of acquisition ($F(2, 337) = 10.43$, $p < 0.001$), significant differences were found between L1-English and L1-All Other Languages ($r = 3.19$, $p = 0.006$ for content word familiarity and $r = -0.13$, $p = 0.008$ for age of acquisition) and between L1-Romance Languages and L1-All Other Languages groups ($r = 4.53$, $p < 0.001$ for content word familiarity and $r = -0.17$, $p < 0.001$ for age of acquisition). This suggests that texts in L1-All Other Languages groups, compared to those in L1-English and L1-Romance languages groups, had a higher content word familiarity score, indicating use of more familiar and basic words, and a lower age of acquisition score, indicating use of words from the lower tier of vocabulary level. The pairwise comparisons between groups on the syntactic and lexical components that showed significant differences are exhibited in Table 5.12 below.

Table 5.12

Pairwise Comparisons between Three L1 Groups

	Are there differences between these groups?		
	Gr1 vs. Gr2	Gr1 vs. Gr3	Gr2 vs. Gr3
Syntactic complexity based on analytic scoring	No	Yes	No
Noun phrase variety factor score*	No	Yes	No
Lexical Sophistication based on analytic scoring	No	Yes	Yes
Content word familiarity	No	Yes	Yes
Age of acquisition for all words	No	Yes	Yes
Lexical Diversity (content word TTR)	No	No	Yes
K1 words (frequency based on GSL 1-1001)	No	No	Yes
AWL words	No	No	Yes

Note. Gr1—L1 English; Gr2—L1 Romance Languages; Gr3—L1 All Other Languages; * significance was based on 0.05 alpha level without Bonferroni correction.

To encapsulate the results of the second research question in a succinct way, I now highlight the key findings. Based on the results of MANOVA, multiple univariate ANOVAs, and Bonferroni post hoc analyses, the key findings were:

- The texts in three groups did not significantly differ in rhetorical features that included source use and integration, balance of summary, evidence, and commentary, and cohesion.
- Significant differences were found in some syntactic features between L1-English and L1-All Other Languages groups only. Texts written in all groups did not differ in clause length, production of unconventional and complex sentences, and noun phrase variety. The syntactic differences between the two groups appear to be in human-judged analytic scoring and noun phrase complexity measures.
- No significant differences were found between L1-English and L1-Romance Languages groups on any of the syntactic and lexical components.
- Texts in L1-All Languages group significantly differed from those in L1-Romance Languages group in all lexical features, suggesting that texts produced by students in L1-All Other Languages group had significantly lower score on lexical sophistication and showed higher percentages of words that are familiar and basic.
- Texts in L1-All Languages group significantly differed from those in L1-English group on only three lexical components—lexical diversity, K1 words, and AWL words, suggesting that texts written by students in L1-All Other Language group had significantly lower use of AWL words, higher use of K1 (basic) words and less lexical diversity compared to the texts in L1-English group.

These findings suggest that all students in the sample of this study, regardless of their home and first language backgrounds, had similar performance on the identified rhetorical features that include source use, balance of summary, evidence, and commentary, and textual cohesion. While there was no significant difference in linguistic features of texts between L1-English and L1-

Romance Languages groups, texts written by students in the L1-All Other Languages group indicate significantly low performance on linguistic features.

Relationships among Rhetorical Features, Linguistic Features, and Holistic Writing Quality (RQ3)

The third research question of the study examined the extent to which lexical sophistication, syntactic complexity, and rhetorical features relate to one another and predict a human-judged holistic writing quality score. Table 5.13 presents bivariate correlations between measured variables that were selected for CFA and structural regression. For the measures of rhetorical features, source use and integration, balanced use of summary evidence and commentary (SEC), and global cohesion were selected. While the first two measures were based on analytic scoring, global cohesion was measured by lexical overlap between paragraphs based using an index obtained from the computational tool TAACO. For syntactic features, I used a syntactic complexity measure based on analytic scoring, ratio of conventional sentences based on manual coding, and noun phrase variety (average score of 4 noun phrase variety measures) derived from TAASSC. For lexical features, I used a lexical sophistication measure based on analytic scoring and two measures obtained from TAALES—AWL words and Age of Acquisition. The selection of the variables was based on the literature, findings of previous studies, and their correlations with one another and the writing quality score. The issue of multicollinearity and non-significant correlations with writing quality were also considered in the selection of the variables. The correlations between all selected variables, except the correlation between noun phrase variety and conventional sentences, were statistically significant. All selected variables were weakly to strongly related to the holistic writing quality score (r = .26 to r = .66).

Table 5.13

Correlation Matrix: Rhetorical, Syntactic, Lexical Measures Used in SEM

		1	2	3	4	5	6	7	8	9
1.	Writing quality score	-								
2.	Source use and integration$_{(as)}$	.64***	-							
3.	Balanced use of SEC[1]$_{(as)}$	.66***	.70***	-						
4.	Global cohesion[2]$_{(nlp)}$	.41***	.31***	.37***	-					
5.	Syntactic complexity$_{(as)}$	.61***	.67***	.69***	0.35***	-				
6.	Conventional sentences$_{(mc)}$	.35***	.32***	.29***	0.19***	0.34***	-			
7.	Noun phrase variety[3]$_{(nlp)}$	.30***	.29***	.30***	0.10*	0.35***	0.02	-		
8.	Lexical sophistication$_{(as)}$	.64***	.66***	.74***	0.37***	0.77***	0.29***	0.39***	-	
9.	AWL covered word$_{(nlp)}$	.26***	.31***	.29***	0.11*	0.25***	0.19***	0.14**	0.37***	-
10.	Age of acquisition$_{(nlp)}$	.30***	.27***	.35***	0.21***	0.28***	0.16**	0.20***	0.43***	0.46***

Note. N=410; ***** p<0.05, ** p<0.01, *** p<0.001; 1. SEC – Summary, Evidence, Commentary; 2. global cohesion was measured by lexical overlap between paragraphs (TAACO variable name is adjacent_overlap_2all_para); 3. Noun phrase variety is average score of four noun phrase variety indices (TAASSC variable names are nominal_deps_NN_stdev, nsubj_NN_stdev, dobj_NN_stdev, pobj_NN_stdev)

To examine how the rhetorical features relate to the lexical and syntactic features and how these features predict human-judged holistic writing quality, I first ran confirmatory factor analysis (CFA) to test three models and then conducted structural regression using the best-fitting model. The model fits of the three competing models are presented in Table 5.14. The three-factor model in which the lexical, syntactic, and rhetorical features are latent constructs that covary with one another had a close fit ($\chi^2 = 104.44$, *df* = 24, CFI = .95, SRMR = .05) and was the best-fitting model. Both model 1 (SRMR=.053; CFI=.925) and model 2 (SRMR=.052;

CFI=.937) had acceptable fits based on SRMR and CFI statistics, but model 1 (RMSEA=.100) had a poor fit and model 2 ((RMSEA=.0.94) had a mediocre fit based on RMSEA. Chi-square difference test showed that the three-factor model was superior ($\Delta\chi^2$=15.1, Δdf=2, p<.001) to the other two models. As the three-factor model with a close fit was superior to the one-factor and two-factor models based on all goodness-of-fit measures used, it was chosen as the final model.

Table 5.14

Model Fits and Comparisons Based on Confirmatory Factor Analysis

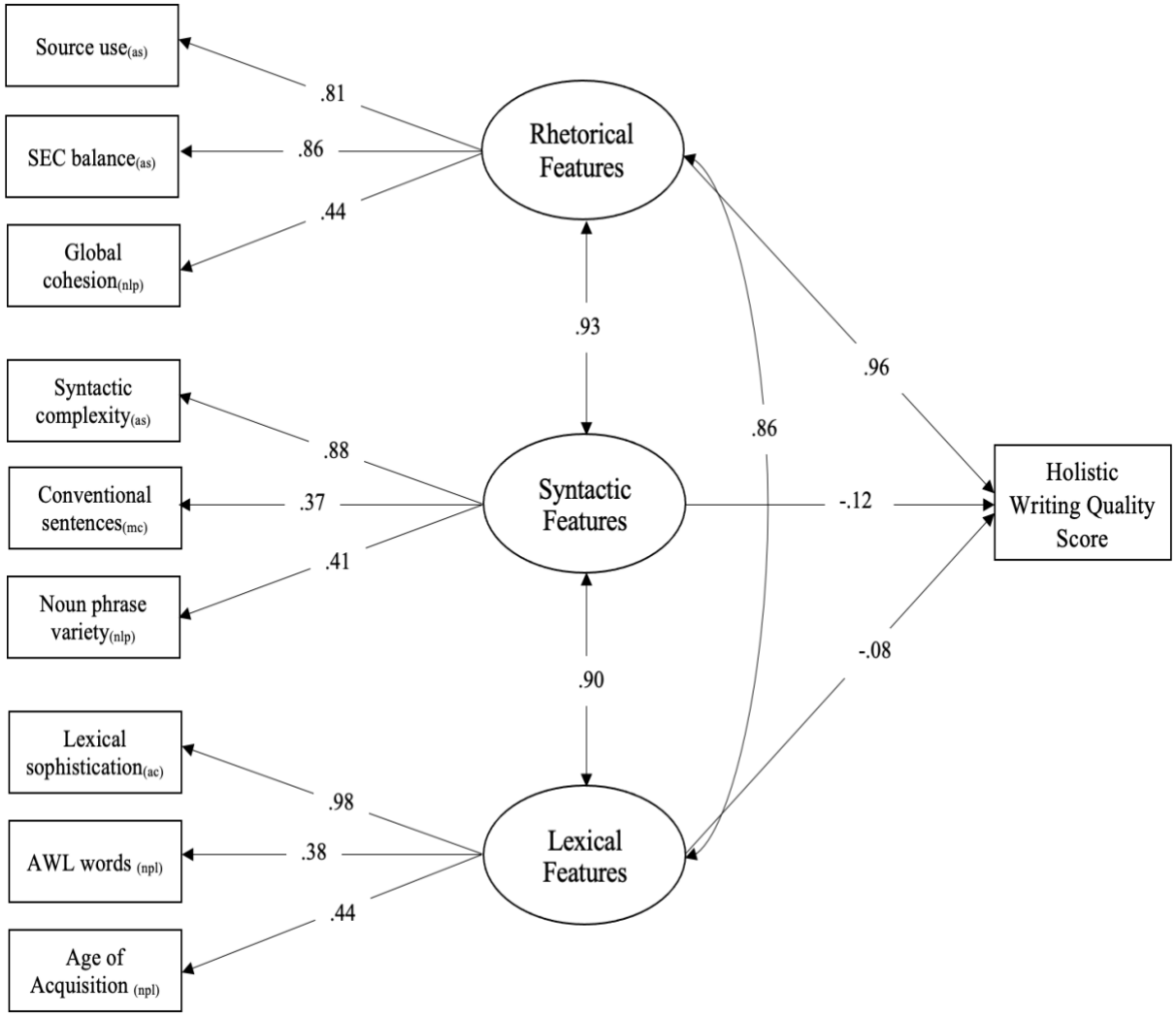
	χ^2	*df*	*p* value	RMSEA	RMSEA 90% CI	CFI	SRMR	Model Comparisons
Model 1: One-factor	137.56	27	<.001	.100	.084; .117	.925	.053	
Model 2: Two-factor	119.56	26	<.001	.094	.077; .111	.937	.052	$\Delta\chi^2$=18, Δdf=1, p<.001
Model 3: Three-Factor	104.44	24	<.001	.090	.073; .109	.946	.050	$\Delta\chi^2$=15.1, Δdf=2, p<.001

The final model was used in structural regression to examine both the relations among the latent constructs of rhetorical, syntactic, lexical features and the extent to which they predict the human-judged holistic writing quality score. Figure 5.5 shows the structural equation model (standardized coefficients presented) with the latent constructs of rhetorical, syntactic, and lexical features as predictors of the holistic writing quality. The fit of the structural regression model was good, χ^2 = 127.93, *df* = 30, CFI = .946, SRMR = .048. The latent construct of rhetorical features was strongly associated with the human-judged holistic writing quality, β = .959, $p < 0.001$. However, the latent constructs of syntactic features (β = -.115, $p = 0.662$) and lexical features (β = -.076, $p = 0.548$) were not significant predictors of holistic writing quality in the structural regression model. The bivariate correlations between the latent constructs demonstrated that all three features were strongly correlated with one another. The latent construct of rhetorical features was strongly correlated with the latent constructs of syntactic features ($r = .925$, p<.001) and lexical features ($r = .864$, p<.001). The latent constructs of

syntactic and lexical features were also highly correlated (r = *.900*, p<.001). This suggests that syntactic and lexical features, although not significant predictors of writing quality score in the structural regression model, were strongly associated with the latent rhetorical features and with one another, indicating shared variance. It is also important to note that the high correlations between the latent constructs may indicate a multicollinearity issue in the structural regression model.

Figure 5.5

Structural Equation Model with Latent Factors as Predictors



Putting together the results of the third research question that investigated how lexical, syntactic, and rhetorical features relate to one another and predict holistic writing quality, I now highlight the key findings.

- The three-factor model in which lexical, syntactic, and rhetorical features are distinct yet related factors fitted well to the data and was superior to the unidimensional and two-factor models, indicating that lexical, syntactic, and rhetorical features are distinct and dissociable dimensions of writing.
- The latent construct of rhetorical features significantly predicted holistic writing quality while the latent constructs of lexical and syntactic features did not predict holistic writing quality in the structural regression model.
- Latent constructs of lexical, syntactic, and rhetorical features were strongly related to each other, indicating shared variance.

While the results showing linguistic features, including both lexical and syntactic latent constructs, being not significant predictors of writing quality were unexpected, it could be further hypothesized that lexical and syntactic features contribute to rhetorical features, which then predicts writing quality. In other words, lexicon and syntax as micro-level features are crucial in expression and forming of ideas at the macro rhetorical level, and the quality of the ideas formed is what ultimately predicts writing score.

Discussion

In this study, I systematically analyzed the syntactic, lexical, and rhetorical features of the text-based analytical essays written by linguistically diverse adolescent students attending public schools in several states in the United States. The purpose of the study was to identify common patterns of the lexical, syntactic, and rhetorical features exhibited in students' academic texts,

analyze whether the identified features vary among L1 groups, and examine how the identified features relate to one another and to the writing quality. In what follows, I discuss the results of each research question, followed by the limitations of the study.

Exploring linguistic and rhetorical features to understand writing challenges (RQ1)

Based on the systematic analyses of the texts involving manual coding, analytic scoring, and NLP tools, I identified several common linguistic and rhetorical features in the text-based analytical essays in the selected sample (n=340). The common linguistic features include patterns of unconventional sentences (run-ons, fragments, and faulty sentences) and one clause simple sentences, overuse of subordination and coordination that are common in colloquial discourse and underuse of varied and sophisticated subordination that are common in academic writing, lack of syntactic variety, dependence on frequent and familiar words, low use of advanced and academic vocabulary, and repetitions of content and function words. While these common features observed in adolescent writers' text-based analytical essays demonstrate successful linguistic performance drawing on their existing linguistic repertoire for clear communication, they also point to challenges in linguistic choices based on genre conventions and audience expectations of academic written discourse. Without knowledge of specialized language conventions in academic written discourse and formal register, developing writers are constrained as they engage in complex writing tasks that require higher-order analytical skills (Kellogg, 2008; Olson et al., 2015; Shaughnessy, 1977).

The common rhetorical patterns that were identified based on the text analytic approach include dependence on summary and retelling rather than argumentation structure, less use of varied cohesive devices and more use of additive connectives, low use of strategic and sophisticated textual borrowings as evidence to support a claim. These rhetorical features

observed in the sample of text-based analytical writing demonstrate that the adolescent writers showed strength in several areas, including comprehension of the source text and proficiency in summary writing. However, these patterns also indicate both the need and the challenge for students to move beyond "knowledge telling" that is characterized by summary and retelling to an advanced level of "knowledge transformation" (Bereiter & Scardamalia, 1987) that involves more complex rhetorical problem-solving skills demanded by the higher-order tasks of analyzing, arguing, interpreting, and using textual evidence strategically for the broader rhetorical goals associated with argument construction. Text-based analytical writing involves developing arguments of literary analysis, requiring students to analyze a source text, construct an argument of interpretive analysis, and use the source text for evidence (Olson et al, forthcoming). To successfully complete these tasks, students need to produce a text that is marked by well-informed and focused argument, evidence-based development, and textual coherence based on logical order.

These results are consistent with the findings of the first two studies of the dissertation, as well as other studies that analyzed the features of text-based analytical essays produced by adolescent writers in secondary school (Chen, et al., 2020; Olson et al., forthcoming). Students' unfamiliarity with specific conventions of the text-based analytical writing genre and the formalized language features of academic written discourse may prevent them from making informed choices with an understanding of communicative goals, rhetorical purposes, and discursive norms based on genre conventions and audience expectations. For students who are still in the process of developing their academic language proficiency, the demands of academic writing are much greater. Effective writing and argumentation in an academic context depend on advanced literacy and analytical skills, and text-based argument is a "threshold concept" that

requires in-depth understanding of rhetorical situations (genre, audience, context) and communicative purposes (Fletcher, 2015). The complexity of argument writing and the shifting instructional focus toward argument literacy have placed greater responsibility for teachers to find effective ways to support students in developing both argumentation skills and academic language. To this end, the study's goal of identifying common linguistic and rhetorical features in students' texts not only paints a descriptive picture of their academic writing but, importantly, provides us with a better understanding of the challenges and needs in students' engagement in academic writing tasks. Such understanding is pivotal in our collective effort of supporting students to develop their academic literacy skills.

Comparing L1 groups on the identified features

The second research question of the study aimed to compare three different L1 groups (L1-English, L1-Romance Languages, L-All Other Languages) on the rhetorical, syntactic, and lexical features that were identified. The results suggest that all students in the sample, regardless of their home and first language backgrounds, had similar performance on the identified rhetorical features that include source use and integration, balance of summary, evidence, and commentary, and textual cohesion. In addition, there were no significant differences between L1-English and L1-Romance languages groups in linguistic features, including both syntactic complexity and lexical sophistication measures of their texts. However, L1-All Other Languages groups differed significantly from both L1-English and L1-Romance Languages groups on some syntactic features and on all lexical components. Regarding the syntactic features, while there was no difference in large-grained, global syntactic measures (mean length of clause, production of unconventional sentences, use of complex and compound sentences) among all three groups, L1-All Other Languages group had significantly lower performance, compared to the other two

groups, on fine-grained phrasal complexity and on the measure of syntactic complexity based on human-judged analytic scoring. On lexical features, texts written by students in L1-All Other Languages had significantly lower performance on all lexical components, including lexical diversity and lexical sophistication measures derived from both human-judged analytic scoring and computational NLP tools.

These findings are novel as the current study, to my knowledge, is the first to examine the rhetorical and linguistic features in tandem among three different L1 groups. Although there is an abundance of research that compares and controls for English only (EO) and English Learner (EL) status at a secondary level and L1 and L2 status at a postsecondary context, to my knowledge, there is no other study that examined differences in rhetorical, syntactic, and lexical features of text-based analytical writing among adolescent students with different home/first language backgrounds at a secondary education level. The results of this study attest to the complexity of and the variability in the writing skills and development of adolescent multilingual students with differing linguistic and educational backgrounds. Since students' L1, L1 language family, and how distant the languages are from one another were factors considered for grouping, the linguistic distance (i.e., linguistic distance between English and Romance language is lower, compared to that of English and all other languages in the third group) may have contributed to the variability among groups. Although linguistic distance may be considered as a contributing factor, students' educational experiences, years of language development, exposure to language and literacy practices, access to resources and support for literacy development are important factors that may explain the differential linguistic performances. For example, L1-All Other Languages group consists of students whose L1s are Somali, Hmong, Arabic, Burmese, Amharic, Karen and other Afroasiatic, Niger-Congo, and Sino-Tibetan languages that represent

communities that have been affected by civil unrest and migration. Many of these students may be English learners who have experienced interrupted formal education and limited schooling opportunities for a variety of reasons including war, migration, poverty, and societal expectations. According to a WIDA report (2015), a vast majority of the students with limited or interrupted formal education (SLIFE) in the United States are enrolled in grades 6 through 12.

Based on the results of the second research question, several conclusions can be drawn about adolescent writers' academic literacy support in a secondary context. As the results of the study show no significant differences in the rhetorical features between all groups and the challenge of developing argumentation of literary analysis among all students, it is important that all adolescent writers, regardless of their language background, need support on moving beyond "knowledge telling" to an advanced level of "knowledge transformation" (Bereiter & Scardamalia, 1987). In addition, the results suggest that sentence boundary issues, lack of syntactic variety, and underuse of sophisticated subordination that reflect syntactic features of academic writing are common challenges among all students, including students whose L1 is English. Thus, all students may benefit from instructional support geared towards building syntactic awareness, knowledge, and variety and developing an in-depth understanding of the syntactic features of academic writing. In relation to differences among L1 groups, the findings of the study suggest that some adolescent multilingual writers whose first language is not English perform at the same level or even above the students whose first language is English, while other multilingual students underperform and are particularly challenged by the language demands of academic writing. Thus, providing additional language support and resources to those students who are still in process of developing proficiency in English is necessary to level the playing field for these students. This is important as youths who cannot effectively convey thoughts and

ideas through writing are likely to receive lower grades in their classes, diminishing their prospects for college and career readiness (Graham & Perin, 2007).

Modeling the adolescent writers' text-based analytical essays

The goal of the third research question was to examine the relations among lexical, syntactic, and rhetorical features and writing quality. Upon testing alternative models using CFA, I selected the best-fitting model that fits the data well and that best represents the dimensions of text-based analytical writing. Using the final model selected, I conducted structural regression to examine the relations among the latent constructs of rhetorical, syntactic, and lexical features, and writing quality. The findings based on CFA suggest that the three-factor model, in which lexical, syntactic, and rhetorical features are distinct yet related constructs of writing, best represents the dimensions of text-based analytical writing. The results of structural regression analysis indicate that the latent construct of rhetorical features predicts writing quality while the latent constructs of lexical and syntactic features are not significant predictors of writing quality in the structural regression model. Although the latent constructs of lexical and syntactic features do not significantly predict writing quality in the model, they are strongly correlated with rhetorical features, indicating shared variance.

These results contribute to expanding previous research on the dimensions of writing and their relations to writing quality. The finding about the relation between the latent construct of rhetorical features and writing quality is consistent with the results of an earlier study by Steiss et al. (2021) that showed the latent constructs of evidence use and ideas/structure were independently related to the holistic writing quality score. However, the findings about the relations between the latent constructs of syntactic and lexical features and writing quality show some inconsistencies with the results of previous studies (Kim & Crossley, 2018; MacArthur et

al., 2019; and Steiss et al., 2021). Studies by Kim and Crossley (2018) and MacArthur et al. (2019) found that latent constructs of syntactic complexity and lexical sophistication were strongly correlated with writing quality score. Steiss et al., (2021) also found that the latent construct of language use was independently related to holistic writing quality score.

Several explanations can be offered regarding the inconsistent results. First, the models tested by Kim and Crossley (2018) and MacArthur et al., (2019) do not include rhetorical features such as source use and balance of summary, evidence commentary (which were included in the model tested and used in this study) but include text cohesion as a distinct dimension alongside the latent variables of syntactic complexity and lexical sophistication. In addition, the variability in the use of measures may have contributed to the inconsistent results. Important considerations pertaining to the variability in measures include selection of indicators that constitute the latent constructs and the types of measures used for the analysis. In Kim and Crossley's model, syntactic complexity is operationalized by the mean length of clause while lexical complexity was operationalized by the lexical decision reaction times. In MacArthur et al.'s model, sentence similarity across paragraphs, left embeddedness (words before main verb), and sentence length are the manifest variables that constitute the latent construct of syntactic complexity. The variables used in both of these studies are quantitative measures obtained by NLP tools. Steiss et al.'s model, on the other hand, includes rhetorical features alongside the language features. However, their latent construct of language use is derived from human-judged analytic scoring only and is a much broader construct that includes fluency, syntax, diction, conventions, and tone.

In measuring syntactic complexity and lexical sophistication, I used a combination of methods including analytic scoring, manual coding, and NLP tools. The selection of variables

used in the modeling part of the study was based on the findings of previous studies, theory, and data features. I made the choice to include measures based on human-judged analytic scoring and manual coding considering the limitations of the current automated tools that provide only quantitative measures based on incidences, occurrences, and frequencies. This unique and novel feature of this study may be a reason for the inconsistent results. In addition to the variability in measurement, the model tested in this study includes rhetorical features that haven't been examined in conjunction with lexical sophistication and syntactic complexity features. To my knowledge, no study has examined the source use and the balance of summary, evidence, and commentary—two important rhetorical features—along with linguistic features.

As the results show a strong correlation between the linguistic features and rhetorical features, I hypothesize that the syntactic and lexical features may directly relate to rhetorical features and indirectly relate to writing quality via the rhetorical features. It is also important to consider the criteria specified in the scoring guide for the holistic writing quality score. The rubric used for the holistic scoring of the essays weighs the rhetorical features more than the linguistic features (the scoring guide is provided in Appendix d). Considering the many factors that contribute to writing quality and the complex relations among these factors, there is a need for more studies that examine linguistic and rhetorical features in conjunction. In this vein, Bazerman et al. (2017) reminds us that writing is a complex, multidimensional act that "involves multiple processes, each of which draws on different domains of development, which may interact among themselves" (p. 354).

Limitations of the study

The current study is subject to several important limitations that need to be addressed, and the results should be interpreted with these limitations in mind. The text-based analytical

essays, the unit of the analysis in the current study, are written in an on-demand writing situation in which two class periods were devoted to reading and analyzing the source text (during the first period) and writing an analytical essay (in the following period). Using texts produced in timed writing conditions is a common trend in research that examines linguistic and rhetorical features of student writing. Timed writing assessment is also a large part of the education system and is often the basis for measuring writing growth and performance. However, timed writing should be considered an important limitation as the time constraint does not allow an opportunity for revision. It is important to recognize that sufficient time for writing and multiple opportunities for revision are paramount when engaging in complex rhetorical problem solving demanded by the text-based analytical writing and academic writing, in general. It is also important to acknowledge that many of these students have the potential to produce high-quality academic writing given ample time, opportunities for revision, and access to resources. Thus, the descriptive picture provided in this study should not be interpreted as an overall profile of students' writing. But rather, the results should be interpreted within the scope of on-demand writing of a specific genre, which is text-based analytical writing.

The second limitation of the study pertains to the lack of availability of data on measures of English language proficiency, classification, and other demographic variables that provide valuable information about students' language and educational experiences. The data used in the study come from multiple school districts in several different states. Although this is a strength as the data include a more representative sample of diverse learners attending public schools in different states, not having consistent measures of students' English language proficiency and classification is a limitation. As assessment procedures and policies for students' language proficiency differ from one jurisdiction and state to another within the United States, there is a

lack of uniformity in the definition and the classification of students' proficiency in English. Due to this limitation, I was not able to control for students' English language proficiency in my analyses. In addition, the demographic variables in the study are limited to students' home and first language, gender, class, grade level, and schools they are enrolled in. Multilingual students in the U.S. secondary education context are a population with unique educational and lived experiences and differing exposure to language and literacy development. Thus, I acknowledge that demographic variables pertaining to their educational experiences and language exposure and proficiency are valuable as they help us better understand the unique literacy backgrounds of this student population.

Lastly, grouping based on the language family of the L1s and their linguistic distances to English is an important limitation. I grouped the texts into three broader L1 language backgrounds to examine whether the rhetorical, syntactic, and lexical features that were identified differed among students with different home/first languages. Although I attempted to analyze the rhetorical, syntactic, and lexical features of the student texts comparing L1 student groups that represent each specific language, I didn't have sufficient data to meet the assumptions of multivariate analysis of variance. This methodological consideration concerning the need for comparable group sizes for statistical robustness led to a choice of grouping based on the linguistic distances of students' first/home languages to English. While this grouping method is far from ideal, it did provide some insight into both the commonality and the variability in the rhetorical, syntactic, and lexical features of diverse adolescent students' academic writing. This insight adds to our understanding that language, literacy, and writing skills of multilingual adolescent writers are indeed variable due to differences in linguistic, educational, and social experiences.

Conclusion

The present study systematically analyzed the rhetorical and linguistic features of culturally diverse adolescent students' academic writing and examined how these features vary among L1 groups and the extent to which these features relate to one another and the writing quality. By doing so, the study addresses several critical gaps in L2 writing research. Most of the extant studies that examine student texts have focused on linguistic features alone without attending to rhetorical aspects of the text. This is an important gap given the complexity and multidimensionality of writing whose quality is dependent on both macro-level rhetorical elements and micro-level linguistic features. Second, the majority of the studies that examine the linguistic features of student writing have analyzed independent writing tasks, as opposed to text-based writing tasks. This shortcoming in the writing literature is critical as genre theorists propose that the differences in writing tasks, genres, and communicative purposes lead to variability in rhetorical and linguistic features of texts (Bouwer, 2015; Swales, 1990). Further, much of the research exploring the linguistic features of students' academic writing have revolved around student writing at the collegiate and pre-collegiate levels in a postsecondary context. Academic texts produced by adolescent multilingual students in a secondary education setting have been given little attention in the analysis of learner corpora. In the field of L2 writing, in general, the adolescent multilingual students in a secondary education setting have been an underexplored population (Fitzgerald, 2017; Ortmeier-Hooper & Enright, 2011; Yi et al., 2018).

Given the paucity of research that focuses on both rhetorical and linguistic features of adolescent multilingual students' academic writing, this study makes several important contributions to the extant literature. First, by systematically examining the rhetorical and

linguistic features of text-based analytical writing, which is a formal, school-based, and academic writing genre that requires source use and integration to advance a claim, we better understand the students' strengths, challenges, and needs when engaging in academic writing. This understanding can inform writing pedagogy and intervention geared towards supporting students' academic literacy skills in the secondary context. Second, examining the differences in rhetorical and linguistic features among students with different linguistic backgrounds helps us understand the variability in language and literacy skills of diverse multilingual students. Finally, modeling the relations among the rhetorical elements, linguistic characteristics, and writing quality contributes to our understanding of the multidimensionality of writing and the complex relations and interactions among macro- and micro-level dimensions of writing.

As developing proficiency in academic writing and language is a requisite skill for college and career success, instructional support that meets the diverse needs of culturally and linguistically diverse students in secondary school is necessary. A rather consequential question, then, is how to best support diverse adolescent students in their effort to become proficient academic writers, critical thinkers, and strategic language users. To meet the goal of better supporting academic language and writing development of linguistically diverse students in a secondary context, it is important to understand their challenges and needs. To this end, this study is a small step towards this bigger goal as it sheds light on what strengths adolescent writers have, where they fall short, what challenges they face, how they differ among each other, and what contributes to their writing quality.

References

Acock, A. C. (2008). *A gentle introduction to Stata* (2nd ed.). Stata Press.

ACT, Inc. (2007). *Writing framework for the 2011 national assessment of educational progress*. National Assessment Governing Board.

Bang, H. J. (2013). Reliability of national writing project's analytic writing continuum assessment system. *Journal of Writing Assessment*, 6, 13–24.

Bazerman. C., Applebee, A. N., Berninger, V. W., Brandt, D., Graham, S., Matsuda, P. K., Murphy, S., Rowe, D. W., Schleppegrell, M. (2017). Taking the long view on writing development. *Research in the Teaching of English,* 51, 351–360.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates.

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

Bouwer, R., Béguin, A., Sanders. T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing,* 32, 83–100.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Routledge.

California Department of Education. (2008a). *English language arts teacher guide for California high school exit examination*. Sacramento, CA: Author.

California Department of Education. (2008b). *Teacher guide for the 2008 California writing standards test in grade 7*. Sacramento, CA: Author.

Chen, V., Olson, C. B., & Chung, H. Q. (2020). Understanding proficiency: Analyzing the characteristics of secondary students' on-Demand analytical essay writing. *Journal of Writing Assessment, 13.*

Cobb, T. (2018). *Compleat lexical tutor.* Retrieved from http://www.lextutor.ca.

Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second-language writing.* Cambridge University Press.

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

Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research,* 11, 415–443.

Crossley, S. A. (2018). How many words needed? Using natural language processing tools in educational data mining. *Proceedings of the 10th International Conference on Educational Data Mining (EDM).* 630–633.

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System,* 41, 965–981.

Crossley, S. A., & Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing,* 38, 46–50.

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

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The tool for the automated analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods,* 51, 14–27.

Crossley, S. A., Kyle, K., Varner, L., Gou, L., & McNamara, D. S. (2014). Linguistic microfeatures to predict L2 writing proficiency: A case study in automated writing evaluation. *Journal of Writing Assessment,* 7, 1–23.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35, 115–135.

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing,* 26, 66–79.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29, 243–263.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics,* 36, 570–590.

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, K., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL* (Monograph Series, No. 30). Educational Testing Service.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16–29.

Deng, Y., Lei, L., Liu, D. (2021). Calling for more consistency, refinement, and critical consideration in the use of syntactic complexity measures for writing. *Applied Linguistics,* 42, 1021–1028.

Doolan, S. M. (2021). An exploratory analysis of source integration in postsecondary L1 and

L2 source-based writing. *English for Specific Purposes,* 62, 128–141.

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

Ferris, D. R. (2014). *Language power: Tutorials for writers.* Bedford/St. Martins.

Fitzgerald, J. (2017). How practice-based research informs adolescent English-language learners' composing and compositions. In K. A. Hinchman & D. A. Appleman (Eds.). *Adolescent literacies: A handbook of practice-based research.* (pp. 357–378). Guilford Press.

Fletcher, J. (2015). *Teaching arguments: Rhetorical comprehension, critique, and response.* Stenhouse Publishers.

Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multidimensional analysis. *Journal of Second Language Writing,* 26, 80–95.

Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. *Spaan Fellow Working Papers in Second or Foreign Language Assessment,* 7, 47–84.

Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly,* 10, 9–27.

Grabe, W., & Kaplan, R.B. (1996). *Theory and practice of writing: An applied linguistic perspective* (1st ed.). Routledge.

Graham, S. & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools — A report to Carnegie Corporation of New York.* Alliance for Excellent Education.

Grant, L. & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9, 123–145.

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18, 218–238.

Halliday, M. A. K. (1994). *Introduction to functional grammar* (2nd ed.). Edward Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar*, (4th ed.). Routledge.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* Longman.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features.* Lawrence Erlbaum Associates.

Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, 25, 4–22.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1, 1–26.

Kim, M., & Crossley, S. A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing*, 37, 39–56.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine-grained indices of syntactic complexity and usage-based indices of syntactic sophistication. Dissertation, Georgia State University. Retrieved from https://scholarworks.gsu.edu/alesl_diss/35.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings and application. *TESOL Quarterly*, 49, 757–786.

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing,* 34, 12–24.

Kyle, K., & Crossley, S. A. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing,* 34, 513–535.

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *Modern Language Journal,* 102, 333–349.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behav Res,* 44, 978–990.

Laufer, N., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics,* 16, 307–322.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15, 474–496.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45, 36–62.

Maamuujav, U., Krishnan, J., and Collins, P. (2019). The utility of infographics in L2 writing classes: A practical strategy to scaffold writing development. *TESOL Journal,* 11, 484.

Maamuujav, U., Olson, C. B., & Chung, H. (2021). Syntactic and lexical features of adolescent L2 students' academic writing. *Journal of Second Language Writing*, 53, 100822.

MacArthur. C. A., Jennings, A., & Phillippakos. Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing*, 32, 1553–1574.

Mann, W.C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward functional theory of text organization. *Text,* 8, 243–281.

McNamara, D. S., Graesser, A. C. McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

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

Morris, L., & Cobb, T (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System*, 32, 75–78.

National Writing Project (2010). *Analytical Writing Continuum*. Available at https://archive.nwp.org/cs/public/download/nwp_file/21466/NWP_Analytic_Writing_Continuum.pdf?x-r=pcfile_d.

Nazzal, J. S., Olson, C. B., & Chung, H. Q. (2020). Differences in academic writing across four levels of community college composition courses. *Teaching English in the Two-Year College,* 47, 263–296.

Olson, C. B., Matuchniak, T., Chung, H. Q., Stumpf, R., & Farkas, G. (2017). Reducing achievement gaps in academic writing for Latinos and English learners in Grades 7–12. *Journal of Educational Psychology,* 109, 1–21.

Olson, C. B., Scarcella, R., & Matuchniak, T. (2015). *Helping English learners to write: Meeting common core standard, grades 6-12*. Teachers College Press.

Olson, C. B., Woodworth, K., Arshan, N., Black, R., Chung, H. Q., D'Aoust, C., Dewar, T., Friedrich, L., Godfrey, L., Land, R., Matuchniak, T., Scarcella, R., & Stowell, L. (2020). The pathway to academic success: Scaling up a text-based analytical writing intervention

for Latinos and English learners in secondary school. *Journal of Educational Psychology,* 112, 701–717.

Olson, C. B., Maamuujav, U., Steiss, J., & Chung, H. (forthcoming). Examining the impact of a cognitive strategies approach to teaching arguments of literary analysis on the writing of mainstreamed English learners in secondary school. *Written Communication.*

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics,* 24, 492–518.

Ortmeier-Hooper, C., & Enright, K. A. (2011). Mapping new territory: Toward an understanding of adolescent L2 writers and writing in US contexts. *Journal of Second Language Writing,* 20, 167–181.

Perin, D., & Lauterbach, M. (2018). Assessing text-based writing of low-skilled college students. *International Journal of Artificial Intelligence in Education,* 28, 56–78

Plakans. L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing,* 26, 561–587.

Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22, 217–230.

Salsbury, T., Crossley, S. A, & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research,* 27, 343–360.

Sanders, T.J.M., & Schilperoord, J. (2006). Text structure as a window on the cognition of writing: How text analysis provides insights into writing products and writing processes. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 386–402). Guilford Press.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective.* Lawrence Erlbaum Associates, Inc.

Shaughnessy, M. P. (1977). *Errors and Expectations: A guide for the teacher of basic writing*. Oxford University Press.

Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, 21, 171–200.

Steiss, J., Krishnan, J., Kim, Y.S., Olson, C. B. (2022). Dimensions of text-based analytical writing of secondary students. *Assessing Writing*, 51, 100600.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

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

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2012). *The nation's report card: Writing 2011* (NCES 2012-470). Washington, DC.

Watanabe, Y. (2001). Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions. Unpublished doctoral dissertation. University of Hawaii, Manoa.

WIDA Consortium. (2015). Focus on SLIFE: Students with Limited or Interrupted Formal Education. Wisconsin Center for Education Research (WCER). University of Wisconsin-Madison, www.wida.us.

Witte, S., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, 32, 189–204.

Wolfe-Quintero, K. Inagaki, S., & Kim, H-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity* (Technical Report No. 17). Second Language Teaching & Curriculum Center, University of Hawaii.

Yi, Y., & Kao, C., & Kang, J. (2018). Middle and high school writing. In J. I. Liontas. (Ed.). *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons, Inc.

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, 47, 100505.

Zhao, R., & Hirvela, A. (2015). Undergraduate ESL students' engagement in academic reading and writing in learning to write a synthesis paper. *Reading in a Foreign Language*, 27, 219–247.

CHAPTER 6

Summary, Implications, and Contributions

The three studies of this dissertation focus on the academic writing of linguistically diverse adolescent students in a U.S. secondary education context. As part of the academic writing genre, the text-based analytical essays (N=410) written by middle and high school students were systematically analyzed with a goal to identity language use and linguistic patterns in students' academic writing and to further examine how linguistic features relate to rhetorical features and writing quality. The key findings of the studies are summarized in Table 6.1 below.

Table 6.1

Summary of Key Findings

Studies	Key findings
Study 1: Examining Lexical Features and Academic Vocabulary Use in Adolescent Multilingual Students' Text-based Analytical Essays	▪ Adolescent multilingual students' text-based analytical writing exhibited low lexical density and diversity. ▪ Higher percentage of frequently used words and lower percentage of academic words indicate reliance on words common in colloquial discourse. ▪ Lexical density and lexical diversity did not predict human-judged holistic writing quality in text-based analytical writing. ▪ Lexical sophistication was significantly correlated with writing quality score. ▪ 44% of the AWL words in student texts were from the source and the prompt.
Study 2: Syntactic and Lexical Features of Adolescent Multilingual Students' Academic Writing	▪ Sentences with boundary issues and underuse of sophisticated subordination were common syntactic features in adolescent students' writing. ▪ Low use of academic vocabulary and dependence of frequent words were observed in students' academic writing. ▪ Complex and conventional sentence, age of acquisition, word concreteness, imageability, and meaningfulness were correlated with writing quality. ▪ Sentence complexity, age of acquisition for word, and word imageability accounted for 41% of the variance in the holistic writing quality score.
Study 3: An Investigation of Linguistic and Rhetorical Features in the Academic Writing of Culturally Diverse Adolescent Students	▪ Unconventional and simple sentences were common syntactic patterns while varied and nonfinite clausal structures were underused. ▪ Low use of academic words was observed in the learner corpus. ▪ Summary structure, as well as general summary/paraphrasing as common source integration, indicate reliance on summary and retelling. ▪ Texts in L1 groups did not significantly differ in rhetorical features, but differences were found in linguistic features with variability among groups. ▪ Lexical, syntactic, and rhetorical features were distinct but related dimensions of writing as the three-factor model best fitted the data. ▪ Rhetorical features significantly predicted writing score while lexical and syntactic features were strongly correlated with rhetorical features.

It is important to highlight some results that were unanticipated and at variance with some of the findings of the earlier studies. First, the results of the third study of the dissertation indicated a non-significant difference in both rhetorical and linguistic features of texts between students with English L1 and multilingual students with Spanish L1. In fact, academic texts of multilingual students with Spanish L1 in the sample of the third study, on average, showed linguistic performance that is similar to, or in some features even higher than, the writing of students with English L1. Earlier studies often reported low writing and language performance and "achievement gaps" of Hispanic students (Kim et al., 2011; Olson et al., 2017, 2020) in a secondary education context. However, this does not seem to be the case in the sample of text-based analytical writing that were analyzed in this dissertation study. An important difference is that the earlier studies used data from the state of California only, while the data used in the third study of the dissertation come from Arizona, Minnesota, Texas, Utah, and Wisconsin. It is possible that many of the Spanish-speaking multilingual students attending public school in these states, unlike those in California, may have developed proficiency in English.

However, texts in the third language group with multilingual students representing other languages, including Somali, Hmong, Arabic, Amharic, Burmese and so on, demonstrated significantly lower performance on linguistic features, compared to the texts produced by their English- and Spanish-speaking peers. It is plausible that these students in the sample of the study are multilingual writers who are still in process of developing their proficiency in English. The data pertaining to students' proficiency in and exposure to English were not available; thus, these explanations are speculative. The results, however, indicate variability in the linguistic needs of a multilingual student population. To better understand this variability among multilingual

students' academic literacy skills and needs, future research can further analyze the factors that contribute to the differential linguistic performances among diverse multilingual students.

Another unique finding pertains to the structural regression model in which the latent constructs of lexical and syntactic features did not predict the human-judged holistic writing score. This was not consistent with the results of the previous studies that found that the latent constructs of lexical sophistication and syntactic complexity predicted holistic writing quality in structural regression models (Kim & Crossley, 2018; MacArthur et al., 2019). It is, however, important to note that the models tested by Kim and Crossley (2018) and MacArthur et al. (2019) do not include macro-level rhetorical features with the exception of textual features of cohesion which was measured by quantitative measures of lexical overlaps and connectives obtained from NLP tools. It is also important to highlight that the lexical and syntactic features are strongly correlated with rhetorical features in the structural regression model tested in this dissertation study. Thus, it can be hypothesized that lexical and syntactic features contribute to rhetorical features, which then predicts writing quality. In other words, lexicon and syntax as micro-level features are crucial in the expression and forming of ideas at macro rhetorical level, and the quality of the ideas formed is what ultimately predicts the writing score. In future research, this hypothesis can be tested with a bifactor model or mediation analysis.

The unique results of the dissertation research attest to the complexity of writing dimensions and development. Nonetheless, the overall findings of the three studies have important implications for writing pedagogy that is geared toward developing adolescent students' academic literacy and for writing research that deepens our understanding of the complex dimensions of writing. In what follows, I discuss these implications and highlight the contributions of the dissertation studies to the field of writing research and pedagogy.

Pedagogical Implications

Curriculum and instruction for literacy development in a secondary education context have placed considerable emphasis on students' academic writing and literacy development. In the United States, for example, the importance of academic literacy development is reflected in the Common Core State Standards (CCSS, 2010) and the Next Generation Science Standards (NGSS, 2013) as the goals of these standards are to prepare students for college and career readiness and to help them succeed in the global economy. These standards emphasize the need for students to develop skills for effective writing and communication, inquiry and problem solving, and analytical and critical thinking. While the standards have placed a premium on academic writing and set a high bar for students, how these ambitious goals may be achieved are not often clear. Thus, the responsibility is on the teachers to find ways to meet the differing and complex literacy needs of culturally and linguistically diverse adolescent students in U.S. secondary schools for improving their academic writing. As adolescent students participate in school-based academic writing activities that are often challenging for them, they need support in their effort to: 1) figure out how meaning is constructed and communicated within a particular rhetorical context; 2) expand, use, and negotiate their linguistic resources; and 3) become articulate in the discursive practice of the academic writing community.

Developing the higher-order skills within an academic context requires an in-depth understanding of the specialized lexical, syntactic, and discursive features of academic texts and discourse. A question of pedagogical import, then, is how teachers in a secondary education context can support diverse students with differing literacy needs in their effort to become proficient academic writers well versed in the specialized language conventions of academic discourse. One approach that has been empirically proven effective is a comprehensive,

cognitive strategies instruction that focuses on higher-level interpretive and analytical aspects of writing, alongside the development of language skills necessary to express the higher-level thinking effectively (Kim et al., 2011; Olson & Land, 2007, 2008; Olson et al., 2012, 2017, 2020, forthcoming). A comprehensive, cognitive strategies approach has been particularly beneficial to multilingual writers who are still in process of developing their proficiency in English. A recent study by Olson et al. (forthcoming) found that the cognitive strategies approach improved all components of writing, including macro-level rhetorical features and micro-level linguistic features, among multilingual students with different levels of English proficiency. The study suggests that all learners, regardless of their language status, can improve their writing when given an opportunity to engage in high-quality, comprehensive instruction that focuses on higher-level interpretive and argumentative aspects of writing. Similarly, Walqui and Bunch (2019) argue for the amplification of the curriculum, rather than reduction and simplification, and emphasize the importance of enacting "stimulating, demanding, well-supported lessons to transform what is currently offered to many English learners" (p. 21).

While many teachers who work with adolescent multilingual writers with different linguistic backgrounds realize the necessity of helping these students develop academic language proficiency, they face considerable challenges in meeting the academic writing, language, and literacy needs of diverse adolescent multilingual students. They may not always be certain about how to integrate strategy and language instruction into their already busy and packed writing curriculum and what particular language and writing features to focus on. For example, in the survey I collected from 23 teachers in 5 states who participated in the third study of the dissertation, teachers reported several challenges including not having enough time to focus on writing strategies and language use, not having adequate training to teach writing and language

to multilingual students, and not knowing how to scaffold and what to teach when it comes to academic writing and language development. The following vignettes are provided to illustrate these points.

One teacher (8th grade, ELA) said,

> *The most challenging aspect of teaching language is how to scaffold it. Some of the vocabulary/content is intimidating to me because of my lack of training/instruction in it. I feel like I know how to write well, but I'm not sure how to teach students how to do it.*

Another teacher reported (12th grade ELA) reported,

> *I don't have training on how to help them, regarding English acquisition issues. They commonly struggle with vocabulary and prepositions/smaller pieces of grammar and sentence fluency.*

These concerns pertaining to the lack of training teachers had to address the language development of multilingual students is supported by the results of the survey in which the majority of the teachers (74% or 17out of 23) reported that they had minimal and no pre-service and in-service training to teach writing and language to multilingual students who are still developing their English proficiency.

Given the importance of students' academic literacy development for college and career success and the challenges secondary school teachers encounter in supporting students in that endeavor, it is important to understand the linguistic needs and academic literacy challenges of students with differing language backgrounds and educational experiences. By systematically analyzing students' academic writing using data representing linguistically diverse students in multiple states, the three studies of the dissertation collectively inform writing pedagogy by shedding light on the academic language and writing needs of linguistically diverse adolescent

students in a secondary education context. Based on the findings of the dissertation studies, several recommendations can be made in addressing the academic literacy needs of students in secondary school. The academic literacy needs include both macro-level rhetorical features that require complex problem-solving skills and micro-level linguistic features that are necessary to express and convey complex ideas.

The results based on the systematic analysis of texts indicate that students have considerable challenges with higher order tasks involved in writing arguments of literary analysis. Developing arguments of literary analysis involves analyzing a source text, constructing an argument of interpretive analysis, and using a source text for evidence. To successfully complete this task, students need to produce a text that is marked by well informed and focused argument, evidence-based development, and textual coherence based on logical order. To address the student challenges and to develop their analytical writing skills, the following recommendations are proposed:

- Provide students multiple opportunities to engage in analytical writing and arguments of literary analysis to develop their rhetorical problem-solving skills
- Teach students various types of argument structures using models, anchor texts, and visuals such as graphic organizers and infographics
- Demonstrate to students how to connect ideas using a variety of cohesive devices (e.g., linking words, synonyms, referring backwards and looking forward) and teach strategies to improve cohesion
- Provide guidance on attributing sources and using source material for textual evidence to support an argument

- Show students various ways to integrate sources, including introduced and embedded quotations
- Regularly engage students in critical discussion, interpretive analysis, commentary writing based on problems, theme, and other literary elements presented in the text.

These suggestions are presented as a guide to move students beyond mere retelling and to engage them in knowledge transformation which involves complex rhetorical problem solving and analytical skills.

The findings of the dissertation studies also point to the need for adolescent writers to develop their linguistic resources and repertoires for academic writing. To be able to produce convincing and coherent academic texts, students need to be conversant with the language conventions and the formal code of academic written discourse. Students need to know not just what linguistic choices they can make, but why they make these choices and when these choices are fitting. To address the linguistic needs of adolescent writers for academic writing, teachers can provide instruction and resources on the following linguistic features:

- Sentence boundaries and avoiding sentence boundary errors, such as fragments, comma splices, and run-on sentences when writing within an academic context
- Types of sentence structures (simple, compound, complex), and purpose and functions of coordination and subordination
- Syntactic variety and ways of varying sentence styles and structure
- Participial phrase, absolute phrase, and apposition (nonfinite clause structures)
- Noun phrase and nominal structures for phrasal complexity and variety
- Academic vocabulary and academic expressions
- Lexical variety

This is not an exhaustive list, but instruction that focuses on developing students' knowledge of and about linguistic features and their functions can help students expand their linguistic resources and, thus, improve their writing quality. This argument is supported by evidence-based recommendations by the What Works Clearinghouse (Graham et al., 2016) and the meta-analysis of research-based approaches that found teaching vocabulary and sentence construction skills improved writing quality (Graham et al., 2015).

As the findings of the dissertation study showed variability in the linguistic performance among diverse students, it is important to note that some multilingual students need additional language support, especially in the area of vocabulary development and syntactic complexity at a phrasal level. The results of the study also indicate that all students, regardless of their language background, can benefit from instruction that focuses on the language features of academic written discourse. For example, the results suggest that unconventional sentences and lack of syntactic variety are common among all students. Thus, instruction and resources on avoiding sentence boundary issues and ways of varying sentences and styles can be integrated into writing pedagogy and made accessible to all learners.

In order to expand adolescent writers linguistic resources and repertoires for academic writing, teachers may need a multifaceted pedagogical approach that provides: 1) exposure to rich materials and activities that represent and align with the content and conventions of academic writing; 2) explicit instruction that draws learners' attention to syntactic structures and lexical use; 3) strategy instruction that shows how language is used to construct meaning; and 4) guided practice that promotes the application of their acquired linguistic knowledge in composing, revising, and editing processes (Maamuujav & Olson, 2019). This way, teachers can help students determine how writers create meaning from and with texts and how linguistic

choices are shaped by socially constructed genre-conventions. Given the complexity of argument writing and the shifting instructional focus toward argument literacy, it is important for teachers to find effective ways to support diverse students in their effort to develop critical thinking and analytical writing skills and to become well-versed and strategic language users.

Research Implications

Previous writing studies have provided insights into the linguistic features of students' writing and the relationship between linguistic features and writing quality; however, most of the extant studies that examine student texts have focused on independent writing tasks, as opposed to source- or text-based writing. This is an important shortcoming in the writing literature as genre theorists suggest that differences in writing tasks, genres, and the communicative purposes lead to variability in rhetorical and linguistic features of texts (Bouwer, 2015; Swales, 1990). In addition, recent L2 studies have found that linguistic features and cohesive properties of texts differ by writing task types (e.g., source-based vs. independent) and genres (Berman & Verhoeven, 2002; Guo et el., 2013; Kim & Crossley, 2018; Kyle & Crossley, 2016; Yoon & Polio, 2017). Importantly, sociocultural theory of writing informs us that writing is a complex, socially situated activity that is influenced by social, cultural, and historical forces and that writers within a specific community are bound to adhere to the established norms and the expectations of the community. For these reasons, examining student texts that are produced in different rhetorical situations and that represent different genres and discursive practices is crucial for a more comprehensive understanding of the complexity of writing and the features of written texts. Text-based analytical writing is a school-based genre that is crucial for academic success. Thus, analyzing texts that represent this important yet lesser-studied writing genre and

task will add to our knowledge about linguistic and textual dimensions of students' academic writing.

Collectively, the studies of this dissertation will inform us of both linguistic needs, as well as strengths, of adolescent multilingual students who are often marginalized in their classrooms and schools. Kanno and Cromley (2015) assert that many developing multilingual students in secondary school are restricted in their access to rigorous curricula and become "long-term" English learners due to instructional and institutional barriers. For many of these students across different school districts and states, school is the only place where they are exposed to academic language. However, they are often placed in sheltered and remedial courses with watered-down academic content while rigorous instruction that prioritizes complex interpretative and rhetorical problem-solving skills is often reserved for honors students or those deemed "English proficient" (Applebee et al., 2003; Kanno, 2022; Short & Fitzsimmons, 2007). Schleppegrell's (2004) remark is noteworthy here; she asserts, "In the absence of an explicit focus on language, students from certain social class backgrounds continue to be privileged and others to be disadvantaged in learning, assessment, and promotion, perpetuating the obvious inequalities that exist today" (p. 3). Meeting the linguistic needs of developing multilingual writers while engaging them in rigorous, meaningful, and socially contextualized literacy instruction and experiences are a way to promote equal access and level the playing field for this student population. The studies of the dissertation can help us understand the linguistic needs of this student population and shed light on how to best support them for developing their academic writing skills.

In the context of standards-based secondary education that places a premium on academic writing and literacy development, understanding how students use language for academic

communication and what their linguistic needs are in writing within an academic context is crucial. In this regard, Bunch et al. (2012) emphasized an important role language and literacy researchers play in "helping to articulate fundamental understandings about the development and use of language and literacy relevant to the challenges facing ELs in light of the new Standards" (p.3). Such understanding is vital in writing research that addresses the academic writing and language development of diverse multilingual students navigating the complex landscape of U.S. secondary education. Such understanding is also crucial in writing policy that are geared towards preparing adolescent multilingual students for college and career success.

Besides the implications pertaining to diverse multilingual students' academic literacy development, the results of the dissertation studies also have important implications in writing research that examines the dimensions of writing, the features that contribute to writing quality, and the complex relations between these features and dimensions. Bazerman et al. (2017) reminds us that writing is a complex, multidimensional act that "involves multiple processes, each of which draws on different domains of development, which may interact among themselves" (p. 354). By examining how lexical, syntactic, and rhetorical features relate to one another and to overall writing quality, the dissertation studies add to our knowledge of the complex and multidimensional system of writing.

More specifically, the methodological approaches and the novel findings of this dissertation research can inform further research in two important areas. The first area pertains to operationalizing the linguistic features of student texts. The use of manual coding and analytic scoring, in addition to NLP tools, for measures of lexical sophistication and syntactic complexity add more nuances to the way these constructs are operationalized. Thus, the manual coding and analytic scoring frameworks developed for the dissertation studies can inform future research

that examines linguistic features and their relations to writing quality. Further, given the novel finding pertaining to the relations of linguistic features to both rhetorical features and writing quality, this dissertation research can propel more research to examine complex relations among linguistic characteristics, macro-level rhetorical and discourse features, and writing quality. Given the paucity of research and complexity of writing dimensions, more research is needed to further explore the relation between linguistic and rhetorical features and how they independently and together contribute to writing quality.

Contributions to the Field

The dissertation studies as a whole make important contributions to the field of writing research and pedagogy by examining the lesser studied writing of an underexplored student population in a secondary education context. To my knowledge, no study has explored lexical and syntactic features in text-based analytical writing, yet this type of writing is an increasingly important genre for academic success. Crossley (2020) contends that "the majority of information available to the field about the interactions between writing quality/development and linguistic features is derived from a single task (independent writing), shedding some confidence on the generalizability of the findings to other tasks" (p. 432). Thus, finding out whether and how texts that represent different writing tasks and genres exhibit commonality and variability in lexical features is crucial for building a more comprehensive and complete understanding of the relation between linguistic features and writing quality.

In addition, academic texts of adolescent multilingual students in a secondary education context have been an underexplored area in L2 writing research. In fact, adolescent multilingual students in the secondary education setting, despite their growing rate, have remained "outside the purview" in L2 research (Ortmeier-Hooper & Enright, 2011, p. 167). Adolescent multilingual

students in this context are often faced with cultural, cognitive, linguistic, communicative, contextual, and affective constraints while grappling with the demands of academic writing and learning an additional language (Olson et al., 2015). Thus, these students constitute a unique population with differing exposure to and experiences of navigating the complex landscape of US secondary education. To my knowledge, there is no other study that focus on both rhetorical and linguistic features using a sample of linguistically diverse students in different geographical regions of the United States.

The dissertation research has the potential to shed light on how to best support adolescent multilingual students in their effort to become proficient academic writers well versed in the specialized language convention of academic discourse. As developing proficiency in academic writing and language is a requisite skill for college and career success, institutional and instructional support are necessary to help diverse adolescent students to develop their language resources and expand their linguistic repertoires. To become successful academic writers, adolescent multilingual students need to develop their linguistic competence, syntactic fluency, and lexical capacity; they need knowledge of language, text, and discourse to be able to write confidently and competently, to be able to compose analytically and persuasively, to meet the standards set forth by CCSS, and to be ready for college and career success. To accomplish this goal, a needs-based approach, based on a systematic examination of students' language use and analytical skill in their academic writing, can help identify the language features and rhetorical components to target in writing instruction. This research is a small step towards this bigger goal as I attempt to systematically examine the linguistic and rhetorical features of diverse adolescent students' academic writing.

References

Applebee, A., Langer, J., Nystrand, M., Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal,* 40, 685–730.

Bazerman. C., Applebee, A. N., Berninger, V. W., Brandt, D., Graham, S., Matsuda, P. K., Murphy, S., Rowe, D. W., Schleppegrell, M. (2017). Taking the long view on writing development. *Research in the Teaching of English,* 51, 351–360.

Berman, R. and Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy,* 5(1), 1–43.

Bouwer, R., Béguin, A., Sanders. T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing,* 32, 83–100.

Bunch, G. C., Kibler, A. K., & Pimentel, S. (2012). *Realizing opportunities for English learners in the common core English language arts and disciplinary literacy standards*. Paper presented at the Understanding Language Conference, Stanford, CA. Retrieved from http://ell.stanford. edu/papers.

Common Core State Standards Initiatives. (2021). *English language arts standards: College and career readiness anchor standards for writing.* Retrieved from http://www.corestandards.org/ELA-Literacy/CCRA/W/.

Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research,* 11(3), 415–443.

Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., . . . Smither Wulsin, C. (2016). *Teaching secondary students to write effectively* (NCEE 2017–4002). National

Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-based writing practices and the common core. *The Elementary School Journal,* 115, 498–522.

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18, 218–238.

Kanno, Y. (2022). Preparing multilingual learners for college and career success [Webinar]. Center for Applied Linguistics. https://www.cal.org/resource-center/research-to-policy-series/multilingual-college-career.

Kanno, Y., & Cromley, J. G. (2015). English language learners' pathways to four-year colleges. *Teachers College Record,* 117, 1–44.

Kim, J., Olson, C. B., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., Collins, P., & Land, R. (2011). A randomized experiment of cognitive strategies approach to text-based analytical writing for mainstreamed Latino English learners in Grades 6 to 12. *Journal of Research on Educational Effectiveness*, 4, 231–263.

Kim, M., & Crossley, S. A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing,* 37, 39–56.

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing,* 34, 12–24.

Maamuujav, U., & Olson, C. B. (2019). Meeting the linguistic needs of adolescent multilingual

writers for academic writing. In J. I. Liontas (Ed.). *The TESOL Encyclopedia of English Language Teaching.* John Wiley & Sons, Inc.

MacArthur. C. A., Jennings, A., & Phillippakos. Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing,* 32, 1553–1574.

Next Generation Science Standards. (2013). Fact Sheet. Retrieved from https://www.nextgenscience.org.

Olson, C. B., & Land, R. (2007). A cognitive strategies approach to reading and writing instruction for English language learners in secondary school. *Research in the Teaching of English,* 41, 269–303.

Olson, C. B., & Land, R. (2008). Taking a reading/writing intervention for secondary English language learners on the road: Lessons learned from the Pathway Project. *Research in the Teaching of English,* 42, 259–269.

Olson, C. B., Kim, J. S., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., & Land, R. (2012). Enhancing the interpretive reading and analytical writing of mainstreamed English Learners in secondary school: Results from a randomized field trial using a cognitive strategies approach. *American Educational Research Journal,* 49, 323–355.

Olson, C. B., Matuchniak, T., Chung, H. Q., Stumpf, R., & Farkas, G. (2017). Reducing achievement gaps in academic writing for Latinos and English learners in Grades 7–12. *Journal of Educational Psychology,* 109, 1–21.

Olson, C. B., Scarcella, R., & Matuchniak, T. (2015). *Helping English learners to write: Meeting common core standard, grades 6-12.* Teachers College Press.

Olson, C. B., Woodworth, K., Arshan, N., Black, R., Chung, H. Q., D'Aoust, C., Dewar, T., Friedrich, L., Godfrey, L., Land, R., Matuchniak, T., Scarcella, R., & Stowell, L. (2020). The pathway to academic success: Scaling up a text-based analytical writing intervention for Latinos and English learners in secondary school. *Journal of Educational Psychology,* 112, 701–717.

Ortmeier-Hooper, C., & Enright, K. A. (2011). Mapping new territory: Toward an understanding of adolescent L2 writers and writing in US contexts. *Journal of Second Language Writing,* 20, 167–181.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective.* Lawrence Erlbaum Associates, Inc.

Short, D., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners–A report to Carnegie Corporation of New York.* Alliance for Excellent Education.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

Walqui, A., & Bunch, G. C. (2019). What is quality learning for English learners? In A. Walqui & G. C. Bunch (Eds.). *Amplifying the curriculum: Designing quality learning opportunities for English learners.* (pp. 21–41). Teachers College Press.

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, 51, 275–301.

CHAPTER 7

Conclusion

Academic writing plays a significant role in access to, persistence of and success in college. Its importance as a job requirement has grown as labor today increasingly involves communication in a written form and transformation and sharing of knowledge with a wide variety of audiences. Many public and private sector employers say that writing effectively in a professional setting is a critical skill that "directly affects hiring and promotion decisions" (Graham & Perin, 2007). Thus, writing well in an academic and professional context has become the cornerstone and an essential skillset required by many professions. As proficiency in academic writing has a direct and immediate impact on one's college and career path, the prospects of finishing school, obtaining a college degree, and getting a job that pays a living wage are diminished for youths who cannot effectively convey thoughts and ideas through writing. These prospects are much dimmer for multilingual students who haven't developed proficiency in English and academic language in a U.S. education context where academic English continues to be a gatekeeper for college admission, hiring decisions, and career promotions.

Despite its importance for college and career success, academic writing is one of the most challenging skills to develop. It is a complex task that requires an in-depth understanding of domain knowledge, genre conventions, rhetorical situations, and linguistic principles. To write well in an academic context, students need to develop nuanced skills to synthesize concrete details and abstract concepts, to present complex ideas in a clear and coherent way, and to make rhetorical and linguistic choices based on genre conventions and audience expectations (Defazio et al., 2010, Maamuujav et al., 2019). These demands of academic writing are particularly

challenging for multilingual students who are still in process of developing their proficiency in the language in which they are writing. Yet, without proficiency in academic English and academic writing, these students cannot have access to college and professional careers. Thus, instructional support in their academic English and academic writing development is an important educational agenda to level the playing field for this student population. As these students participate in school-based academic writing activities to develop their academic language and writing skills, they need support in their effort: 1) to figure out how meaning is constructed and communicated within a particular rhetorical context; 2) to expand, use, and negotiate their linguistic resources; and 3) to become articulate in the discursive practice of the academic writing community.

To effectively support developing multilingual writers' academic language and writing development, it is crucial to understand where they fall short, what challenges they face, and which types of support they need when engaging in a school-based, academic writing tasks. Such understanding can inform educational interventions, curricular approaches, and instructional practices geared toward developing students' academic literacy skills. This dissertation research addresses the academic literacy challenges and needs of linguistically diverse adolescent students, specifically focusing on the multilingual students from different language backgrounds. The studies of the dissertation examine how students use language, what rhetorical choices they make when writing in an academic context, and how the linguistic and rhetorical features contribute to writing quality. By identifying the language and rhetorical features of their written texts, the studies collectively can help determine the language features and components to target in writing instruction. By examining the complex relations among linguistic features, rhetorical elements, and writing quality, the studies contribute to our understanding of what constitutes

effective writing for the text-based analytical writing, which is an important academic writing genre that is aligned with the CCSS's emphasis on developing the ability to analyze, argue, interpret, and use evidence to support ideas about the meaning and message of a literary work.

The focus of the dissertation research is academic language use and academic writing features of multilingual students who are navigating the complex landscape of U.S. secondary education. This is an area that continues to be underexplored although scholars and researchers have been raising concern that adolescent multilingual students in a U.S. secondary education context and their academic language needs have remained outside the purview of second language writing research (Fitzgerald, 2017; Olson et al., 2020; Ortmeier-Hooper & Enright, 2011; Yi et al., 2018). Thus, this dissertation research addresses the gap in L2 writing research. In addition, most of the studies on academic language and literacy development of multilingual students are small-scale with data collected from one classroom or one school (Olson et al., 2020). This dissertation research uses a large data of linguistically diverse students, with L1s representing 27 different languages, in multiple school districts in 6 states. This is an important factor in the generalizability of the findings, as well as in informing effective pedagogical interventions to address the academic literacy needs of adolescent multilingual students across the nation.

The most comprehensive national document that guides writing and literacy pedagogy is the Common Core State Standards (CCSS) for English Language Arts & Literacy in History/Social Studies, and Technical Subjects, which has been adopted by 41 states (Common Core State Standards Initiative, 2022). However, the CCSS explicitly states that defining the range of institutional and instructional supports that are available and appropriate for English learners is "beyond the scope of the Standards" (CCSS, 2013, p.5). Thus, the responsibility is on

the teachers to find ways to meet the different and complex needs of adolescent English language learners for academic writing. With lack of guidance, expertise, resources, and time to develop instructional materials, secondary teachers in English Language Arts, as well as content area classrooms, find it challenging to address the linguistic needs of diverse multilingual students. In fact, the results of my survey given to teachers in 5 states indicate that the lack of English proficiency and differing linguistic needs of multilingual students pose immense challenges to teachers, especially since the majority of the teachers have not had adequate training to teach academic language and writing to multilingual students. The dissertation research addresses this pressing issue facing teachers in secondary school and provides some guidance by identifying students' academic literacy needs.

Critical Perspectives on Academic Language and Literacy Development

Although the focus of this dissertation is consequential as it pertains to addressing the linguistic needs and, thereby, leveling the playing field for developing multilingual writers, it is important to attend to the criticism of and the contention surrounding academic English. Academic English has drawn criticism for its hegemonic status as some scholars have problematized the prioritization of academic English for stigmatizing and suppressing linguistic practices of language-minoritized populations in U.S. society (Flores & Rosa, 2015; Garcia, 2020; Garcia & Solorza, 2020). The raciolinguistic ideologies, the term coined by Flores and Rosa (2015) to refer to the interrelationships between racialized bodies and linguistic practices, extend the critique of academic language, claiming that it legitimizes the social and educational exclusions of many language-minority students. These scholars share the sentiment that language practices of minoritized groups are devalued and underprivileged as a consequence of the prioritization of academic English in a formal schooling context. These critical views toward

academic English are important as they illuminate the ideologies and the power underlying academic English and suggest how the linguistic practices of language-minoritized students might become devalued in a context of formal schooling.

While it is crucial to be mindful of these critical stances in both research and pedagogy, the deeper systemic and structural barriers confronting multilingual students cannot be ignored. For example, one of the most pressing barriers that multilingual students encounter in a largely monolingual U.S. education system, at both secondary and postsecondary levels, is meeting the rigid criteria of assessment of all sorts that have a direct effect on their college and career path. Academic English continues to be a gatekeeper for access to college and professional careers, and high-stakes assessment and tasks require academic English while norm-based monolingual criteria are in place to measure students' academic language and literacy skills. Within such a system, multilingual students who haven't developed proficiency in academic English continue to be marginalized, excluded, and restricted in their access to college and career opportunities. This brings us to the paradoxical nature of linguistic social justice, which perhaps is the root of the contention between those who support academic English instruction and those who reject it.

By promoting academic English, are scholars and classrooms teachers contributing to the social exclusions of language-minority students? On the other hand, would deprioritizing academic English perpetuate the inequalities that permeate the society? These are challenging and paradoxical questions educators and researchers will have to grapple with. However, a point that scholars on the opposite sides of the issue may agree on is that there is a need for a collective effort to meaningfully support multilingual students in their endeavor to expand their linguistic repertoires. Language educators recognize the value of helping students to become language users who are able to make linguistic choices strategically. In a multicultural pluralistic society

in which multilingual students "shuttle between communities and enjoy multiple memberships" (Canagarajah, 2002, p. 35), strategic use of language for effective communication is a desirable skill. Strategic language users are conversant with distinct language conventions and make linguistic choices considering the linguistic practices of the discursive communities they shuttle between. They are aware of their audience and the rhetorical situation and tactically make linguistic choices to negotiate meaning in different communicative contexts. Access to and participation in academic communities and disciplinary discourse entail being proficient in the linguistic practices of these communities. Thus, instead of rejecting academic English, we need a way to support students in their effort to become well-versed in academic English while upholding the rich linguistic practices of their homes and communities. This is an important pedagogical goal that aligns with the principles of linguistic diversity and social justice.

Moving Forward

It is important for educators, researchers, and policy makers to be mindful that academic language can be "key to both success and oppression in schools" (Proctor et al., 2019). Hence, considering the current raciolinguistic ideologies and practices are essential when implementing curriculum, designing syllabi, and selecting methodological approaches. We move toward this direction when instructional practices effectuate effective and equitable learning opportunities that not only attend to the linguistic needs of multilingual students for academic literacy but foster cultural and linguistic pluralism and critical language awareness. Recent scholarship on Culturally Sustaining Systemic Functional Linguistics (CSSFL) is an effort to center the dynamic cultural and linguistic practices of multilingual students while supporting them in "building up their rhetorical, civic and academic repertoires within their new cultural context" (Harman & Burke, 2020, p. 18). As implied from its name, CSSFL draws on Culturally Sustaining Pedagogy

(CSP) advocated by Paris and Alim (2014; 2017) and Systemic Functional Linguistics (SFL) proposed by Halliday (1994) to foster linguistic pluralism, shift towards heteroglossic language practices, and raise critical language awareness while simultaneously supporting students to expand their linguistic repertoires to meet the language demands of academic and disciplinary practices.

Moving forward, a collective effort is necessary to foster a more inclusive language education that addresses both the raciolinguistic ideologies of linguistic practices and the linguistic needs of multilingual students to engage in academic discourse. Attending to the raciolinguistic ideologies associated with linguistic practices and addressing the linguistic needs of multilingual students for academic writing are two important areas in research and pedagogy. Although this dissertation research does not explore the raciolinguistic ideologies, it was conducted with an awareness of and knowledge about the criticism and contention surrounding academic language and literacy of multilingual and multidialectal students. The dissertation research, however, addresses the second area pertaining to the linguistic needs of multilingual students for academic writing. In that, the dissertation makes some contribution to the collective effort as a small step towards the broader goal of supporting diverse multilingual students for their effort to develop academic language and literacy skills. Future research can address both areas by examining how multilingual students draw on their full linguistic resources and repertoires available to them to develop their academic language skills and how teachers can leverage their existing linguistic resources and integrate their dynamic cultural and linguistic practices into classroom.

Finally, I conclude this dissertation by revisiting the Writer(s)-Within-Community (WWC) model (Graham 2018a, 2018b) that informs this dissertation research. The model

proposes that writing is a complex task involving multidimensional processes that are shaped by both social factors, individual variability, and the characteristics of community within which it occurs (Graham, 2018b). Participation in a particular community of practice and disciplinary culture requires writers to acquire knowledge of norms and expectations, become articulate in the discursive practice of the community, figure out how meaning is constructed and communicated within a particular rhetorical context, and learn to use and negotiate the resources that are available within the community. Thus, providing access to and opportunities for participating in different types of academic reading and writing experiences that promote critical inquiry and analytical thinking is an important step towards ensuring equity and advancing social justice.

The WWC model also highlights the complexity and multidimensionality of writing whose quality is dependent on both macro-level rhetorical elements and micro-level linguistic features. Collectively, the three studies of the dissertation contribute to our understanding of the complexity, multidimensionality, and variability of writing by adding to the knowledge base about the lexical, syntactic, and rhetorical features of academic writing and the challenges multilingual students face. By systematically examining the rhetorical and linguistic features of text-based analytical writing, this dissertation research contributes to our understanding of students' strengths, challenges, and needs when they write within an academic context. This understanding can better inform writing pedagogy and intervention geared towards supporting students' academic literacy skills in a secondary education context. By examining the differences in rhetorical and linguistic features among students with different linguistic backgrounds, the study helps us understand the variability in language and literacy skills of diverse multilingual students. Finally, by modeling the relations among linguistic characteristics, rhetorical elements, and writing quality, this dissertation contributes to our understanding of multidimensionality of

writing and the complex relations and interactions among macro- and micro-level dimensions of writing.

As adolescent students in secondary schools are challenged by meeting the demands of academic writing, classroom teachers and literacy researchers throughout the nation are searching for research that provides them with knowledge base to prepare diverse adolescent students to become confident and competent academic writers who can meet state and national standards and who are well-prepared for college and career pursuits. Access to and participation in academic communities and disciplinary discourse entail being well-versed in linguistic practices of these communities. Thus, meaningfully supporting students in their effort to become well-versed academic writers and strategic language users is an important educational goal. This dissertation research makes a small contribution to this bigger goal.

References

Defazio, J., Jones, J., Tennant, F., Hook, S.A. (2010). Academic literacy: The importance and impact of writing across the curriculum – a case study. *Journal of the Scholarship of Teaching and Learning*, 10, 34–47.

California Common Core State Standards. (2013). *English language arts & literacy in history/social studies, science, and technical subjects.* Retrieved from https://www.cde.ca.gov/be/st/ss/documents/finalelaccssstandards.pdf.

Canagarajah, S. (2002). Multilingual writers and the academic community: towards a critical relationship. *Journal of English for Academic Purposes*, 1, 29–44.

Common Core State Standards Initiative. (2022). *Standards in your state.* Retrieved from http://www.corestandards.org/standards-in-your-state/.

Fitzgerald, J. (2017). How practice-based research informs adolescent English-language learners' composing and compositions. In K. A. Hinchman & D. A. Appleman (Eds.). *Adolescent literacies: A handbook of practice-based research.* (pp. 357–378). Guilford Press.

Flores, N., & Rosa. J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review* 85, 149–171.

Garcia, O. (2020). The education of Latinx bilingual children in times of isolation: Unlearning and relearning. *MinneTESOL Journal*, 36, 1–8.

Garcia, O., & Solorza, C. R. (2020). Academic language and the minoritization of U.S. bilingual Latinx students." *Language and Education*, 35, 505–521.

Graham, S. (2018a). A writer(s)-within-community model of writing. In C. Bazerman, V. W. Berninger, D. Brandt, S. Graham, J. Langer, S. Murphy, ... M. Schleppegrell (Eds.), *The lifespan development of writing* (pp. 272–325). National Council of English.

Graham, S. (2018b). A revised Writer(s)-within-community model of writing. *Educational Psychologist,* 53, 258–279.

Graham, S. & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools — A report to Carnegie Corporation of New York.* Alliance for Excellent Education.

Halliday, M. A. K. (1994). *Introduction to functional grammar* (2nd ed.). Edward Arnold.

Harman, R., & Burke, K. (2020). *Culturally Sustaining Systemic Functional Linguistics: Embodied Inquiry with Multilingual Youth*. Routledge.

Maamuujav, U., Krishnan, J., and Collins, P. (2019) The utility of infographics in L2 writing classes: A practical strategy to scaffold writing development. *TESOL Journal,* 11, 484.

Olson, C. B., Woodworth, K., Arshan, N., Black, R., Chung, H. Q., D'Aoust, C., Dewar, T., Friedrich, L., Godfrey, L., Land, R., Matuchniak, T., Scarcella, R., & Stowell, L. (2020). The pathway to academic success: Scaling up a text-based analytical writing intervention for Latinos and English learners in secondary school. *Journal of Educational Psychology,* 112, 701–717.

Ortmeier-Hooper, C., & Enright, K. A. (2011). Mapping new territory: Toward an understanding of adolescent L2 writers and writing in US contexts. *Journal of Second Language Writing,* 20, 167–181.

Paris, D., & Alim, S. H. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard Educational Review,* 84, 85–100.

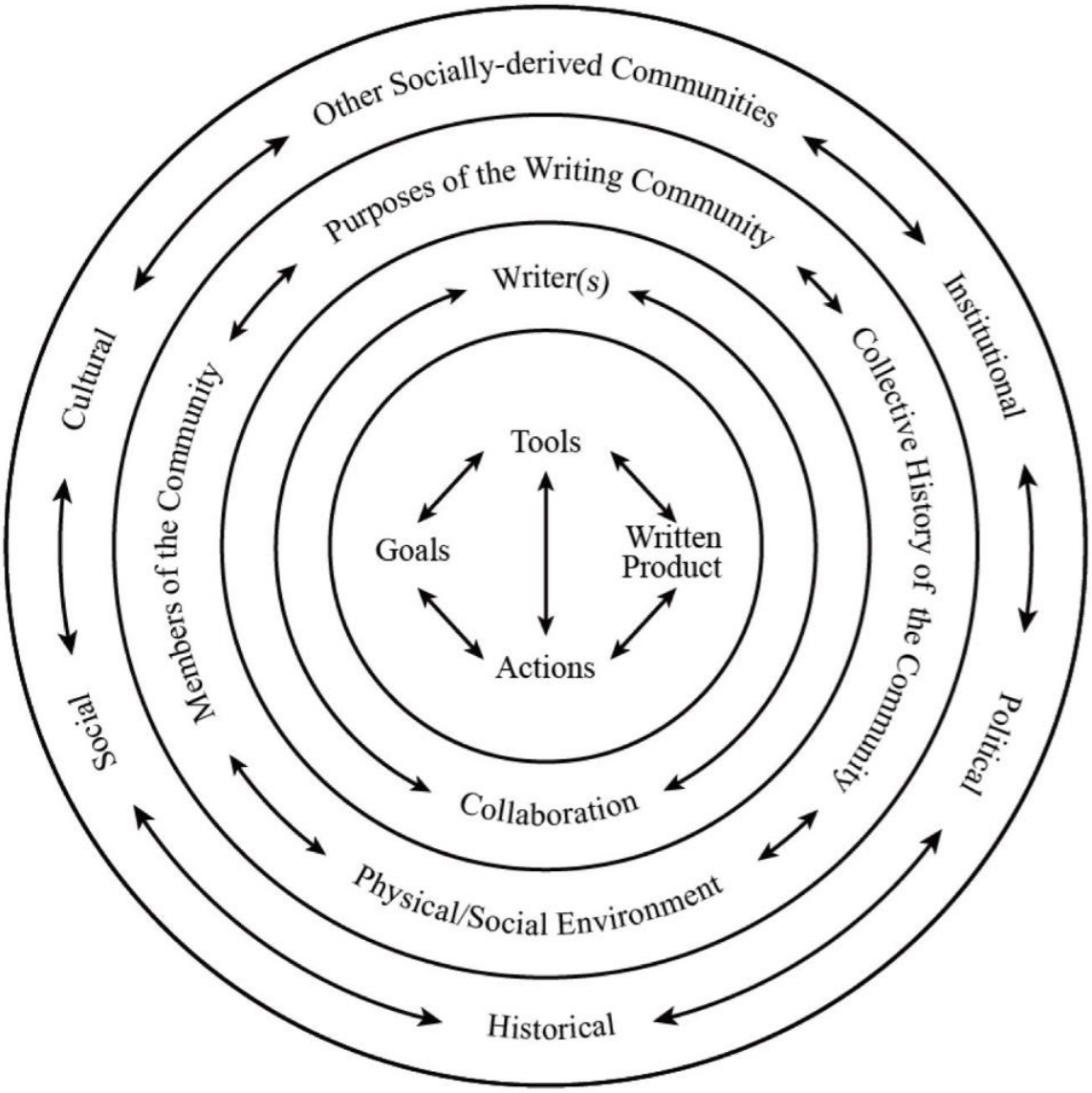
Paris, D., & Alim, S. H. (2017). *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world.* Teachers College Press.

Proctor, C. P., Silverman, R. C., Harring, J. R., Jones, R. L., & Hartranft, A. M. (2019). Teaching bilingual learners: Effects of a language-based reading intervention on academic language and reading comprehension in grades 4 and 5. *Reading Research Quarterly,* 55, 95–122.

Yi, Y., & Kao, C., & Kang, J. (2018). Middle and high school writing. In J. I. Liontas (Ed.). *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons, Inc.

APPENDIX A

Writer(s)-Within-Community Model of Writing



Note. Reprinted from Graham, S. (2018b). A Revised Writer(s)-Within-Community Model of Writing. *Educational Psychologist,* 53, 258–279.

APPENDIX B

Writing Prompts

Prompt 1 — "The Man in the Water"

Writing Situation
After the crash of Air Florida Flight 90 in 1982, Roger Rosenblatt, an award-winning journalist, wrote an article for *Time Magazine* about a man who risked his life in order to save his fellow passengers from the icy waters of the Potomac River. In the end, this man lost his own life in the process of saving others. When a journalist's purpose is strictly to inform, he or she will present the facts objectively. However, Rosenblatt does more than this. He carefully crafts his text to create an impression on the reader.

Writing Directions
After reading "The Man in the Water," select one important theme to write an essay about. Create a theme statement which expresses the author's main point, lesson, or message in the article. Your theme statement will be the thesis of your essay—the claim you make about the writer's message or main idea.

As you develop the main body of your essay, pay special attention to:

- Rosenblatt's description of the man in the water's actions after the plane crash
- The language Rosenblatt uses to describe nature and the relationship between the man in the water and nature (including similes, metaphors, symbols, personification, or other figurative language)
- Rosenblatt's response to the fact that the man in the water lost his life in the process of saving others

In your conclusion:

- Discuss Rosenblatt's purpose in writing "The Man in the Water."
- Revisit the message he wants his readers to take away from reading his article and explain why it is especially significant.

Remember: There is no one theme and therefore no "right" answer to this prompt. What is important is to support your ideas with evidence from the text. Proofread your paper carefully to be sure that it follows the conventions of written English.

What is a Theme?
The theme of a written text is the writer's message or main idea. The theme is what the writer wants you to remember most. Most stories, novels and plays, and sometimes poems have more than just one theme. A character might say something about life that is clearly important. For example, in E.B. White's *Charlotte's Web*, Wilbur says at the end, "Friendship is one of the most satisfying things in the world." That's a statement of one of the book's themes. But, often, you have to be a bit of a detective to discover the theme or themes. The author leaves clues, but it is up to you to put them together and decide what the important message or lesson is.

The article you just read was nonfiction. Although some nonfiction texts are written solely to present facts and information, others are also intended to present the writer's message and influence readers' ideas about people, places, or events. Therefore, nonfiction texts can also contain themes.

(Adapted from Great Source *Reader's Handbook*)

Writing Prompt 2 — "Sometimes, the Earth is Cruel"

Writing Situation
Two days after the Haiti earthquake on January 12, 2010, Leonard Pitts, an award-winning journalist, wrote an article for the *Miami Herald* in which he describes the Haitian people's response to the tragedy which struck their country. When a journalist's purpose is strictly to inform, he or she will present the facts objectively. However, Pitts does more than this. He carefully crafts his text to create an impression on the reader.

Writing Directions
After reading "Sometimes, the Earth is Cruel," select one important theme to write an essay about. Create a theme statement which expresses the author's main point, lesson, or message in the article. Your theme statement will be the thesis of your essay—the claim you make about the writer's message or main idea.

As you develop the main body of your essay, pay specific attention to:

- Pitts' description of the Haitian people's actions after the earthquake
- The language Pitts uses to describe nature and the relationship between the Haitian people and nature (including similes, metaphors, symbols, personification, or other figurative language)
- Pitts' response to the way the Haitian people deal with their tragedy

In your conclusion:

- Discuss Pitts purpose in writing "Sometimes, the Earth is Cruel."
- Revisit the message he wants his readers to take away from reading his article and explain why it is especially significant.

Remember: There is no one theme and therefore no "right" answer to this prompt. What is important is to support your ideas with evidence from the text. Proofread your paper carefully to be sure that it follows the conventions of written English.

What is a Theme?
The theme of a written text is the writer's message or main idea. The theme is what the writer wants you to remember most. Most stories, novels and plays, and sometimes poems have more than just one theme. A character might say something about life that is clearly important. For example, in E.B. White's *Charlotte's Web*, Wilbur says at the end, "Friendship is one of the most satisfying things in the world." That's a statement of one of the book's themes. But, often, you have to be a bit of a detective to discover the theme or themes. The author leaves clues, but it is up to you to put them together and decide what the important message or lesson is.

The article you just read was just nonfiction. Although some nonfiction texts are written solely to present facts and information, others are also intended to present the writer's message and influence readers' ideas about people, places, or events. Therefore, nonfiction texts can also contain themes.

(Adapted from Great Source *Reader's Handbook*)

APPENDIX C

Instructions for Prompt Administration

To the teacher: Please read the instructions carefully before administering the writing task in your class.

Prior to Prompt Administration

- Please **read all these instructions** before administering the reading and writing tasks in your classes**.**
- Review your materials and make sure you have enough of the following materials for your focal class size:
    - Parent Notification Letters (please send one home to each student)
    - Nonfiction Article
    - Pathway Project Reading and Writing Assessment Packet
    - Writing Prompt Packet
    - Class Roster (one per class)
    - Teacher Survey (one per class)

Prior to Day 1

- Distribute Parent Notification Letters (titled Study Information Sheet) a day before administering the reading and writing task

---

Day 1: Reading and Reading Assessment

Reading and Reading Assessment Packets

- Distribute nonfiction article and Reading Assessment Packets to students and ask them to fill out their name, district I.D., school, and class information on the first page. This I.D. is crucial for the study.
- Read out loud the directions to students at the beginning of the Reading Assessment Packet:
  You will have two class periods for this reading and writing assessment. Today, you are working on the reading assessment activity. First, you will read a nonfiction article. Then you will respond to questions that will help you think about what you have read. During another class period, I will return your notes to you and you will need to write an essay about this non-fiction article.
- Next, read the nonfiction article.
- Follow the reading with a discussion of difficult vocabulary words and remind students that italicized words in the text are defined at the end of the packet.
- Have students review the texts and make notes in the margins.
- Return to the Reading Assessment Packet and have students respond to a series of questions and activities to guide them through the text. Encourage students to move onto the next question approximately **every 5-10 minutes**. Give students the remainder of the period to complete the activities

Suggested Timeline:

- 5 minutes: Pass out the nonfiction article, and the Reading Assessment packets. Read the directions out loud.
- 10 minutes: Read the text out loud and go over difficult vocabulary words.
- 5 minutes: Summarize events.
- 5 minutes: Sketch/Visualize
- 5 minutes: Analyze Author's Craft
- 5 minutes: Analyze Theme
- 5 minutes: Evaluate & Reflect

*Teachers, we ask that you **do *NOT* help the students** in any way during the reading or writing time.

- **At the end of Day 1**, collect the **Reading and Reading Assessment packets** and store them safely for distribution on Day 2.

Day 2: Writing (In Response to the Reading and Reading Assessment Prompts)

Writing In Response to The Prompt

- Redistribute the Day 1 Reading and Reading Assessment packets to students.
- Distribute Writing packets to students.
- Ensure that all students have a pen to complete the Writing Prompt Packet.
- Ask students to complete the cover page on the packet.
- Tell students that:

  Today is our writing day. Before you begin answering your prompt, follow along as I read these instructions.
- Read out loud the writing situation and directions on the third page of the Writing Prompt Packet.
- Tell students to use the space on the following **lined** pages to write an essay in response to the reading they have been given. Remind them again to write in pen.
- Suggest that they reread the prompt silently and underline what they are supposed to do.
- Students may take a few minutes to plan their essay on the Planning Page on page 8 before they write.
- Remind them to review the Reading Assessment packets to generate ideas for writing.

*Teachers, you can define a vocabulary word students don't understand. But please don't coach the students and tell them what to write!!

- Give them the full class period to write.
- Place all completed Writing packets in the Prompt Return envelope.
- Collect all packets and sort as follows:

Return to UCI	Keep
Completed Writing Packets	Packets for absent students (to be made up within 2 weeks)
Unused Reading Packets	Completed Reading Packets
Unused Reading Assessment Packets	Completed Reading Assessment Packets
Unused Writing Packets	
Class Roster	
Teacher Survey	

Make Ups

- During the following two weeks, please try to give each absent student the chance to make up the prompt activity. You must provide two full class periods.
- If, for any reason, a student is unable to complete the assessment, please explain why in the space given on the bottom of the cover sheet (even if the writing packet is blank).
- Please put all make-up packets in the Make-Up Prompt Return envelope and return to UCI.
- We will expect one Writing Prompt packet returned for each student on the class roster.

On behalf of the entire Pathway to Academic Success research team, we thank you!

APPENDIX D

Analytical Writing Assessment Rubric for Text-based Analytical Writing

Scoring Guide for "Sometimes, the Earth is Cruel" and "The Man in the Water'

Note: Papers at all levels of achievement described below will contain some or all of the characteristics listed as criteria for each particular score.

6 Exceptional Achievement

- Writer introduces the subject, giving enough background for the reader to follow the interpretation he/she offers in response to the prompt. • Overall, writer presents an especially thoughtful theme statement which expresses the author's main point, lesson, or message.
- Writer clearly and carefully addresses all parts of the writing task:
  - Writer clearly discusses the author's description of the Haitian people's actions after the earthquake or the man in the water's actions after the plane crash.
  - Writer thoughtfully analyzes the language Pitts or Rosenblatt uses to describe nature and the relationship between either the Haitian people and nature or the man in the water and nature (including similes, metaphors, symbols, personification, or other figurative language).
  - Writer thoughtfully discusses Pitts' or Rosenblatt's response to the way the Haitian people respond to their tragedy or to the fact that the man in the water lost his life in the process of saving others.
  - Writer thoughtfully discusses the author's purpose in writing his article, restating his message, and explaining why that message is especially significant.
- Writer skillfully weaves numerous references from the text into the paper to support his/her ideas.
- Writer interprets well and brings the paper to a logical conclusion.
- Writer uses precise, apt, or descriptive language and sentence variety. • Paper has few errors in the conventions of written English.

Note: The writer does not need to discuss these issues in a specific order to receive a high score as long as he/she addresses all parts of the prompt.

5 Commendable Achievement

- Writer introduces the subject, giving some background for the reader to follow the interpretation he/she offers in response to the prompt.
- Overall, writer presents a thoughtful theme statement which expresses the author's main point, lesson, or message.

- Writer addresses all parts of the writing task to some extent:
    - Writer discusses the author's description of the Haitian people's actions after the earthquake or the man in the water's actions after the plane crash.
    - Writer analyzes the language Pitts or Rosenblatt uses to describe nature and the relationship between either the Haitian people and nature or the man in the water and nature (including similes, metaphors, symbols, personification, or other figurative language).
    - Writer discusses Pitts' or Rosenblatt's response to the way the Haitian people respond to their tragedy or to the fact that the man in the water lost his life in the process of saving others.
    - Writer discusses the author's purpose in writing his article, restating his message, and explaining why that message is especially significant.
- Writer weaves several references from the text into the paper to support his/her ideas.
- Writer interprets reasonably well and brings the paper to a logical conclusion.
- Writer uses some precise, apt, or descriptive language and sentence variety.

4 Adequate Achievement

- Writer orients the reader adequately by giving at least some introductory context.
- Writer may begin unsteadily but reaches a focus or point.
- Overall, writer presents an adequate theme statement which expresses the author's main point, lesson, or message. This statement may not be in the introduction.
- Writer addresses most parts of the writing task (at least 3 of the 4 tasks).
    - Writer adequately discusses the author's description of the Haitian people's actions after the earthquake or the man in the water's actions after the plane crash but not in much detail.
    - Writer adequately discusses but may not analyze the language Pitts or Rosenblatt uses to describe nature and the relationship between either the Haitian people and nature or the man in the water and nature (including similes, metaphors, symbols, personification, or other figurative language). The writer may only briefly refer to the language, discuss only one example, or quote the language without really analyzing it.
    - Writer provides an adequate discussion of Pitt's or Rosenblatt's response to the way the Haitian people respond to their tragedy or to the fact that the man in the water lost his life in the process of saving others.
    - Writer discusses the author's purpose but may not thoroughly discuss why that message is especially significant.
- Writer weaves some references from the text into the paper to support his/her ideas.

- Writer interprets in less depth than a 5 paper. While the paper has a conclusion, the development of the paper toward that conclusion may be less logically organized.
- Paper has less apt, precise or descriptive language than a 5 or 6 paper and little sentence variety.
- Paper has some errors in the conventions of written English, but none that interfere with the message.

3 Some Evidence Achievement

- Writer introduces the topic perfunctorily or simply dives in--answering the questions without developing a clear introduction.
- Overall, writer's essay may not include a theme statement; the essay may be superficial or rely on the retelling of events; or a one sentence theme statement may be tacked on at the end.
- Writer responds to some but not all of the writing tasks in the prompt:
    - Writer briefly summarizes but does not clearly describe how the Haitian people or the man in the water respond to the tragedy.
    - Writer summarizes but does not discuss in any depth how the author uses language to portray the relationship between the Haitian people and nature or the man in the water and nature. Writer may not identify uses of figurative language.
    - Writer may fail to discuss Pitt's or Rosenblatt's response to how the Haitian people deal with the tragedy or the fact that the man in the water lost his life in the process of saving others.
    - Writer may fail to discuss author's purpose or may fail to discuss why that message is especially significant.
- Writer uses few, if any, references from the text into the paper to support his/her ideas.
- Writer may have limited facility with descriptive language and write simple sentences (i.e. no sentence variety).
- Writer seems to lack skill in presenting his/her own ideas and may fall back on plot summary. The writer's conclusion may be slightly off base or inadequately developed.
- Paper may have errors in the conventions of written English, some of which interfere with the reader's comprehension.

2 Little Evidence of Achievement

- Writer provides no introduction, or it is brief and unfocused.
- Writer does not seem to understand what a theme is and simply retells what happened.
- Writer may fail to discuss how the authors describe the actions of either the Haitian people or the man in the water.

- Writer may not understand how the author is using language to portray man versus nature.
- Writer may focus solely on what happened and not to the author's response to how people deal with tragedy.
- Writer may not understand the author's purpose in writing the article or why the theme is significant.
- Writer talks in generalities and fails to provide specific references to the text.
- Conclusion may be abrupt or missing.
- Writer has extremely limited facility with language and little command of sentence structure.
- Paper has many errors in the conventions of written English, many of which interfere with the writer's message.

1 Minimal Evidence of Achievement

- Context/introduction is missing, abrupt or confusing.
- Writer merely retells the story briefly and does not address the prompt.
- Writer misreads or has a very limited understanding of the article.
- Writer has very poor command of how to construct an essay.
- Paper has so many errors in the conventions of written English that the writer's meaning is obscured.
- Writer may also just copy down chunks of text that are taken directly from the article.

APPENDIX E

Analytic Scoring Framework

Adapted from Pathway's Analytic Coding Framework that was generated based on Pathway's Prompts and Rubrics (NWP C3WP UST; NWP AWC-LA, & Literacy Design Collaborative)

Who are you?

☐ Scorer 1

☐ Scorer 2

Essay code: Copy and paste the essay code here

_____

1. **Lexical Sophistication:** How does the essay demonstrate command of diction and word choice? (Discount language "borrowed" from source text)
   - Effective: It uses precise, apt, descriptive, and/or sophisticated words appropriately to engage the reader and enhance meaning.
   - Moving from Competent to Effective.
   - Competent: It uses some precise, apt, descriptive, and/or appropriate words to communicate meaning. There might be a few lexical errors that do not interfere with meaning.
   - Moving from Developing to Competent.
   - Developing: It uses language that is functional and achieves purpose but is ordinary and lacks precision and variety. There might be occasional lexical errors, but these errors often do not interfere with meaning.
   - Moving from Ineffective to Developing.
   - Ineffective: It has very limited, flawed, or simplistic words, and frequent lexical errors that may interfere with the meaning.
2. **Syntactic Complexity and Variety:** How does the essay demonstrate syntactic complexity and variety?
   - Effective: It demonstrates effective phrasing and includes sentences that vary in structure and length, creating extremely effective structure.
   - Moving from Competent to Effective.
   - Competent: It usually has effective phrasing (although a few lapses in syntax may be present) and it has some variation in structure and length but lacks complexity in both structure and variety.
   - Moving from Developing to Competent.

- Developing: Phrasing is adequate, with some lapses in syntax; the essay has little variation in sentence structure and length. The essay doesn't exhibit use of varied syntactic structures but use basic and frequent structures.
- Moving from Ineffective to Developing.
- Ineffective: Sentences have structural and word placement problems that result in confusion and unnatural phrasing, and the essay is marked by a pattern of simple and choppy sentences.

3. **Cohesion:** How are the ideas and sentences in the text connected together to form a cohesive text?
   - Effective: The sentences and ideas are cohesive, and a wide range of cohesive devices are used appropriately (appropriate repetition of key noun/words, linking adverbials for transition signals, reference words and consistent use of pronouns, appropriate use of ellipsis and substitution). Ideas flow well and are tightly connected throughout the essay.
   - Moving from Competent to Effective.
   - Competent: The sentences and ideas are usually cohesive (a few lapses in flow may be present), and several cohesive devises are attempted. May not have sophisticated linking adverbials/words to signal transition but other cohesive devices are used appropriately. Ideas usually connected although some lapses might be present.
   - Moving from Developing to Competent.
   - Developing: Some appropriate cohesive devices are used to connect ideas although some may be inadequate or overused; Some omissions of cohesive devices may affect flow and connected of ideas. The reader often has to guess how the ideas are connected.
   - Moving from Ineffective to Developing.
   - Ineffective: Lacks the use of appropriate cohesive devices; thus, it extremely difficult to follow. Ideas are rather random and do not follow logical progression.

4. **Source use:** Overall, how would you describe the student's use of source material?
   - Effective: Skillfully integrates varied and appropriate source material in a purposeful manner. A combination of quotation, paraphrasing, summary is integrated seamlessly and are used strategically and purposefully.
   - Moving from Competent to Effective.
   - Competent: Mostly integrates varied and appropriate source material in a purposeful manner although certain type of sources integration (quotations, paraphrasing etc) might be dominant. Usually, the source material is well integrated with some introductory phrases and reporting verbs although there might be some floating quotations.
   - Moving from Developing to Competent.

- Developing: Uses some source material but does not always use it in a purposeful manner. Excessive summary and/or floating/random quotations that do not support claim may be present.
- Moving from Ineffective to Developing.
- Not used: Provides limited reference to source material and fails to refer to the source to provide context, support claim, and advance an argument.

5. **Balance of summary, evidence, and commentary:** To what extent does the essay demonstrate a balanced use of summary, evidence, and commentary?
   - Effective: Skillfully balances, integrates, and weaves together summary, evidence, and commentary when appropriate.
   - Moving from Competent to Effective.
   - Competent: The essay has some balance of summary, evidence, and commentary when appropriate.
   - Moving from Developing to Competent.
   - Developing: The essay integrates at least two elements (summary, evidence, commentary) although one of the elements may dominate. Predominantly integrates one element.
   - Moving from Ineffective to Developing.
   - Ineffective: The essay relies on only one element. It may only summarize the article, or it may only draw on the writer's ideas without a reference to the source.

APPENDIX F

Manual Coding Framework and Guidelines

Code each sentence based on the following criteria and guidelines

Level 1 Coding: Determine if the sentence is conventional or unconventional. Unconventional sentences are run-ons, comma splices, fragments, and problematic sentences that have structural and semantic problems that interfere with the meaning it attempts to convey.

- ☐ Conventional
- ☐ Unconventional

Level 2 Coding: Code both conventional and unconventional sentences for the subcategories.

Code the conventional sentences based the following subcategories:	Code unconventional sentences based on the following subcategories:
☐ Run-on/Comma Splice	☐ Simple or One-clause Structure
☐ Fragment or Incomplete Sentence	☐ Coordinated[@] or Compound Sentence
☐ Faulty/Problematic Sentence	☐ Subordinated* or Complex Sentence
	☐ Coordinated and Subordinated Sentence

Note. Examples of each sentence subcategory is provided in the next page. *Include both finite and nonfinite clauses in subordinated structure. [@]Identify the conjunction used to coordinate the sentence.

Level 3 Coding: Code all subordinated and coordinated-subordinated sentences based on the following subcategories.

- ☐ Finite adverbial clause* (two clauses)
- ☐ Finite complement clause* (two clauses)
- ☐ Finite noun modifier* (two clauses)
- ☐ Multiple finite dependent clauses (3 or more clauses)
- ☐ Nonfinite adverbial
- ☐ Nonfinite complement
- ☐ Nonfinite noun modifier
- ☐ Multiple nonfinite dependent clauses
- ☐ A combination of finite and nonfinite dependent clauses*
- ☐ Subordinate and coordinated sentences*

Note. *Identify the connective used. Identify grammatical functions of the finite and nonfinite dependent clauses for 1) combination of finite and nonfinite clauses, and 2) subordinated and coordinated clauses.

Level 4 Coding: Code the sentences based on their rhetorical functions using the categories below.

- ☐ Summary/paraphrased sentence
- ☐ Floating quotation
- ☐ Quoted with introductory phrase and reporting verb
- ☐ Embedded quote (quotation integrated into student's own sentence)
- ☐ Verbatim copied
- ☐ Patchwritten* sentence
- ☐ Interpretive (claim, opinion, hypothetical, anecdotal, commentary)

Note. *If 3 or more words are different (added, deleted, or rearranged) in the original sentence, consider this patchwritten sentence.

Refer to the sample sentences below:

Examples of Unconventional Sentences	
Run on/Comma Splice	○ Haiti people were poor they had to do what they need everyone will help people if they were injured or had been stuck. ○ In "Sometimes the Earth is cruel" the author explains the situation of the country of Haiti one of the themes in this article is human competition. ○ Nature is like a human, it can either destroy itself or keep building and rebuilding.
Fragment	○ Displaying his bravery & compassion for the others. ○ Because you can lose everything with just one blink of an eye. ○ Which then ties back to the theme of an eye for an eye. ○ The balance between good and evil. ○ To just ignore the cruelties of the earth's nature.
Faulty/Problematic	○ The Problem is the fall rains will not stop. ○ Bad things happen to people but people always overcame for explain people go back to their regular life.

Examples of Simple, Coordinated, and Coordinated + Subordinated Sentences	
Simple/One-clause	o His description disasters are so powerful. o Adults and kids face hard stuff here and there. o It will get rebuilt by someone else. o The man in the Water has his own natural powers.
Coordinated/Compound	o He chose to "challenge death" *and no one really knows why*. o Many innocent people were dead *and their houses were destroyed*. o The damage can happen anytime anywhere *and it will never let you know when and where.* o This man lost the battle but *he was able to save 5 other men in the process.* o He didn't have super strength or laser eyes, but he had a kind heart and a brave soul. o It was not his duty nor was he obligated to do so. o Not only is this story a real life attention grabber, but it also tells a story about the good in people.
Coordinated + Subordinated	o He knew [<u>he would die and suffer</u>] *but that didn't matter to him* [<u>because he knew [that he had to sacrifice his life for others]</u>] *and that's a brave thing.* o The author knows [<u>that the people can't control the way the earth functions</u>], *so the only way* [<u>*to handle their loved one's death*</u>] *is by* [<u>*only grieving and memorializing the lost/dead.*</u>] o <u>[Once help arrived]</u> they threw a floatation ring [<u>to drag the cold passengers to shore</u>], *but one passenger kept* [<u>*passing the ring to other passengers beside him*</u>.]

Examples of Subordinated (Finite and Nonfinite)	
Finite adverbial	○ No one would have blamed him if he chose to do so. ○ The Earth is cruel because we don't take care of it. ○ When something tragic happened, they just move along with there business and forget the past.
Finite complement	○ I think the theme is courage and bravery. ○ In the article the theme is that anyone can be a hero. ○ Additionally, the author says that there has been a history of political instability. ○ We learn how to get through obstacles and how to change our thought of mind.
Finite noun modifier	○ The language he uses to write this article is more of a poetic and a first-person kind of view. ○ However, human cruelty is something that can be avoided, justified, punished, or ignored completely. ○ He was just an ordinary man who was responsible for the emotional impact of the disaster.
Multiple finite clauses	○ In conclusion, the power of nature is kindness and caring [because the man in the water helped people] [that were drowning] [even though he was also fighting for survival.] ○ We know [what to do] [when history repeats itself.] ○ I think [the theme is [that [no matter how much you fall,] you have to get back up.]
Nonfinite adverbial	○ The author uses figurative language to describe the destructive ways of the water and nature. ○ The hurricanes destroy people's houses, leaving them homeless. ○ The article informs us about the plane crash while delivering the message about an ordinary man and his selfless actions.
Nonfinite complement	○ Using repetition gives his article a good flow. ○ It is important to know about problems of other people. ○ Pitt's purpose in this article is to express human stubbornness and determination.
Nonfinite noun modifier	○ These are disasters authored by human hands and by human greed and corruption. ○ This man was just an ordinary middle-aged man planning to relax on his flight like everybody else.
Multiple nonfinite clauses	○ The purpose of the article was [to inform about the event] [while delivering the message about an ordinary man and his selfless actions.] ○ [Losing his life while saving others] he set a great example for everyone else [to do great things not for their own satisfaction but for the good for others.]
Finite + nonfinite	○ Rosenblatt carefully constructed the image of man [standing up against nature] in order [to show that, [even though a civilian lost the physical battle,] the emotional battle had been won in the minds of those [who saw him.]] ○ I think [that Pitt is trying [to say how the Earth can be cruel] [by showing us the cruel thing (the Earth has done)].] ○ This story was written [to raise awareness about [what's happening in Haiti]]

Examples of Quotations	
Embedded quote:	o This man was as Rosenblatt described “ordinary man on an ordinary plane with other ordinary passengers.”
Introduced quote:	o The author said, “Sometimes the skies turn barren and will not rain.”
Floating quotation:	o “Sometimes the skies turn barren and will not rain.”

APPENDIX G

Summary of Data

Total Number of Essays		410
By States		
	Arizona	119
	California	70
	Minnesota	82
	Texas	81
	Utah	25
	Wisconsin	33
By Gender		
	Female	216
	Male	194
By Grade		
	7th grade	38
	8th grade	87
	9th grade	75
	10th grade	87
	11th grade	30
	12th grade	93
By Home/First Language (L1)	Spanish	222
	English	92
	Somali	16
	Hmong	12
	Arabic	11
	Burmese	9
	Amharic	8
	Karen	5
	Vietnamese	5
	Russian	3
	Swahili	3
	Assyrian	2
	English/Spanish	2
	French	2
	Malay	2
	Rohingya	2
	Yoruba	2
	Oromo	2
	Oromo/Amharic	1
	Cantonese/Taiwanese	1
	Falam	1
	Ibibio	1
	Laotian	1
	Micronesian	1
	Mina	1
	Tagalog	1
	Tigrinya	1
	Twi	1

APPENDIX H

Study Information Sheet

University of California, Irvine
Study Information Sheet

The Pathway to Academic Success: A Cognitive Strategies Approach to Text-Based Analytical Writing to Improve Academic Outcomes for Secondary English Learners

Sub-study 1: Examining Linguistic Features of Student Writing

RESEARCH TEAM

Undarmaa Maamuujav
Ph.D. Student
School of Education, UCI
umaamuuj@uci.edu

Carol Booth Olson
Principle Investigator
Professor Emerita, School of Education
Director, UCI Writing Project

Huy Chung
Research Director,
School of Education, UCI

Study Location:
School of Education
University of California, Irvine

In the instance of parental permission, "You" refers to "Your child."

- Please read the information below and ask questions about anything that you do not understand. A researcher listed above will be available to answer your questions.

- You are being asked to participate in a research study. Participation in this study is voluntary. You may choose to skip a question or a study procedure. You may refuse to participate or discontinue your involvement at any time without penalty or loss of benefits. You are free to withdraw from this study at any time. **If you decide to withdraw from this study, you should notify the research team immediately**.

- You are being asked to participate in a research study to examine linguistic features of student writing in an effort to identify common challenges, as well as unique language issues, students have when writing academic essays. The study is being conducted by UCI across eight universities (UC Irvine, Arizona State University, Brigham Young University, Texas State University-San Marcos, Northeastern Illinois University, University of Minnesota, University of Nevada-Las Vegas, University of Wisconsin-Milwaukee).

- You are eligible to participate in this study if your teacher consented to participate in this study.

- The research procedures involve reading two texts and filling out a graphic organizer on one day and writing an essay about the two texts on the second day.
- Possible risks/discomforts associated with the study are anxiety associated with writing; however, these risks may be minimal.

- There are no direct benefits from participation in the study. However, this study may explain a students' linguistic needs for academic writing in English and the factors that contribute to high-quality writing.

- There are no alternative procedures available. The only alternative is not to participate in this study.

- All research data collected will be stored securely and confidentially at UCI behind a locked door inside of a locked cabinet. Only study codes will be sharable.

- The research team, authorized UCI personnel, and regulatory entities, may have access to your study records to protect your safety and welfare.

- While the research team will make every effort to keep your personal information confidential, it is possible that an unauthorized person might see it. We cannot guarantee total privacy.

- The researchers intend to keep the research data for approximately 7 years.

- You will not be compensated for your participation in this research study.

- If, during the course of this study, significant new information becomes available that may relate to your willingness to continue to participate, this information will be provided to you by the research team listed at the top of the form.

- If you have any comments, concerns, or questions regarding the conduct of this research please contact the researchers listed at the top of this form.

- Please contact the UCI Institutional Review Board by phone, (949) 824-6662, by e-mail at IRB@research.uci.edu or at 141 Innovation Drive, Suite 250, Irvine, CA 92697 if you are

unable to reach the researchers listed at the top of the form and have general questions; have concerns or complaints about the research; have questions about your rights as a research subject; or have general comments or suggestions.

What is an IRB? An Institutional Review Board (IRB) is a committee made up of scientists and non-scientists. The IRB's role is to protect the rights and welfare of human subjects involved in research. The IRB also assures that the research complies with applicable regulations, laws, and institutional policies.