UNIVERSITY OF CALIFORNIA,
IRVINE

Language Use and Self-efficacy in Writing of Linguistically Diverse Students

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Jiali Wang

Dissertation Committee:
Professor Young-Suk Grace Kim, Chair
Professor Elizabeth D. Peña
Professor Emerita Carol Booth Olson
Regents and the Emily Warner Professor Steve Graham
Associate Professor Michael Hebert

2024

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Acknowledgments

Completing this dissertation would not have been possible without the unwavering support of my dissertation committee, friends, colleagues, and family. First and foremost, I would like to express my deep gratitude to Dr. Young-Suk Grace Kim, my dissertation chair for guiding me from the very beginning of my experiences of research. Her thoughtful insights, constructive feedback, and rigorous requirements were pivotal in honing my research and writing skills. Next, I am profoundly grateful for Dr. Elizabeth Peña, my co-advisor, who nurtured my passion and supported my exploration of the communication sciences and disorders field. The learning and collaboration in the HABLA lab have taught me invaluable lessons that will guide me throughout my future career. My gratitude also extends to Dr. Carol Booth Olson, the PI of my GSR for her generous support that made it possible for me to be involved in research and collaboration opportunities that are critical for my growth as a scholar. I am also deeply grateful for Dr. Steve Graham's guidance and wisdom in writing research, support in realizing my own research ideas, and the passion I've gotten from him in being a responsible and rigorous scholar. I'm also grateful to Dr. Michael Hebert for his constructive feedback and support on my dissertation work to improve its quality.

I also thank my friends and colleagues for their professional and personal help. I'm especially thankful for my most supportive friends and colleagues, Joseph Hin

Yan Lam and Molly Ann Leachman, for their consistent support and encouragement throughout each step of my pursuit of this PhD. Joseph has provided me with tremendous help, resources, and motivation when I struggled and supported me to become an independent researcher. Molly has accompanied me through numerous academic and life hardships over the years and helped me become a better person. I'm thankful for the faith and patience they have in me and their invaluable and reliable support whenever I'm in need. I also appreciate Danyang Wang as both a wonderful friend and supportive mentor for her encouragement when I'm in doubt of myself. I'm deeply thankful to Jacob Steiss, Jenell Krishnan, and Huy Chung for the opportunities and mentoring from the WRITE Center, which allowed me to develop many new and essential skills. I'm also grateful to other important friends and colleagues, Michael Cooper, Sirui Wan, Yi Feng, Yucheng Cao, Rocío Castiñeira Seoane, Alex Choi-Tucci, Albert Li, Cecilia Perez, and Mariana Gomez for their support that helped me overcome obstacles and the collaboration opportunities they brought me, and Mia Liu, Tianshu Yang, and Qingan Zhou for being my remote but reliable friends.

Last but not least, I would like to thank my family, my mom and dad, my grandma and grandpa. I'm forever grateful for your countless sacrifices and love. And thank you Taifei, Aolifu, and Xiang for always being there as my family while I'm away from home.

Vita

Jiali Wang

UNIVERSITY OF CALIFORNIA, IRVINE.	Irvine, CA
Ph.D. Candidate in Education (HDiC concentration)	2019 - 2024
VANDERBILT UNIVERSITY	Nashville, TN
M. Ed. in English Language Learners	2017 - 2019
BEIJING JIAOTONG UNIVERSITY	Beijing, China
B.A. in Linguistics and Literature	2013 - 2017

PUBLICATIONS

Peer-reviewed

Steiss, J., **Wang, J.**, Kim, Y., Olson. B. C. (in press). Challenges and Dimensions of Secondary Students' Source-Based Argument Writing in History. *Written Communication*. Manuscript accepted for publication.

Wang, J., Kim, Y.-S. G., & Cho, M. (2024). Lexical and Syntactic Features in Narrative and Opinion Genres and Their Relations to Writing Quality in Fourth Grade Writing. *Journal of Research in Reading*.

Steiss. J., Krishnan, J., & **Wang, J.** (2024). Designing Writing Prompts to Elicit Students' Historical Thinking. *The Social Studies*.

Steiss, J., Tate, T. P., Graham, S., Cruz, J., Hebert, M., **Wang, J.**, Moon, Y., Tseng, W., Warschauer, M. (2024). Comparing the Quality of Human and ChatGPT Feedback on Students' Writing. *Learning and Instruction*.

Bedore, L., Peña, E. D., Sjovabasappa, P., Perez C., Granados, A., Leachman, M., & **Wang, J.** (2023) Nature of Vocabulary Change following Language and Literacy Together Intervention. *American Journal of Speech-Language Pathology*, 1-16. https://doi.org/10.1044/2023_AJSLP-23-00042

Cho, M., Kim, Y., & **Wang, J.** (2023). Perspective Taking and Language Features in Secondary Students' Text-Based Analytical Writing. *Scientific Studies of Reading, 27*(3), 199-214. https://doi.org/10.1080/10888438.2022.2132864

Under Review

Lam, J. H. Y., **Wang, J.**, Pratt, A. S., & Peña, E. D. *Reading Comprehension Intervention in Children with Developmental Language Disorder: A Scoping Review*. Manuscript submitted for publication.

Lam, J. H. Y., **Wang, J.**, Steiss, J., Tate, T, & Kim, Y. -S. G. *Text-based Argumentative Writing Profiles of Monolingual and Bilingual Secondary Students*. Manuscript submitted for publication.

Lam, J. H. Y., **Wang, J.**, Wang, D., Bedore, L. M, & Peña, E. D. Spanish and English Morphosyntax Changes in Bilingual School-age Children With and Without DLD: A One-year Longitudinal Study. Manuscript submitted for publication.

Seoane, R. C., **Wang, J.**, Cao, Y., Kim, Y.-S. G. Unpacking the Relation between Oral Language and Written Composition: A Meta-analysis. Manuscript submitted for publication.

Wang, J., Graham, S., Kim, Y.-S. G., & Steiss, J. Testing the Dimensionality and Generality Hypothesis Using an Argumentative and History Specific Writing Self-efficacy Scale. Manuscript submitted for publication.

Wang, J., Kim, Y.-S. G., Lam, J. H. Y., & Leachman, M. A. A Meta-Analysis of Syntactic Features in Writing and Writing Outcomes. Manuscript submitted for publication. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4790890

PRESENTATIONS

1. **Wang, J.**, Graham, S., Steiss, J., & Kim, Y.-S. G. (2024). *An investigation of Dimensionality and Validity of Domain-general and Domain-specific Writing Self-efficacy Scales*. [Oral paper presentation]. 2024 Society for the Scientific Study of Reading Annual Conference. Copenhagen, Denmark.

2. Lam, J. H. Y., **Wang, J.**, Steiss, J., & Kim, Y.-S. G. (2024). *Text-based Argumentative Writing Profiles of Monolingual and Bilingual Secondary Students.* [Oral paper presentation]. 2024 Society for the Scientific Study of Reading Annual Conference. Copenhagen, Denmark.

3. **Wang, J.**, Lam, J. H. Y., Wang, D., Anaya, J. B., Bedore, L. M., & Peña, E.D. (2024). *Spanish and English Morphosyntax Changes in Bilingual School-age Children With and Without DLD*. [Conference session]. 44th Annual Symposium on Research in Child Language Disorders. Madison, WI.

4. **Wang, J.** & Kim, Y.-S. G. (2024). *Writing Self-efficacy in Secondary Students of Various Levels of Language Proficiencies*. [Poster presentation]. 44th Annual Symposium on Research in Child Language Disorders. Madison, WI.

5. Gomez, M., **Wang, J.**, Lam, J. H. Y., Ye, S., Rhodes, K. T., Peña, E. D. (2024). A Meta-analysis of the Relationship between Language and Mathematics in Bilinguals and the Task Effect. [Poster presentation]. 44th Annual Symposium on Research in Child Language Disorders. Madison, WI.

6. Ye, S., Lam, J. H. Y., Gomez, M., **Wang, J.**, Rhodes, K. T., Peña, E. D. (2024). *Revisiting Cummins's Threshold Hypothesis: the Relationship between Language and Mathematics in Bilinguals.* [Poster presentation]. 44th Annual Symposium on Research in Child Language Disorders. Madison, WI.

7. Moon, Y., Steiss, J., **Wang, J.**, & Collins, P. (2024). *Effectiveness of a Writing Intervention in History Classrooms: Do They Vary by Teacher and Student Factors?* [Research Roundtable]. 2024 American Educational Research Association. Philadelphia, PA.

8. Steiss, J., Tate, T. P., Graham, S., **Wang, J.**, Moon, Y., Tseng, W., Warschauer, M., & Hebert, M. (2024). *Comparing the Quality of Human and ChatGPT Feedback on Students' Writing*. [Oral presentation]. American Educational Research Association. Philadelphia, PA.

9. Pham, J., Wu, J., Wang, M., Steiss, J., Moon, Y., & **Wang, J.**. (2024). *Investigating the Role of Teaching Strategies and Self-efficacy in Student Writing Development in Argumentative Writing in History* [AERA Undergraduate Education Research Poster] 2024 American Educational Research Association. Philadelphia, PA.

10. Lam, J. H. Y., **Wang, J.**, Leachman, M. A., Pratt, A. S., & Peña, E. D. (2023). *A Systematic Review on Reading Comprehension Intervention in Children with Developmental Language Disorders*. [Poster presentation]. 2023 ASHA Convention. Boston, MA.

11. **Wang, J.**, Kim, Y.-S. G., Lam, J. H. Y., & Leachman, M. A. (2023). *A Meta-Analysis of Syntactic Features in Writing and Writing Outcomes.* [Poster presentation]. 2023 ASHA Convention. Boston, MA.

12. Perez, C., Shivabasappa, P., Granados, A., Leachman, M., **Wang, J.,** Bedore, L. M., & Peña, E. D. (2023). *Nature of vocabulary change following Language and Literacy Together intervention.* [Poster presentation]. 43rd Annual Symposium on Research in Child Language Disorders. Madison, WI.

13. Cho, M., Kim, Y., & **Wang, J.** (2021). *Perspective Taking and Language Use in Secondary Students' Text-based Analytical Writing*. [Oral paper presentation]. Society for the Scientific Study of Reading Conference 2021.

14. **Wang, J.**, Kim, Y.-S. G., & Cho, M. (2021). *An Investigation of Linguistic Features Across Narrative and Opinion Genres in 4th Grade Writing*. [Poster presentation]. 2021 Society for the Scientific Study of Reading Annual Conference.

RESEARCH EXPERIENCE

UNIVERSITY OF CALIFORNIA, IRVINE   Irvine, CA

HABLA (Human Abilities in Bilingual Language Acquisition) Lab

Spanish-English Bilingual Sentence Repetition Project   Apr. 2024-Now

TELL Mandarin Bilinguals Project   Nov. 2022 - Mar. 2023

Language Sample Analysis   Apr. 2020 - 2021

The National WRITE Center   Mar. 2020 – *Dec. 2023*

UCLAIMS (Using College-Level Argumentation to Improve Multiple Source Writing)

C3WP (College, Career, and Community Writer's Program) project

VANDERBILT UNIVERSITY   Nashville, TN

    LLD-CLED Research Group   Sep. 2018 - May, 2019

    IES WRITING PROJECT   Feb. - Mar. 2018

    MORPHCAT PROJECT   Feb. - Mar. 2018

TEACHING EXPERIENCE

UNIVERSITY OF CALIFORNIA, IRVINE   Irvine, CA

Teaching Assistant – EDUC15 Statistics for Educational Research   Mar. 2024 – *June. 2024*

Teaching Assistant – EDUC25 Introduction to Education   Jan. 2023 – Mar. 2024

Teaching Assistant – EDUC10 Educational Research Design   Apr. - Jun. 2023

Teaching Assistant – EDUC30 21st Century Literacies.   Jun. - Aug. 2020

Teaching Assistant – EDUC151/PSB192V/LSCI182V Language & Literacy.   Jan. - Mar. 2020

Reader – EDUC326 Curriculum and Methods for Elementary School Reading.   Sep. - Dec. 2019

PEABODY COLLEGE, VANDERBILT UNIVERSITY   Nashville, TN

Teaching Assistant – EDUC6530 Educational Linguistics and SLA   Oct. 2018 - Jan. 2019

ABSTRACT OF THE DISSERTATION

Language Use and Self-efficacy in Writing of Linguistically Diverse Students

by

Jiali Wang

Doctor of Philosophy in Education

University of California, Irvine, 2024

Professor Young-Suk Grace Kim, Chair

This dissertation outlines three studies investigating linguistically diverse students' (including bilingual and monolingual students) language use and self-efficacy in writing. Bilingual students are those who speak a language other than English, including students with various levels of English proficiency and language experience. The first two studies (Study 1 and Study 2) examine the relations of language use features to writing performance, while the third study examines which dimensions of writing self-efficacy are more crucial for educators to target. The Direct and Indirect Effects Model of Writing (DIEW; Kim & Graham, 2022) and the Writer(s)-Within-Community serve as the guiding theoretical frameworks for this dissertation. According to DIEW and WWC, many linguistic, socioemotional, and cognitive skills and knowledge are needed to write proficiently. The present dissertation focuses on language skills and self-efficacy in writing. Study 1 is a meta-analysis that aims to investigate the relations between syntactic features and writing performance systematically. Previous studies have inconsistent results regarding the relation between syntactic complexity and writing quality and the moderating roles of measurement (e.g., genre, syntactic complexity measures) and writers' (i.e., language proficiency, age) in the relation. The results show a weak relation between syntactic features and

writing performance, and the relation can be moderated by measurement and writers' characteristics. Study 2 examines the predictive accuracy of machine learning in predicting multiple dimensions of writing quality (structure, evidence use, historical thinking, and language use) in secondary history source-based argument writing. In addition, based on linguistic features that contribute to the prediction of writing quality, the study further examines language use patterns in bilingual and monolingual writers with different EL designations. Results indicated that machine learning models can explain a higher variance in writing quality scores. English Learners (EL) performed lower on lexical sophistication and global cohesion, but bilingual students designated as proficient in English performed equivalently with their monolingual peers. Using the same data as in Study 2, Study 3 examines secondary students' writing self-efficacy. The study examines (1) in which dimension(s) students feel less efficacious in writing, (2) which dimension(s) is more predictive of students' writing quality, and (3) to what extent students of different EL designations perform differently. Results showed that students had the lowest writing self-efficacy in self-regulation, which was the dimension that was significantly associated with writing quality. In addition, EL students and students designated as Reclassified Fluent English Proficient (RFEP) had lower writing self-efficacy in ideation and revision, whereas their writing self-efficacy in self-regulation was equivalent to students who were initially fluent in English since they entered school.

These findings suggest that the text-based linguistic features can provide information about students' language skills and areas of improvement in an authentic writing context. Students, in general, had lower writing self-efficacy in self-regulation, which is a critical skill to navigate through the writing process. Students who had the experience of being an EL may have

lower writing self-efficacy in certain aspects but had equivalent writing self-efficacy in self-regulation.

INTRODUCTION

Writing requires individuals to master many cognitive, linguistic, and literacy skills to express themselves appropriately across different settings and contexts (Graham, 2018; Kim & Graham, 2022; Kim et al., 2014; Wagner et al., 2011). Examining language use features in written compositions provides crucial insights for educators about the specific linguistic demands of writing. The measures and approaches to assessing syntactic features in writing have garnered considerable attention recently (e.g., Jagaiah, 2020; Kyle & Crossley, 2018; Troia et al., 2019; Wang et al., 2024). Linguistic features, including syntactic, lexical, and cohesion features, offer rich information about students' language and literacy skills and have been the subject of extensive study in recent literature (e.g., Crossley et al., 2019; Crossley & Kyle, 2018; Kyle, 2016; Kyle et al., 2021; Lu, 2017; McNamara et al., 2014).

Specifically, high-quality writing usually demands the flexible use of a range of syntactic features, from simple to complex, tailored to the writers' goal and audience (Beers & Nagy, 2009; Kim & Graham, 2022). In Study 1, we systematically and comprehensively examined the relationships between syntactic features in writing and writing performance—focusing on elements such as writing quality—and investigated how these relationships are moderated by specific syntactic complexity measures, text-based or not, writing outcomes, and writing genres, as well as writer characteristics like grade level and language proficiency. The relationship between syntactic accuracy in written composition and writing quality has not been widely examined, while more studies have explored the relationship between syntactic complexity and writing quality. However, many review articles, while extensive (Crossley, 2020; Crowhurst, 1983; Lu, 2017; Jagaiah et al., 2020), have not estimated the magnitude of these relationships or discussed moderating effects comprehensively. Our study aims to document a comprehensive

picture of the relationship between syntactic features and writing performance, examining both the magnitude of these relationships and the factors that may influence them. This understanding is crucial for educators to know which syntactic features are essential for writing proficiently across various genres.

Study 2 expands the examination to include a comprehensive list of linguistic features—such as lexical, syntactic, and cohesion elements—rather than focusing solely on syntactic features in argumentative writing within the context of history. Given the significant role of linguistic features in writing, this study aims to determine the extent to which these features can explain variances in writing quality outcomes. Previous studies have consistently shown that linguistic features can explain about 50%-60% of the variance in writing quality (e.g., Crossley & McNamara, 2011a, 2012; Crossley et al., 2014; Guo et al., 2013). In Study 2, we investigated secondary students' language features generated by natural language processing software in source-based argument writing and assessed how these features predict various dimensions of writing quality, such as structure, evidence use, historical thinking, and language use, using machine learning approaches. We also examined how these features differ for bilingual and monolingual secondary students of different levels of language proficiency, seeking to understand how linguistic challenges differ for various writing tasks and diverse writers.

Examining the relations between linguistic features and writing can generate insights regarding linguistic demands in writing and how linguistic challenges differ for various writing tasks and diverse writers. Writing models and empirical studies have suggested that a variety of measurement and writers features contribute to different relations between linguistic features and writing performance, including the specific linguistic measures examined (Kyle & Crossley, 2018), writing tasks (Kim & Graham, 2022), writing genre (Wang et al., 2024), writing outcomes

(Wagner et al., 2011), writers' developmental phases (Wagner et al., 2011), and language proficiency (Lu, 2017). Thus, study 1 and study 2 consider these measurements and writers' characteristics when examining linguistic features in writing and writing quality. In study 1, measurement and writers' characteristics are considered by examining the moderating effects of these factors. In study 2, We examined bilinguals with different levels of language proficiency, including those designated as having limited proficiency (EL) and those classified as proficient in English, including RFEP and Initially Fluent English Proficient (IFEP). There is a lack of research on the writing development of bilingual secondary students in the US context, and study 2 aims to fill this gap in the literature and examine linguistic features in writing for bilingual and monolingual students with various language proficiency levels.

Study 3 also endeavors to understand the writing development of linguistically diverse students. Study 3 addresses three critical gaps in the literature: (1) differences in dimensions of writing self-efficacy among secondary students, (2) the relation between dimensions of writing self-efficacy and dimensions of argumentative writing quality in history, and (3) differences in writing self-efficacy across linguistically diverse students with different English language designations in the US context. With an understanding of which aspects students are more or less confident in and which dimensions of writing self-efficacy are more crucial to students' writing performance, educators can target specific aspects of writing instruction. This understanding can inform writing interventions and instructions by highlighting the nuanced relationship between writing self-efficacy and writing quality.

Aligned with study 1 and study 2, study 3 also takes on a perspective to understand heterogeneity in linguistically diverse students. Students' language and literacy skills differ due to their different experiences in both school and home contexts (de Jong, 2004; Williams &

Lowrance-Faulhaber, 2018). Investigations of the writing skills of linguistically diverse students tend to focus on viewing them as a homogeneous group and use a strict cutoff to determine individual students' language proficiency as either proficient or not. In study 3, we distinguished between originally fluent students (i.e., IFEP) when they entered school and students designated as Reclassified Fluent English Proficient (i.e., RFEP) and ELs. By doing so, we recognized diverse linguistic experiences in student populations.

CHAPTER 1

A Meta-analysis of Relations of Syntactic Features in Written Composition and Writing Performance and How the Relations Vary by Writer and Measurement Features

Introduction

Writing requires an individual to master many cognitive, linguistic, and literacy skills to express themselves appropriately across different settings and contexts (Graham, 2018; Y.-S. G. Kim & Graham, 2022; Y.-S. G. Kim et al., 2014; Wagner et al., 2011). Examining language use features in student written composition can inform educators of the specific linguistic demands in writing. The measures and approach to assessing syntactic features in writing have also received much attention recently (e.g., Jagaiah, 2020; Kyle & Crossley, 2018; Troia et al., 2019; Wang et al., 2024).

Two commonly examined syntactic features in written composition are syntactic complexity and syntactic accuracy. Syntactic complexity, defined as the variety and degree of sophistication (complexity at the phrase and the clausal level) of the syntactic structures (Ortega, 2003), has been widely examined in student writing. Syntactic complexity in written composition reflects one aspect of writers' language use in writing, and it varies based on writing purpose, genre, topic, and audience. Writers employ sentence structures associated with communicative needs to fulfill the communicative demands of specific registers (Biber & Gray, 2011). Thus, research examining syntactic complexity can reveal the extent to which students use various grammatical structures in their writing and test the differential demands of various writing genres. Another syntactic feature examined in writing research is syntactic accuracy, which refers to the extent to which the text is free from grammatical errors. Whereas a plethora of research examined the

relation between syntactic complexity and writing quality, only a few studies have explicitly examined the relation of syntactic accuracy with writing quality (e.g., Troia et al., 2019; Wang et al., 2024). Syntactic accuracy is still an important element contributing to writing quality for monolingual writers and language learners (Troia et al., 2019). In this meta-analysis, we examined the relations between syntactic features in writing and writing performance (e.g., writing quality) systematically and comprehensively and investigated the extent to which the relations are moderated by measurement (specific syntactic complexity measures, text-based or not, writing outcomes, and writing genre) and writer characteristics (grade level and language proficiency).

Theoretical Framework

The role of language skills in writing has been recognized in theoretical models: in the cognitive model of writing (Hayes, 1996) and writer(s)-within-community model of writing (Graham, 2018), linguistic knowledge is in students' long-term memory and is retrieved during the writing process, especially when writers translate their ideas into sentences (Berninger et al., 2002). In the current study, we used the Direct and Indirect Effects model of Writing (DIEW; Y.-S. G. Kim & Graham, 2022; Y.-S. G. Kim & Park, 2019) as a primary theoretical framework because it explicitly recognizes the role of various dimensions of language skills and discourse knowledge in writing and posits differential relations (moderation—see below). Another key reason we selected this model as the theoretical framework is that the model proposes dynamic relations between component skills (e.g., language skills) and writing outcomes by different factors, such as development phases and measurement. As a component of language skills, grammatical knowledge allows students to express their ideas accurately, effectively, and richly (Y.-S. G. Kim & Graham, 2022). High-quality writing usually requires flexible use of a range of

6

syntactic features from simple to complex, depending on the goal and audience of writing (Beers & Nagy, 2009; Y.-S. G. Kim & Graham, 2022). Low proficiency in constructing complex sentences may inhibit students from translating thoughts into sentences, and students equipped with more grammatical knowledge can allocate more cognitive resources for other crucial parts of the writing process (e.g., ideation; Graham, 2018). This may lead to improvement in the overall quality of the writing.

Discourse knowledge is also crucial for writing quality because different genres demand different text structures and language features (Y.-S. G. Kim & Graham, 2022). Different genres have various demands and expectations on writers, and writers acquire these expectations as they learn to write. Certain syntactic features are helpful and functional for specific purposes in certain genres (Biber & Gray, 2011; Schleppegrell, 1998). Thus, to navigate writing in a particular genre, writers need knowledge of text structure and language features associated with the genre (Y.-S. G. Kim & Graham, 2022).

Syntactic Features and Writing Quality

The relation between syntactic accuracy in written composition and writing quality is not widely examined. Studies conducted with students from upper elementary grade levels found that syntactic accuracy, operationalized as the percentage of grammatical sentences, predicted narrative writing quality (Troia et al., 2019; Wang et al., 2024). An analysis of TOEFL essays written by EFL learners found that students who scored higher on holistic scores also scored higher on syntactic accuracy (Cumming et al., 2005). No genre or task differences have been found (Cumming et al., 2005; Scott & Windsor, 2000).

More studies have examined the relation between syntactic complexity in written composition and writing quality. Research on syntactic complexity was initially conducted to understand students' (especially students with limited language proficiency) grammatical knowledge. T-unit length, the most commonly examined syntactic complexity measure, was found to have high positive correlations with grammatical knowledge (Loban, 1976). Thus, young monolingual writers or bilingual writers with limited language proficiency are the main targets of studies examining the relation between syntactic complexity and writing quality. Such studies have the potential to identify students' language needs more efficiently. However, previous studies failed to establish a clear relation between syntactic complexity and writing quality for two reasons. The first reason is the inconsistent results across studies: only some studies have found a positive relation between syntactic complexity and writing quality (e.g., Beers & Nagy, 2009; Grobe, 1981; Ortega, 2003; Stewart & Grobe, 1979). Secondly, the relation also varies by other factors, such as grade level, language proficiency status, the type of syntactic complexity measures, and writing genres. In the sections below, we introduce potential moderators related to student and measurement characteristics that may impact the relation between syntactic features and writing outcomes.

Student Characteristics

Grade Level as a Proxy for Developmental Phase

As foundational language acquisition occurs throughout childhood, students' ability to formulate sentences grows (Ortega, 2003). Thus, grade level or age is a factor that impacts students' syntactic complexity in written compositions. In the US context, studies found that students of higher grade levels have more complex syntax in their writing (Beers & Nagy, 2011; Crowhurst,1980; Crowhurst & Piche, 1979; Crossley et al., 2011; Ferris, 1994; Wagner et al.,

2011). As students advance to higher grade levels, children write longer T-units, produce a greater number of clauses, and increase the variety of clause types (e.g., Beers & Nagy, 2011; Crossley et al., 2011; Ferris, 1994; Wagner et al., 2011; Loban, 1976). The pattern is also consistent for adult writers (e.g., MacArthur et al., 2019).

Because of the development of students' sentence construction skills and language use, the relation between syntactic features and writing quality is likely to vary as a function of developmental phases. Novice young writers are more likely to be confined by insufficient syntactic skills to express their ideas. A positive relation between syntactic complexity and writing quality was found among younger writers. A study (Steward & Grobe, 1979) examined three syntactic complexity measures in expository writing with students in Grades 5, 8, and 11. Results indicated a positive relation only for fifth graders, no matter which syntactic complexity measures were used. While several studies found that the number of words before main verbs was an indicator of better writing (McNamara et al., 2010; Crossley et al., 2011; McNamara et al., 2013), some other studies did not find a relation between syntactic complexity and writing quality (Perin & Lauterbach, 2018).

DIEW hypothesizes differential relations between language skills and writing quality based on development stages (dynamic relations as a function of development; Y.-S. G. Kim & Graham, 2022). As transcription skills improve and become more automatic, other skills and knowledge, including syntactic complexity and accuracy, can influence writing quality to a greater extent. It is important to note, however, that this does not imply that syntactic skills in earlier phases of writing development do not impact writing quality. Instead, it suggests that their role is significantly constrained by transcription skills, such that differences in syntactic skills in oral language contexts are not fully reflected in written composition in the beginning phase of

writing development. The current study focuses on syntactic features in written compositions, where transcription skills are already taken into account (students' transcription skills already have an impact on their language use and overall writing outcomes). Therefore, the relation between syntactic features and writing quality is not expected to be weaker in earlier phases. In fact, this relation might be stronger during earlier stages of development, as there is likely to be greater variation in syntactic features in written composition during an earlier phase as children rapidly develop syntactic skills.

Language Proficiency

The relation between language skills and writing ability has been indicated to be prominent with students with limited language proficiency (Lu, 2017). According to the DIEW model, language skills are imperative to write high-quality essays (Y.-S. G. Kim & Graham, 2022). Therefore, language proficiency has an impact on students' writing performance, where the writing performance of students with limited language proficiency may depend more on their language proficiency and the extent to which they are able to utilize their language skills to communicate their ideas. As a part of language skills in writers, the ability to use syntactic structures flexibly and appropriately may thus be more crucial in writers with limited language proficiency to achieve high writing performance. More consistent results were reported regarding the relation between syntactic complexity and writing quality in writing in the EFL context than in an ESL or monolingual-dominant context (Lu, 2017). For learners with limited language proficiency, it was found that better essays had longer and more varied syntactic structures (Li, 2015; Lu, 2011; Ortega, 2003), greater clausal subordination (Biber et al., 2016; Grant & Ginther, 2000; Li, 2015), more complex clause structure (Crossley & McNamara, 2014; Li, 2015), more complex phrasal structure (Crossley & McNamara, 2014; Guo et al., 2013; Kyle &

Crossley, 2018; Taguchi et al., 2013), and more incidence of passive structures (Biber et al., 2016; Ferris, 1994). These studies used a variety of measures for phrasal, clausal, T-unit, and sentence-level syntactic complexity measures generated by a variety of software. However, it should be noted that most of these studies were conducted with adult learners; few have been done with younger learners.

Measurement Features

When examining relations between skills, it is essential to consider the role of measurement and whether the relations may differ as a function of measurement (see dynamic relations as a function of measurement; Y.-S. G. Kim & Graham, 2022). Measurement in this study refers to the measurement of syntactic complexity (i.e., the type of syntactic complexity) and writing performance.

Syntactic Complexity Measures

Researchers have examined numerous syntactic measures and sought ways to investigate the nature of the relation between syntactic complexity and writing, which makes comparing results across various measures challenging. The most commonly used traditional syntactic complexity measures are T-unit length, clause length, and the number of clauses per T-unit. However, since these measures examine syntactic complexity at relatively larger grain sizes (at T-unit or clause level), they cannot identify specific syntactic complexity features of texts, and interpretation of these measures is challenging (Norris & Ortega, 2009). Common use of these traditional measures can be one of the reasons for the mixed findings in the literature examining the relation between syntactic complexity and writing quality. Kyle and Crossley (2018) used a variety of syntactic complexity indices at the phrase, clause, and sentence levels. They found that

11

the phrase and clause level measures predicted more variance in holistic writing quality scores in TOEFL independent argumentative essays than a traditional measure (words per clause). Kyle and Crossley (2018) suggest that more fine-grained measures may be able to explain more variance in writing quality and may have a stronger relation to writing quality. Given the potential limitations of traditional syntactic complexity measures, studies in second language writing started to examine syntactic complexity at clause and phrase levels (e.g., Biber et al., 2016; Crossley & McNamara, 2014). In addition, since distinct syntactic complexity measures tap into different aspects of syntactic complexity, multiple studies (Beers & Nagy, 2009, 2011) indicate that different syntactic complexity measures perform differentially in the same writing prompt. Thus, the relation between syntactic complexity measures and writing quality may also differ. Therefore, it is necessary to examine the impacts of different measures on the relation between syntactic complexity and writing quality.

Writing Task

Measurement of construct also includes the nature of writing tasks (Y.-S. G. Kim & Graham, 2022). For example, when writing tasks include reading source materials, reading comprehension skills are particularly important to the writing process and outcomes. Language demands in the text-based and non-text-based tasks may differ in the following aspects: (1) text-based tasks require writers to summarize, manipulate, and reorganize the content from sources, whereas non-text-based tasks require writers to generate ideas and express them using their own language; (2) while text-based tasks demand both comprehension and production, non-text-based tasks only demand production (Cumming et al., 2005; M. Kim & Crossley, 2018). Because of differential linguistic demands in the two types of tasks, it is possible the relation between syntactic features and writing outcomes may differ.

12

Syntactic features in source-based writing have not been widely examined as writing that is not source-based (Cumming et al., 2016) despite research indicating that different linguistic features predict text-based and non-text-based writing (Guo et al., 2013; M. Kim & Crossley, 2018). Guo and colleagues (2013), using TOEFL writing tasks, found that noun phrase complexity significantly related to writing quality in both source-based and non-source-based tasks, whereas subordination was negatively correlated with writing quality only in the non-text-based writing task. In another study using TOEFL writing tasks that compared text-based and non-text-based writing tasks (M. Kim & Crossley, 2018), words per clause was found to be significantly related to both tasks, words per sentence and words per T-unit were significantly related to only the text-based task, complex nominals per clause and coordinate phrases per clause were significantly related to only the non-text-based. However, due to the limited number of studies conducted with source-based writing tasks, it is challenging to conclude whether syntactic features may contribute to these two types of writing tasks differently.

Writing Genres

Empirical evidence suggests that the demands of syntactic features vary across genres (e.g., Beers & Nagy, 2009, 2011; Olinghouse & Wilson, 2013; Scott & Windsor, 2000). A literature review found that the argumentative genre had the highest values for T-unit length and clauses per T-unit across the nine studies examined (Jagaiah et al., 2020). Studies have also found that narrative writing has the least demands on syntactic skills (e.g., Crowhurst & Piche, 1979), whereas argumentative writing has greater needs for writers to employ more complex syntactic structures because argumentation requires writers to convey meaning efficiently by tying similar or competing ideas closely together in a single T-unit (Jagaiah, 2020). For example, Beers and Nagy (2011) found that argumentative essays had more clauses per T-unit than other

genres (narrative, compare/ contrast, and descriptive). They also found that descriptive texts had more words per clause than argumentative and narrative texts (though only in grade seven; Beers & Nagy, 2011). Similarly, Scott and Windsor (2000) found that expository writing had higher syntactic complexity (measured by T-unit length) than narrative writing for 9- to 11-year-olds. Taken together, extant evidence suggests that the informational genre tends to have higher demands on syntactic complexity than the narrative genre, especially when syntactic complexity was measured by T-unit length and clauses per T-unit.

Studies have shown that syntactic features in written composition predict writing quality differently across genres. Qin and Uccelli (2016) found that words per clause was significantly correlated with writing quality in argumentative writing but not narrative writing for Chinese secondary EFL learners. Grobe (1981) used the same three measures (T-unit length, words per clause, clauses per T-unit) in narrative writing with students in Grades 5, 8, and 11 and did not find a significant relation between syntactic complexity and narrative writing quality. Stewart and Grobe (1979), who worked with students in the same three grade levels but examined expository writing, found a positive relation between syntactic complexity in all three measures only in Grade 5. Crowhurst (1980) examined the relation between T-unit length and writing quality in Grades 6, 10, and 12 in narrative and argumentative compositions. It was found that while written compositions with longer T-units had significantly higher writing scores in the argumentative task in Grades 10 and 12, compositions with longer T-units did not have higher scores in the narrative task. Similarly, Beers and Nagy (2009), in a study where they examined three different measures of syntactic complexity (T-unit length, words per clause, clauses per T-unit) in narrative and argumentative genres for seventh and eighth-grade students, found the same pattern. However, they found that T-unit length was positively correlated with writing

quality in the narrative task but was negatively correlated with writing quality in the argumentative task. The authors pointed out that there is awkward and repetitive use of subordinate clauses, such as the sentence structure of "I think X because Y," where the use of subordinations does not necessarily allow students to elaborate on the content or argue effectively. In summary, the literature suggests that the relation between syntactic complexity and writing varies by genre, although results are inconsistent across studies.

Writing Outcomes

Previous studies have proved that writing quality, writing productivity or text length, and writing fluency are related but dissociable skills (Y.-S. G. Kim et al., 2014; Puranik et al., 2008; Wagner et al., 2011) and the contributions of language and literacy skills differ depending on writing outcomes (see the dynamic relations as a function of measurement of DIEW, Y.-S. G. Kim & Graham, 2022). Specifically, language skills play a greater role in writing quality than writing productivity, whereas transcription skills are more important to writing productivity (Y.-S. G. Kim & Graham, 2022). By examining how certain syntactic features relate to writing outcomes, we can better understand how syntactic demands may limit various aspects of writing performance.

The Current Study

Using a meta-analysis approach, the current study aims to examine the relations between syntactic features and writing quality and potential moderation by writers and measurement factors. While review articles exist (Crossley, 2020; Crowhurst, 1983; Lu, 2017; Jagaiah et al., 2020), these articles did not estimate the magnitude of the relations between syntactic features and writing outcomes and have only discussed moderating effects to a limited extent. In order to

document a comprehensive picture of the relation between syntactic features and writing performance, we examine the magnitude of the relations and moderators that may impact the relation. Understanding which syntactic features are critical to various writing tasks for various populations can improve educators' understanding of necessary language skills to write proficiently in various genres. The following research questions guided the present study:

1. What is the magnitude of the relation between syntactic complexity and accuracy and writing performance?

2. Do the relations between syntactic complexity (including words per unit, subordination, noun phrase complexity, left embeddedness, and words per sentence) and accuracy and writing vary as a function of writers' characteristics (grade level, language proficiency status) and measurement (the type of syntactic complexity measures, text-based or not, writing outcomes, and writing genre)?

Methods

Search Procedures and Inclusion Criteria

The electronic databases to conduct the main literature search included Educational Resources Information Center (ERIC), APA PsycInfo, Linguistics and Language Behavior Abstracts (LLBA), Dissertations & Theses Global, ProQuest Dissertations & Theses A & I, and Sociological Abstracts, through ProQuest electronic databases. The search date was limited to January 1, 1960 to December 31, 2021 because studies investigating syntactic features in writing emerged in the 1960s (Crowhurst, 1983). There was no restriction on the age of the population. The Boolean search terms were as follows: "ab(("gramma* complex*" OR "complex* gramma*" OR "synta* complex*" OR "complex* synta*" OR "MLU" OR "text complex*" OR "sentence complex*" OR "t-unit*" OR "synta* density" OR "claus* density" OR "synta* accur*" OR

"gramma* accur*")) AND ab((writ*))”. The initial search yielded 2,951 articles. There were no restrictions on publication type.

Inclusion criteria were as follows: (a) both syntactic features and writing quality were measured; (b) participants of various levels of language proficiency and participants with disabilities (e.g., ADHD, developmental language disorder, dyslexia) were included, but not participants identified with severe disabilities and sensory impairments (e.g., participants with traumatic brain injuries, any form of aphasia, or Down syndrome); (c) sample sizes were no fewer than four; (d) the study reported adequate information to calculate effect sizes (i.e., zero-order correlations) for the relation between syntactic features and writing performance (if not, the primary author was contacted); (e) the study was published in English; and (f) the collected data were not impacted by writing intervention; if a writing intervention was present, only pretest or control group data would be included.

Study Selection and Exclusion

As illustrated in the PRISMA chart (Figure 1.1), the database search yielded 2951 studies, and 772 duplicates were deleted. Rayyan, an online meta-analysis tool, was used for blind screening and establishing inter-rater reliability. During title/abstract screening, 622 additional documents out of 2951 articles were excluded because they (1) did not include any writing ($n = 283$); (2) were methodological or theoretical ($n = 109$); (3) did not measure syntactic features ($n = 56$); (4) were of wrong publication type, such as book reviews, letters to the editor ($n = 51$); (5) used a public corpus without identifying information about participants ($n = 35$); (6) were qualitative studies ($n = 25$); (7) were review studies ($n = 31$); (8) had fewer or equal to four participants ($n = 17$); (9) used a population with other more specific needs/conditions ($n = 15$). An interrater agreement of 95% was achieved in the title/abstract

screening stage using 4% of all articles ($n = 120$). Discrepancies were resolved through discussion before coders began independent coding.

Figure 1.1

PRISMA Flow Diagram Showing the Searching and Screening Processes

Identification of new studies via databases

Identification

Records identified from ERIC, APA PsychInfo, LLBA, Dissertations & Theses Global, ProQuest Dissertations & Theses A& I, and Sociological Abstracts (n = 2951)

Records removed *before screening*: Duplicate records removed (n = 772)

Screening

Records screened (n = 2179)

Records excluded (n = 622):
No writing (n = 283)
Methodological /theoretical (n = 109)
No syntactic features (n = 56)
Wrong publication type (n = 51)
Unclear corpus studies (n = 35)
Qualitative studies (n = 25)
Review studies (n = 31)
Insufficient participants (n = 17)
Extreme population (n = 15)

Reports sought for retrieval (n = 1557)

Reports not retrieved (n = 66)

Reports assessed for eligibility (n = 1491)

Reports excluded (n=1406):
No Pearson's correlation (n = 676)
Construct not assessed (n = 582)
Duplicate (n = 71)
No original data (n = 48)
Insufficient participants (n = 12)
Reanalysed data (n = 6)
Unclear assessments/ aggregate scores (n = 6)
Extreme population (n = 4)
Intervention with no pre-test or control (n = 1)

Included

Primary studies included in final sample via screening (n = 96)

Additional search methods:
Backward search (n = 3)
Contacting authors (n = 7)
Journal hand search (n = 15)

Total studies included in review (n = 121)

Studies included during the title/abstract screening were screened by reading the full texts to examine further whether inclusion criteria were met. Interrater reliability for full article

screening was 95.45% on 10% of the included articles ($n = 154$). Correlation values were calculated for articles that reported raw scores ($n = 7$). In addition, for all articles published within the past ten years (starting from 2012), authors of the articles that clearly measured our target skills (syntactic features and writing outcomes) but did not report correlations were contacted for information on Pearson correlations.

Beyond electronic database searches, we also searched for articles from other sources. First, the reference lists of literature review articles on the topic were checked (Crossley, 2020; Crowhurst, 1983; Jagaiah et al., 2020; Lu, 2017; Ortega, 2003). We also manually searched three journals likely to publish research articles on the topic, including *Journal of Second Language Writing, Reading & Writing, and Journal of Speech, Language, and Hearing Research*. Total number of studies included through database search, reference chasing, and journal search was 121.

Coding Procedures

All studies included were coded for effect sizes, sample size, and sample characteristics, including age, language proficiency status (language learning context, language proficiency), socio-economic status, gender, race, primary language, and disability status. We also coded types of syntactic complexity measures, writing genres, types of writing outcomes, language of the writing samples, and reliability information of the syntactic and writing performance measures. Two PhD students with a background in linguistics double-coded all studies. Interrater reliability (exact agreement) using 20% of the studies ($n = 24$) was 90.8%. Discrepancies were discussed and resolved.

For the analysis, measures reported in fewer than three studies (e.g., relative clauses/ T-unit, adverbial clauses/ T-unit) were excluded. In addition, measures operationalized as a mixture

of different syntactic complexity features were also excluded. In total, twelve studies were excluded after coding, which led to a total number of 110 studies for data analysis. Effect sizes of the number of words per clause and the number of words per T-unit were combined because both measured length and their effect sizes were very similar. This category was referred to as *words per unit* hereafter.

Student Features

Students' grade level was operationalized in two approaches: as a categorical variable and a continuous variable. We used a categorical variable because the moderating effect of grade level may not be linear. Grades were divided into five categories: primary grades (Grades K-2), upper elementary (Grades 3-5), middle school (Grades 6-8), high school (Grades 9-12), and adults (undergraduate level and beyond). We used a continuous variable because some categories may have a small number of subsamples. For the continuous variable, the weighted grade level of the sample was used.

Language proficiency was operationalized in two ways. First, the language learning context of the sample was recorded. Foreign language (FL) context is when learners included write in a foreign context where the target language in the writing task is not the official language of the country. Second language (SL) context refers to writers in a context where the target language is the official language, but some students were identified as limited in English proficiency. Contexts with monolingual and bilingual students were coded as mixed, and contexts with only monolingual speakers were coded as monolingual.

Measurement Features

Multiple measurement features were examined, including the type of syntactic complexity measures, text-based or not, writing genres, and writing outcomes (see Appendix A).

Types of syntactic complexity measures were coded. Subordination measures included dependent clauses per clause, dependent clauses per T-unit, number of clauses per T-unit, subordinating conjunctions, and percentage of sentences containing embedded clauses. Noun phrase complexity included word per noun phrase, complex nominal phrase/T-unit, complex nominal phrase/clause, incidence of prepositions, and number of modifiers per noun phrase.

Writing genres were coded into the following categories: narrative, argumentative, expository, compare and contrast, problem solution, informational (did not specify which subgenre or uses multiple different subgenres of informational tasks), mixed, and others (e.g., poem, diary). Due to limited effect sizes, compare and contrast, and problem-solution were merged with the informational category.

Writing outcomes were coded into writing quality, productivity, and fluency. Writing quality includes either holistic or analytic scoring (commonly examined analytic aspects include content, organization, coherence, and language use; writing conventions, such as punctuation and capitalization, were not counted toward writing quality). Writing productivity measures included measures of different grain sizes, such as the number of clauses, T-units, and sentences. Writing fluency refers to productivity during a certain amount of time. Writing productivity measures with a time limit of five minutes or less were coded as writing fluency.

Data Analysis Strategies

The first and second research questions were analyzed using Fisher's z scores based on the study's sample size and reported correlation values. This was to ensure standardization of the

scores so that no sample was over- or underrepresented. We also calculated sample variance and used this instead of sample size for the calculations. Calculations were run using R with the packages *metafor* (Viechtbauer, 2010) and *robumeta* (Hedges et al., 2010). We used *robumeta* because it is an estimator of Robust Variation Estimation. This estimator considers the sample size of each study when calculating the weighted effect sizes. Unlike traditional techniques that assume effect sizes are independent of each other, the *robumeta* package accounts for the nested nature of data (Hedges et al., 2010). Because many studies examined multiple syntactic complexity measures, we used the *robumeta* package, which can deal with both types of dependency in data. We ran the overall correlation of the study in both of *metafor* and *robumeta* to ensure the reliability of the calculated values. Because studies with statistically significant results are more likely to be published, we also ran publication bias tests using funnel plots and Egger's regression test (Sterne & Egger, 2005).

To answer the first research question, we estimated the weighted effect sizes of the included studies for each syntactic measure. If a study provided more than one measure of the syntactic features, a mean value was calculated using the *robumeta* package. For the second research question, meta-regressions were run where moderators (grade level, language proficiency status, syntactic complexity measures, text-based or not, writing genre, and writing outcomes) were tested. Results were only reported for categories with sufficient effect sizes ($df >$ 4). Because we cannot assume whether moderation effects are the same across syntactic features, we ran separate moderation effect models for each syntactic complexity feature examined (excluding left embeddedness due to a small number of subsamples it appears; $k = 7$; see Appendix B for a list of articles).

Results

Characteristics of Included Studies

This meta-analysis included 109 studies with 871 effect sizes clustered within 162 subgroups/subsamples and a total number of participants of 24,628 (sample sizes range from 5 to 2,916). Included studies spanned from the years 1977 to 2022. Out of the 162 subgroups, 81 were from journal articles, 55 from dissertation studies, nine from research reports or conference presentations, 11 from book chapters, and six from other studies. The majority of the studies were conducted in the US ($n = 68$), followed by China ($n = 7$), Japan ($n = 5$), and UK ($n = 3$). One-hundred-and-one subgroups reported students' gender information (students were balanced in gender), and 147 subgroups reported students' grade levels. Only 65 studies described students' academic ability (e.g., whether students' academic performance is low, average, high, or mixed). Only seven subgroups provided specific information about the number of students with learning disabilities; 25 subgroups excluded students with learning disabilities. Twenty-nine subgroups reported the overall socioeconomic (SES) status of their sample: 9 of them were from middle SES background, six from medium and low SES, five from medium and high SES background, five from a mixed SES background, and four from low SES background.

Among the 162 subsamples, 90 samples reported language learning context: 37 were in a Foreign Language context, 14 were in a Second Language context, 19 were in a mixed context, and 20 were in monolingual contexts. Regarding students' language proficiency, only 84 subsamples provided information about students' language proficiency. Of these, 44 predominantly consisted of students with limited language proficiency.

Among the 156 subsamples that mentioned the specific format of their prompts, 26 were text-based and provided text(s) that students need to read before proceeding to the writing task. Regarding the genre of writing tasks, 38 subsamples used argumentative tasks, 37 used narrative

24

tasks, 32 used expository tasks, 15 reported informational tasks but did not specify which subgenre was used, 15 used a mixture of multiple types of writing tasks, 6 used contrast and compare, 5 used other genres, one used problem and solution task, and 16 did not report which genre was used. Of the 147 subgroups that provided information on writing prompts, only 22 used norm-referenced writing tasks. Only 90 out of 162 subsamples reported reliability of writing outcomes, while only 49 reported reliability information of syntactic features. Most subsamples' ($n = 147$; 90.7%) writing tasks were in English. Only 21 subsamples specified that digital writing format was used. Of the 38 studies that provided information about whether spelling was corrected, eight corrected spelling before generating writing outcome measures.

Research Question 1: Magnitude of Relations Between Syntactic Features and Writing Outcomes

Magnitudes of zero-order correlation for each syntactic measure were reported in Table 1.1. Syntactic accuracy ($b = .25, p < .001$) and syntactic complexity ($b = .16, p < .001$) across different measures had a weak but significant relation to writing outcomes. As for individual syntactic complexity measures, except for left embeddedness, all syntactic complexity measures had a weak but significant relation to writing outcomes: noun phrase complexity ($b = .24, p < .001$), words per unit ($b = .19, p < .001$), subordination ($b = .14, p < .001$), and words per sentence ($b = .10, p = .02$).

Table 1.1

The Relations of Various Syntactic Features With Writing Outcomes

Variable	b	SE	CI.LB	CI.UB	p
Syntactic accuracy	.25	.07	.11	.40	.001**

Syntactic complexity	.16	.02	.13	.20	<.001***
Noun phrase complexity	.24	.04	.15	.33	<.001***
Words per unit	.19	.03	.13	.24	<.001***
Subordination	.14	.03	.07	.20	<.001***
Words per sentence	.10	.04	.02	.18	.02*
Left embeddedness	.10	.04	-.02	.22	.08

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Research Question 2: Factors that Moderate the Relations Between Syntactic Features and Writing Outcomes

Moderation Effects for Syntactic Accuracy

Characteristics of Writers. When grade level was operationalized as a categorical variable, there was a statistically significant stronger correlation between syntactic accuracy and writing performance among primary grade students ($r = .35$; $p = .03$ (.17 plus .18—see Table 2)) compared to adults ($r = .18$, $p = .03$; see intercept in the first panel of Table 1.2). For other grade levels, the relation between syntactic accuracy and writing outcomes was not different from that of adult writers. However, when grade level was operationalized as a continuous variable, there was no difference in the strength of the relation. In addition, the relation between syntactic accuracy and writing outcome was much weaker for contexts with only monolingual speakers ($r = -.03$, $p = .03$) compared to FL contexts ($r = .43$; $p = .03$). The moderation effect of language proficiency was not statistically significant.

Characteristics of Measurement. Writing genres and different types of writing outcomes did not moderate the relation between syntactic accuracy and writing outcomes.

Table 1.2

Multilevel Random Effects Model: Meta-regression of Moderators of Syntactic Accuracy

Moderator	k	Intercept (SE)	Estimate (SE)	I^2
Grade level (categorical)	145			90.59
Adults		.18***(.07)		
High school			.20(.17)	
Middle school			.07(.18)	
Upper elementary grades			-.11(.12)	
Primary grades			.17(.07)*	
Grade level (continuous)	136	.16(.13)	.01(.01)	91.14
Language learning context	77			91.17
Foreign language		.43*(.13)		
Second language			.15(.22)	
Monolingual			-.40(.16)*	
Language proficiency	78			93.61
Proficient		.26(.18)		
Limited			.11(.22)	
Measurement features: writing genres	84			88.86
Narrative		.19(.11)		
Expository			.01(.13)	
Measurement features: writing outcomes	145			89.06
Quality		.30(.08)***		
Productivity			.23(.58)	
Fluency			-.17(.10)	

Note. Categories were not included if df < 4; * *p* < .05. ** *p* < .01. *** *p* < .001.

Moderation Effects for Noun Phrase Complexity

Due to limited effect sizes, fewer moderation effects were examined for the relation between noun phrase complexity and writing outcomes (see Table). Among the moderators examined, all moderators significantly moderated this relation. This could be due to smaller variances generated by a smaller sample of studies because not many studies reported effect sizes of noun phrase complexity (most likely because such measures were not available in early years).

Characteristics of Writers. Grade level (measured by a continuous variable) was not a significant moderator. The relation was stronger for samples in which the majority of writers were of limited language proficiency compared to samples in which writers were predominantly monolingual speakers ($r = .26$, $p = .005$).

Characteristics of Measurement. Genre was a significant moderator: the relation between noun phrase complexity and writing outcomes was stronger for expository texts compared to narrative texts ($r = .43$, $p = .04$) The relation between syntactic complexity and writing performance was also moderated by how writing outcome was measured: the relation between noun phrase complexity and writing performance was weaker when writing outcome was measured by productivity ($r = .11$, $p = .03$), compared to when writing outcome was measured by writing quality.

Table 1.3

Multilevel Random Effects Model: Meta-regression of Moderators of Noun Phrase Complexity

Moderator	k	Intercept (SE)	Estimate (SE)	I^2
Grade level (continuous)	43	-.20(.29)	.04(.02)	79.20
Language proficiency	43			66.55
Proficient		-.01(.05)		

	k	Intercept (SE)	Estimate (SE)	I^2
Limited			.27(.06)**	
Measurement features: writing genres	39			73.43
Narrative		.10(.08)		
Argumentative			.05(.10)	
Expository			.33(.13)*	
Measurement features: writing outcomes	47			78.18
Quality		.29(.06)***		
Productivity			-.18(.08)*	

Note. Categories were not included if df < 4;* $p < .05.$ ** $p < .01.$ *** $p < .001.$

Moderation Effects for Words per Unit

None of the moderators examined moderated the relation between words per unit and writing outcomes (see Table 1.4).

Table 1.4

Multilevel Random Effects Model: Meta-regression of Moderators of Words per Unit

Moderator	k	Intercept (SE)	Estimate (SE)	I^2
Grade level (categorical)	330			84.68
Adults		.11(.04)***		
High school			.15(.07)	
Middle school			.07(.06)	
Upper elementary grades			.08(.07)	
Grade level (continuous)	345	.17(.02)***	.00(.00)	87.39
Language learning context	225			83.39
Foreign language		.16(.04)		
Second language			.05(.12)	
Monolingual			-.07(.06)	
Mixed			.09(.07)	
Language proficiency	216			87.39

Proficient		.20(.05)		
Limited			-.01(.06)	
Measurement features: writing genres	279			87.94
Narrative		.17(.04)***		
Argumentative			-.11(.05)	
Expository			.11(.08)	
Informational			-.12(.07)	
Measurement features: writing outcomes	120			88.58
Quality		.21(.03)***		
Productivity			-.07(.05)	
Measurement features: text-based	330			88.26
Not text-based		.15(.02)***		
Text-based			. 14(.08)	

Note. Categories were not included if df < 4; * $p < .05$. ** $p < .01$. *** $p < .001$.

Moderation Effects for Words per Sentence

None of the moderators examined moderated the relation between words per sentence and writing outcomes (see Table 1.5).

Table 1.5

Multilevel Random Effects Model: Meta-regression of Moderators of Words per Sentence

Moderator	k	Intercept (SE)	Estimate (SE)	I^2
Grade level (categorical)	52			76.00
Adults		.14(.04)**		
High school			-.10(.08)	
Grade level (continuous)	71	.08(.04)*	.00(.00)	90.03
Language learning context	44			66.70
Foreign language		.16(.04)**		
Second language			-.09(.09)	

Language proficiency	42		89.08
Proficient		.19(.15)	
Limited			-.04(.15)
Measurement features: writing genres	59		95.22
Narrative		.10(.09)	
Argumentative			.03(.12)
Expository			.05(.10)
Informational			-.06(.12)
Measurement features: writing outcomes	63		92.64
Quality		.14(.04)**	
Productivity			-.07(.07)
Measurement features: text-based	69		94.64
Not text-based		.10(.04)*	
Text-based			.15(.06)

Note. Categories were not included if df < 4; * $p < .05$. ** $p < .01$. *** $p < .001$.

Moderation Effects for Subordination

Characteristics of Writers. The relation between subordination and writing outcomes was weaker in the context with a mixture of monolingual and bilingual students compared to the FL context ($r = -.09$, $p = .03$; see Table 1.6).

Characteristics of Measurement. The nature of the writing task was a significant moderator: when the writing task included reading source materials, the relation between subordination and writing outcomes was stronger ($r = .24$, $p = .03$).

Table 1.6

Multilevel Random Effects Model: Meta-regression of Moderators of Subordination

Moderator	k	Intercept (SE)	Estimate (SE)	I^2
Grade level (categorical)	211			85.90

Adults		.08(.03)*		
High school			.09(.11)	
Middle school			-.02(.06)	
Upper elementary grades			.06(.07)	
Grade level (continuous)	220	.15(.06)*	.00(.01)	85.74
Language learning context	143			86.51
Foreign language		.14(.07)		
Second language			-.02(.13)	
Monolingual			-.06(.08)	
Mixed			-.23(.08)*	
Language proficiency	153			88.39
Proficient		.12(.05)*		
Limited			.03(.08)	
Measurement features: writing genres	189			87.50
Narrative		.08(.05)		
Argumentative			-.05(.08)	
Expository			.03(.07)	
Informational			.10(.12)	
Measurement features: writing outcomes	232			82.71
Quality		.09(.03)**		
Productivity			.04(.04)	
Measurement features: text-based	212			86.25
Not text-based		.08(.03)		
Text-based			.16(.06)*	

Note. Categories were not included if df < 4; * $p < .05$. ** $p < .01$. *** $p < .001$.

Syntactic Complexity Measures

Moderation effects of syntactic complexity measures were reported in Table 1.7. The relations between syntactic complexity and writing outcomes differed for words per sentence,

whereas no difference was found across other measures. The relation between words per sentence and writing outcome ($r = .09$, $p = .03$). was significantly lower than between words per unit.

Table 1.7

Multilevel Random Effects Model: Meta-regression of Moderators Controlling for Syntactic Complexity Measures

Variable	Model 1	Model 2	Model 3	Model 4
Words per unit (Intercept)	.19(.03)***	.19(.03)***	.19(.03)***	.19(.03)***
Subordination	-.04(.05)	-.05(.05)	-.06(.05)	-.06(.05)
Noun phrase complexity	.05(.05)	.05(.05)	.05(.05)	
Left embeddedness	-.09(.05)	-.10(.06)		
Words per sentence	-.10(.04)*			

Sensitivity Analysis

Given that the majority of studies' writing tasks were in English, we do not have sufficient effect sizes to test the language of writing tasks as a moderator. However, we conducted a sensitivity analysis to understand whether the relation between syntactic features and writing outcomes may differ by language. Interpretation of this sensitivity analysis should be taken cautiously because we classified all languages other than English as one category, despite differential syntactic features and functions for different languages. In addition, since most of the moderation effects with words per unit, subordination, and words per sentence were not significant, there may be excess variability in effect sizes. Thus, we conducted sensitivity analyses based on study characteristics (publication year, reporting of reliability, and publication type) for key moderators (grade level, language proficiency status, writing genres, and writing outcomes) to understand potential sources of variances that may have masked significant

moderators. Note that all sensitivity analyses were confined to words per unit and subordination measures due to the limited effect sizes of the other syntactic measures.

Language of the Writing Task

For words per unit, the overall weighted effect sizes estimations with ($r = .19$, $SE = .03$, 95%CI = [.13, .24], $p < .001$) and without subgroups where writing tasks were not in English ($r = .19$, $SE = .03$, 95%CI = [.14, .24], $p < .001$) were identical. Similarly, the difference between overall weighted effect sizes estimations with ($r = .19$, $SE = .03$, 95%CI = [.13, .24], $p < .001$) and without subgroups where writing tasks were not in English ($r = .19$, $SE = .03$, 95%CI = [.14, .24], $p < .001$) were negligible. The results indicated that the relations between syntactic features and writing outcomes did not significantly differ by language.

Publication Year

With the emergence of tools that can generate linguistic features automatically and the availability of digital writing, the ease and reliability of syntactic measures may improve. As we include articles from the 1970s, it is possible that with advances in technology and improvement in accuracy in calculating syntactic features, the reliability of syntactic measures increases. Thus, we observed a smaller range of years, 2009 to 2022, as defined by the emergence of studies using automated tools (e.g., Coh-Metrix, L2SCA, TAASSC). However, after adding this limitation to the inclusion criteria for moderation analysis, the results did not change, indicating that publication year was not likely to be the reason for insignificant moderation effects.

3.4.3 Reporting of Reliability

Due to the lack of reporting on reliability for both syntactic measures and writing outcomes, it is likely that studies that did not report reliability may have low reliability with their

measures, which may potentially lead to biased correlations. Thus, we ran a sensitivity analysis to understand whether insignificant moderation effects may have been masked by lack of reliability in syntactic or writing measures.

When limiting the analysis to studies that have reported reliability for syntactic measures, we found a significant moderation effect of genre where the relation between word per unit and writing outcomes was weaker informational genre than narrative genre ($b = -.22, p =. 009$). Other results remained the same. When limiting the analysis to studies that have reported reliability for writing outcomes, we found a significant moderation effect of grade levels where the relation between word per unit and writing outcomes was weaker for primary grade students than adults ($b = -.16, p =. 004$). In addition, a significant moderation effect of writing outcome was found where the relation between word per unit and writing outcomes was weaker for writing fluency and writing quality ($b = -.11, p <.001$). Other results remained the same.

Publication Type

We also ran a sensitivity analysis by only including studies that were peer-reviewed journal articles. By limiting publication type to journal articles, a significant moderating effect of genre was found for the relation between words per unit and writing outcomes: the relation was stronger for expository texts than narrative texts ($b = .30, p =. 009$). In addition, a significant moderation effect was found between subordination and writing outcomes: the relation between subordination and productivity was stronger than the relation between subordination and writing quality ($b = .11, p = .05$). Other results remained the same. The results indicated that some of the variances from non-peer-reviewed studies may have masked moderation effects. However, overall, traditional syntactic complexity measures were not sensitive to changes across grade levels, language proficiency levels, and measurement features.

We ran Egger's test and made a funnel plot (see Figure) to examine potential publication

bias in the included studies. The funnel plot was symmetrical, and Egger's test was not

statistically significant ($b = .20$, $p = .51$), indicating no publication bias.

Figure 1.2

Funnel Plot
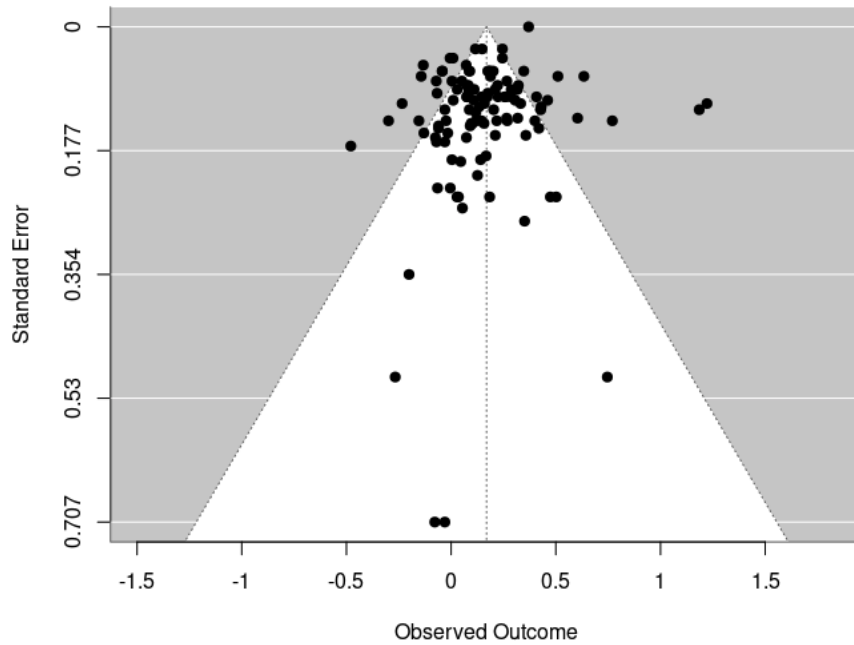
Discussion

The primary purpose of this meta-analysis was to examine the relation between syntactic features in writing and writing outcomes and the moderation effects of both writers' features (grade level, language proficiency status) and measurement features (the type of syntactic complexity measures, text-based or not, writing genre, and writing outcomes).

Overall Relations between Syntactic Features in Written Composition and Writing Performance

There was a weak relation between syntactic accuracy and the majority of syntactic complexity measures and writing outcomes. Although the relation found is weak, this result confirmed the critical role of syntactic skills, specifically in writing (Berninger et al., 2002; Graham, 2018; Y.-S. G. Kim & Graham, 2022) for several reasons. First, text-based linguistic measures are indirect measures of students' language skills. Despite the crucial role of language in writing, objective syntactic features can only partially capture students' ability to utilize syntactic structures in writing. Second, as have been indicated in previous research, writing is a complex activity and numerous skills contribute to the process (e.g., Graham, 2018; Wagner et al., 2011). Thus, when a specific skill has a weak relation with writing outcomes, it may still signal its importance during the writing process. Third, although the ability to write complex sentences is crucial for communicating complex ideas, writing complex sentences does not necessarily lead to higher writing performance and may sometimes even obscure meaning-making (Beers & Nagy, 2009). Given the above reasons, a weak relation between syntactic features in writing and writing performance indicates the importance of syntactic skills in writing: writing sentences proficiently and accurately allows writers to convey their intended ideas accurately and richly (Beers & Nagy, 2009; Y.-S. G. Kim & Graham, 2022), leading to

37

higher writing quality. Thus, it is crucial to address syntactic demands in student writing to improve overall writing quality.

Moderating Effects by Writers' Characteristics

We found moderation effects of writers' characteristics among some syntactic features. The moderation effect of grade level (a proxy for developmental phase) was only found in syntactic accuracy: syntactic accuracy was more strongly related to writing performance for students in primary grades (Kindergarten to Grade 2) than adults. This indicates that syntactic accuracy was particularly important to writing in the early stages of students' writing development. These findings align with the importance of syntactic skills in writing as articulated in theoretical models (see the Theoretical Framework section). Children in primary grades are rapidly developing foundational aspects of syntactic knowledge and, therefore, may make more grammatical errors (Datchunk et al., 2021), which may interfere with high-quality writing. Previous studies have indicated syntactic complexity measures might be sensitive to students' syntactic development over time, although these studies mostly used cross-sectional data (Beers & Nagy, 2011; Bulté & Housen, 2014; Crossley et al., 2011; Crossley & McNamara, 2014; Crowhurst, 1980b; Crowhurst & Piche, 1979; Hunt, 1970; Jagaiah et al., 2020; Wagner et al., 2011). However, no moderation effects of grade level have been found in the current study, which may be due to large variances across studies. Given the large number of studies investigating the relation between syntactic features and writing performance, the sources of variations can be diverse, such as writing tasks and characteristics of writers.

Students' language proficiency status moderated the relation between syntactic features and writing performance. The relation between syntactic features and writing outcomes was stronger for students in foreign language contexts than for students in monolingual contexts or a

context with both monolingual and bilingual students. In addition, the relation between noun phrase complexity and writing performance was stronger for students with limited language proficiency. These results are likely attributed to the fact that lack of language proficiency may constrain their language use (especially noun phrase complexity) in writing and consequently influence writing quality.

Moderating Effects by Measurement Features

Another important goal of the study was to investigate measurement factors (the type of syntactic complexity measures, text-based or not, writing genre, writing outcomes) that may change the relation between syntactic features in writing and writing performance. As noted earlier, DIEW posits that measurement of constructs influences the nature of relations between constructs (Y.-S. G. Kim & Graham, 2022). The relation between syntactic features and writing outcomes differed across different types of syntactic complexity measures: the relation was weaker for words per sentence. This is because words per sentence captures coordination structures (i.e., two independent clauses connected by a coordination conjunction) in addition to complex syntactic features that may make sentences longer. Thus, writing samples with higher values in words per sentence may employ more coordination structures, which technically is not a complex syntactic feature (Norris & Ortega, 2009) and does not enable writers to embed more ideas and details into their sentence structures.

As for the moderating effect of text-based tasks versus not, the relation between subordination and writing performance is stronger when the writing task is text-based. It is possible that text-based writing tasks may provide students with information to elaborate on, which can elicit more subordinating structures. In addition, restating and reorganizing

information from texts (especially when writers cite content from sources) may also require more usage of subordination structures.

Regarding relations of syntactic features to various writing outcomes, we found that the relation between noun phrase complexity and productivity was weaker compared to the relation between noun phrase complexity and writing quality. The use of complex noun phrases may signal a higher level of formality (Biber & Gray, 2011). In addition, Biber and Gray (2011) indicated that expansion caused by noun modifiers is more likely to be accompanied by expansion in meaning. Therefore, the differential relations of noun phrase complexity to writing outcomes may reveal the functions of noun phrases to signal formal tone in writing and elaborate on key content.

In addition, the relation between syntactic features and writing performance varied by writing genre, which is in line with previous studies' findings (Jagaiah et al., 2020). Specifically, the results indicated that the relation between noun phrase complexity and words per unit and writing performance is stronger in expository than narrative texts. This suggests that the demands for complex noun phrases and complex syntactic structures are higher in expository texts than in narrative texts. Noun phrase complexity and words per unit have been associated with academic style and elaboration in writing (Biber & Gray, 2011; Staple et al., 2016), indicating that expository genre may be more aligned with writing in academic context and needs more elaboration to be high quality written compositions.

Overall, the results suggest measurement factors are critical in moderating the relation between syntactic features and writing outcomes, underscoring the dynamic relations between language and writing as a function of measurement factors and different dimensions of writing outcomes (Y.-S. G. Kim & Graham, 2022).

40

One important finding of the current study is that many moderators of traditional syntactic features (i.e., words per unit, subordination measures, words per sentence) were not significant, although previous studies suggest that there may be differential relations in terms of grade levels, language proficiency, and measurement features (Beers & Nagy, 2009, 2011; Jagaiah et al., 2020; Y.-S. G. Kim & Graham, 2022; Steward & Grobe, 1979). Among all syntactic measures, we only found one syntactic complexity measure of which the majority of the moderators were significant: noun phrase complexity features. This could be due to several reasons. The first potential reason is the importance of complex noun phrase structures in writing development, especially in school contexts (Biber & Gray, 2011; Staples et al., 2016). Given that noun phrases allow writers to pack dense information effectively, using such structures may be increasingly critical in school settings. A second reason may be the smaller variance introduced by a smaller number of samples (21 subsamples in total) included. A third reason could be due to study quality. The results may differ by publication types based on our findings from the sensitivity analysis. Although our analysis indicated that there is no publication bias, when running a sensitivity analysis for nonsignificant moderators, we found that it is likely that published articles are more likely to detect certain moderation effects to some extent. In addition to publication type, the reporting of reliability of syntactic and writing measures is another factor of study quality that might have impacted our findings. Our sensitivity analysis indicated that when limiting studies to those that have reported reliability, we are more likely to find significant moderation effects, which indicate that the insignificant moderation effects may be due to lack of reliability in syntactic and writing measures in previous studies. Nevertheless, there could be a plethora of other factors that mask potential moderating effects of the relation between syntactic features and writing outcomes.

Lastly, previous studies have indicated that noun phrase complexity measures may be more sensitive to development. Although some subordination measures are also fine-grained, previous studies have shown that subordination measures are unable to show meaningful differences with writing development (Bulté & Housen, 2014; Casal & Lee, 2019; Crossley & McNamara, 2014). In contrast, noun phrase complexity measures have been found to differ across different levels of writing quality (Casal & Lee, 2019; Kyle & Crossley, 2018; Guo et al., 2013) and more sensitive to age (Ansarifar et al., 2018; Bulté & Housen, 2014). The results of the study align with Norris and Ortega's (2009) proposed three-staged trajectories for second-language syntactic development where coordination, subordination, and phrasal-level complexity are developed sequentially: syntactic structures (coordination and subordination) that are developed during earlier stages may not be sensitive enough to assess students' current syntactic skills. Since studies conducted on the topic of syntactic features in writing mostly feature younger students still developing their language skills or writers with limited language proficiency, it is reasonable that the results of the current study align with the proposed syntactic development trajectories for bilingual learners (Biber et al., 2014; Staples et al., 2016).

Limitations and Future Directions

Future studies should explore which are valid and reliable measures of syntactic complexity beyond traditional measures (e.g., T-unit length, clauses per T-unit). The results of the current study indicate a reliance on the more traditional measures that may not be as sensitive in serving assessment purposes. It is also critical to understand which ones may be more sensitive to students' writing development, with considerations of writing in different genres and tasks (Casal & Lee, 2019).

It is also worth noting that, generally, studies that examine syntactic features in writing did not report students' language proficiency consistently. Many studies did not report the language proficiency level of their student sample, and the studies that did report this information did not have a consistent standard in reporting students' proficiency, making comparisons across studies challenging. Future studies examining the topic should explicitly and comprehensively report students' language abilities so that cross-study comparisons are possible. Another significant gap in the current literature is the lack of representation of students with learning disabilities. Most of the studies either did not report information on students' special education status or excluded these students from their sample. Future research on syntactic features in writing of students with learning disabilities is warranted. In addition, potentially due to large variations across studies, the moderation effect of grade level was not found in most cases. Future studies should consider using longitudinal design or consistent writing tasks and syntactic measures so that the role of developmental phases in the relation between syntactic skills and writing quality can be understood.

Implications for Practices

The study indicates the crucial role of syntactic abilities in achieving high writing performance. Therefore, supporting students' accuracy in using syntactic features and increasing the variety of students' syntactic structures they can draw on to accomplish diverse functions in writing. In addition, the study also highlights specific syntactic demands students may encounter at different developmental stages across different writing genres. Specifically, our results indicate that syntactic accuracy is especially critical for students in primary grades, and expository genres may have higher demands in using complex syntactic structures. For students with limited language proficiency, acquiring certain syntactic features (complex noun phrases,

43

subordination) and improving syntactic accuracy may be even more critical than other students because their relative lower language skills may impact the process where they translate their ideas into sentences. Based on previous research on supporting grammar to improve students' writing, grammar instruction should be integrated and contextualized so that students can learn different functions and purposes associated with different meaning-making processes (Jones et al., 2013). For example, subordination clauses can help writers to introduce other perspectives or cite evidence from sources, which aligns with our finding indicating the association between subordination structures and writing performance is stronger in text-based writing tasks than non-text-based writing tasks.

The study has important implications for assessing writing. For instance, given that noun phrase complexity may be more crucial in expository genre than narrative genre, special attention should be paid to students' ability to write complex noun phrases in expository writing tasks so that detailed information can be efficiently packed into sentences. When it comes to text-based writing, students' ability to use subordination structures should be considered when assessing whether students have sufficient linguistic resources to deal with the specific demands of text-based writing tasks. Findings in the current study also indicate that being able to use linguistic resources proficiently in one writing task/genre does not guarantee writers' linguistic literacy (where they need to employ certain linguistic features corresponding to distinct demands and purposes of different tasks/genres; Ravid & Tolchinsky, 2002) and the ability to use language proficiently in other tasks/genres. Our finding also indicates that monitoring the ability to use more complex syntactic structures when necessary for certain writing tasks is crucial for writing performance of bilingual students with limited language proficiency. The current study suggests that noun phrases complexity and subordination structures may be features to pay

44

attention to during formative assessment. Future studies are warranted to understand other

specific syntactic features that are crucial for writing assessment of students with limited

language proficiency.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?. *Reading and Writing*, *22*(2), 185-200. https://doi.org/10.1007/s11145-007-9107-5

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., ... & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology*, *94*(2), 291. https://doi.org/10.1037/0022-0663.94.2.291

Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language & Linguistics*, *15*(2), 223-250. https://doi.org/10.1017/S1360674311000025

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, *37*(5), 639-668. https://doi.org/10.1093/applin/amu059

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42-65. https://doi.org/10.1016/j.jslw.2014.09.005

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing*, *44*, 51-62. https://doi.org/10.1016/j.jslw.2019.03.005

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415-443. https://doi.org/10.17239/jowr-2020.11.03.01

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. https://doi.org/10.1016/j.jslw.2014.09.006

Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, *28*(3), 282-311.

Crowhurst, M. (1980). Syntactic complexity and teachers' quality ratings of narrations and arguments. *Research in the Teaching of English*, 223-231.

Crowhurst, M. (1983). Syntactic complexity and writing quality: A review. *Canadian Journal of Education/Revue canadienne de l'education*, 1-16. https://doi.org/10.2307/1494403

Crowhurst, M., & Piche, G. L. (1979). Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. *Research in the Teaching of English*, *13*(2), 101-109.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, *10*(1), 5-43. https://doi.org/10.1016/j.asw.2005.02.001

Cumming, A., Lai, C., & Cho, H. (2016). Students' writing from sources for academic purposes: A synthesis of recent research. *Journal of English for Academic Purposes*, *23*, 47-58.

Datchuk, S. M., Hier, B. O., & Watts, E. A. (2021). Accounting for levels of language in narrative and expository writing: A skills analysis of second-grade student writing. *The Elementary School Journal*, *121*(4). https://doi.org/10.1086/714051

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, *28*(2), 414-420.

Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist*, *53*(4), 258-279.https://doi.org/10.1080/00461520.2018.1481406

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, *9*(2), 123-145.

Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English*, 75-85.

 *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. https://doi.org/10.1016/j.asw.2013.05.002

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing process. *The Sciences of Writing*. Hillsdale, NJ: Lawrence Erlbaum Associates, 3-30.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39-65. https://doi.org/10.1002/jrsm.5

Jagaiah, T., Olinghouse, N. G., & Kearns, D. M. (2020). Syntactic complexity measures: variation by genre, grade-level, students' writing abilities, and writing quality. *Reading and Writing*, *33*(10), 2577-2638. https://doi.org/10.1007/s11145-020-10057-x

Jones, S., Myhill, D., & Bailey, T. (2013). Grammar for writing? An investigation of the effects of contextualised grammar teaching on students' writing. *Reading and Writing*, *26*, 1241-1263.

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. https://doi.org/10.1016/j.asw.2018.03.002

*Kim, Y.-S. G., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranik, C. (2014). Evaluating the dimensionality of first-grade written composition. *Journal of Speech, Language, and Hearing Research*, *57*(1), 199-211. https://doi.org/10.1044/1092-4388(2013/12-0152)

Kim, Y.-S. G., & Graham, S. (2022). Expanding the direct and indirect effects model of writing (DIEW): Dynamic relations of component skills to various writing outcomes. *Journal of Educational Psychology*, *114*(2), 215-238. https://doi.org/10.1037/edu0000564

*Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, *102*(2), 333-349. https://doi.org/10.1111/modl.12468

Li, H. (2015). *Relationship between measures of syntactic complexity and judgments of EFL writing quality*. In Proceedings of 2015 Youth Academic Forum on Linguistics, Literature, Translation and Culture. American Scholars Press, 216-222.

Loban, W. (1976). *Language Development: Kindergarten through Grade Twelve*. NCTE Committee on Research Report No. 18.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36-62.

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, *34*(4), 493-511. https://doi.org/10.1177/0265532217710675

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*(6), 1553-1574.

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

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

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555-578. https://doi.org/10.1093/applin/amp044

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, *26*, 45-65.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492-518. https://doi.org/10.1093/applin/24.4.492

Perin, D., & Lauterbach, M. (2018). Assessing text-based writing of low-skilled college students. *International Journal of Artificial Intelligence in Education*, *28*(1), 56-78.

Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing*, *33*, 3-17.

Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language*, *29*(2), 417-447.

Schleppegrell, M. J. (1998). Grammar as resource: Writing a description. *Research in the Teaching of English*, 182-211.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, *43*(2), 324-339. https://doi.org/10.1044/jslhr.4302.324

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and

genre. *Written Communication*, *33*(2), 149-183.

https://doi.org/10.1177/0741088316631527

Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication and other bias in

meta-analysis. *Publication Bias in Meta-analysis: Prevention, Assessment and*

Adjustments, 99-110. https://doi.org/10.1002/0470870168.ch6

*Stewart, M. F., & Grobe, C. H. (1979). Syntactic maturity, mechanics of writing, and teachers'

quality ratings. *Research in the Teaching of English*, *13*(3), 207-215.

*Troia, G. A., Shen, M., & Brandon, D. L. (2019). Multidimensional levels of language writing

measures in grades four to six. *Written Communication*, *36*(2), 231-266.

https://doi.org/10.1177/0741088318819473

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of*

Statistical Software, *36*, 1-48. https://doi.org/10.18637/jss.v036.i03

*Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor,

P. T. (2011). Modeling the development of written language. *Reading and Writing*, *24*(2),

203-220. https://doi.org/10.1007/s11145-010-9266-7

Wang, J., Kim, Y. S. G., & Cho, M. (2024). Linguistic features in narrative and opinion genres

and their relations to writing quality in fourth grade writing. *Journal of Research in*

Reading, *47*(2), 220-239.

CHAPTER 2

Predicting Bilingual and Monolingual Secondary Students' Writing Quality in English Using Linguistic Feature: A Machine Learning Approach

Multiple theoretical models of writing have emphasized the importance of language skills in writing, indicating that the ability to employ linguistic resources accurately, proficiently, and flexibly is critical to becoming successful writers (Berninger et al., 2002; Berninger & Winn, 2006; Juel et al., 1986; Graham, 2018). Investigating linguistic features in writing is one critical approach to understanding writers' language skills in writing because it allows researchers and educators to observe language use in authentic contexts and examine register flexibility (i.e., patterns of language use that vary by register and genre; Hyland, 2006) in specific genres.

Linguistic features can provide rich information about language and literacy skills of students. Numerous linguistic features in writing, including syntactic, lexical, and cohesion features, have been examined in previous studies (e.g., Crossley et al., 2019; Crossley & Kyle, 2018; Kyle, 2016; Kyle et al., 2021; Lu, 2017; McNamara et al., 2014). Previous studies consistently showed that linguistic features could explain about 50%-60% of the variance in writing quality (e.g., Crossley & McNamara, 2011a, 2012; Crossley et al., 2014; Guo et al., 2013; June et al.,2019; M. Kim & Crossley, 2018; Lee et al., 2021). However, most of these studies used traditional linear regression models, limiting the number of measures that can be examined simultaneously, and may also have limited the potential of using linguistic features to predict writing quality scores. In the current study, we investigated secondary students' language (syntactic, lexical, cohesion, source use, and others) features generated by natural language processing (NLP) software in source-based argument writing in the context of history class to understand the extent to which these features predict different dimensions of writing quality of

source-based argumentative writing in history (i.e., structure, evidence use, historical thinking, and language use) using machine learning approaches. Machine learning approaches can enhance the prediction of writing quality compared to traditional linear regression models. Research examining how linguistic features predict writing outcomes can expand researchers' understanding of how to evaluate writing more accurately and efficiently and improve educators' understanding of which linguistic features are critical to master to be successful writers in a specific genre.

After identifying key features that contribute to the prediction of various dimensions of writing quality, we investigated how these features differ for bilingual and monolingual secondary students. Bilingual students are broadly defined as those who speak more than one language. In this study using data from students in the US where the societal language is English, students whose home language is not English are classified in their English proficiency, including those designated as having limited proficiency (English Learners; EL), and those that were classified as proficient in English, including Reclassified Fluent English Proficient (RFEP) and Initially Fluent English Proficient (IFEP). Because of bilingual students' experience with other languages in addition to English and various levels of English language proficiency, their language use patterns and demands may differ from English-only (EO) students. As illustrated in a meta-analysis (Williams & Lowrance-Faulhaber, 2018), bilingual students use their entire linguistic repertoire to write. However, there is a lack of research on the writing development of bilingual secondary students in the US context. The current study aims to fill this gap in the literature and examine linguistic features in writing for bilingual and monolingual students with various language proficiency levels.

Linguistic Features and Writing Quality

Theoretical models have recognized the importance of language skills in writing. In the Writer(s)-Within-Community Model of Writing (Graham, 2018), language skills, as one crucial aspect of long-term memory, impact the richness of writing. According to the Direct and Indirect Effects model of Writing (DIEW; Y.-S. G. Kim & Graham, 2022), language skills such as vocabulary and grammatical knowledge (morphosyntactic and syntactic knowledge) contribute to two essential writing processes: translation and evaluation. Language skills are important to establish coherence, partly via cohesion. Text coherence is a critical standard used to evaluate writing quality and is a strong indicator of writing quality (Crossley & McNamara, 2010; 2011a; Y.-S. G. Kim & Graham, 2002; McCulley, 1985). Writing lacking coherence will most likely fail to communicate its intended message (Bamberg, 1983). Cohesion is a necessary condition for achieving coherence, marked by using lexical or grammatical cues that link chunks of text together (Crossley & McNamara, 2010).

Although language skills mainly develop in early childhood, language skills develop through adolescence, including language features in writing (Crossley et al., 2011). In fact, language skills were shown to be uniquely related to writing quality for adolescents (Dockrell et al., 2009). However, fewer studies have examined secondary students' language use in writing, and the majority of extant studies focused on Natural Language Processing (NLP) tools and undergraduate students or adult English as a Foreign Language (EFL) learners (e.g., MacArthur et al., 2019; Crossley & McNamara, 2011a). In addition, NLP features predict writing quality scores for essays that are both source-based and not source-based (Crossley & McNamara, 2011a; Guo et al., 2013). Source-based argumentative essays are not as widely examined as essays that are not source-based (Cumming et al., 2016) despite research indicating linguistic

55

features predicting their writing quality may differ (Guo et al., 2013). Below, a review of linguistic features (lexical, cohesion, syntactic, and other features such as readability measures) in writing will be presented.

Lexical Features

Lexical sophistication and lexical diversity features are commonly examined lexical text features. Studies have found that high-quality essays tend to use more diverse and sophisticated vocabulary in argumentative essays that are source-based or not source-based (e.g., Crossley & McNamara, 2011a, 2012; Cumming et al., 2005; Guo et al., 2013; Kyle & Crossley, 2016; MacArthur et al., 2019; McNamara et al., 2010). McNamara and colleagues (2010) found lexical diversity (measured by the Measure of Textual Lexical Diversity [MTLD]) and word frequency accounted for a 36% variance in writing quality (essay length is not accounted for in this study). Crossley and McNamara (2011a) found that lexical diversity (measured by *D*) explained the most variance in writing quality compared with other linguistic measures in a study examining senior high school students' argumentative writing in the EFL context. Similarly, a study using a sample of U.S. undergraduate students found that the number of different words explained more variance in writing quality than all other linguistic features examined (Crossley & McNamara, 2012). The contribution of lexical sophistication is also consistent across studies despite its relatively smaller contribution compared with lexical diversity (Crossley & McNamara, 2011a, 2012). Guo and colleagues (2013), using Test of English as a Foreign Language (TOEFL) essays from EFL learners, found that lexical sophistication (as measured by word familiarity and frequency of content words) was a significant predictor for writing quality in the source-based argumentative task.

Cohesion Features

Cohesion is commonly defined as the presence or absence of lexicogrammatical cues that connect text as a whole (Crossley & McNamara, 2010; 2011a). In this study, we will combine Halliday and Hasan's (1976) and Crossley and colleagues' (2016) typology of cohesion features. Crossley and colleagues (2016) categorized cohesion features based on whether these features measure cohesion across sentences, paragraphs, or the whole text. According to them, there are three types of cohesion: local, global, and overall cohesion (Crossley et al., 2016). Local cohesion refers to cohesion at the sentence level. The sentence level is the smallest unit of analysis for cohesion analysis using NLP tools. Global cohesion refers to cohesion between larger units of text, usually paragraphs. Unlike local and global cohesion, which examine cohesion features by dividing the writing into chunks and comparing overlaps, overall cohesion features are the incidence of certain features throughout the text. Although categorizing cohesion features as local, global, and overall cohesion features can help researchers distinguish between cohesion measures that examine at the sentence level, passage level, and across the whole text, this categorization cannot effectively capture how cohesion features perform differently. Thus, the current study will also stick to Halliday and Hasan's (1976) typology for clarity. The cohesion features that will be examined in the study include lexical cohesion, connectives and conjunctions, reference, and others (temporality, causality, and intentionality) across sentences, paragraphs, and the whole text.

Overall, studies conducted with college writers show inconsistent results between local word overlap measures and writing quality/coherence measures (Crossley & McNamara, 2010; 2011a; 2016; Guo et al., 2013; M. Kim & Crossley, 2018; MacArthur et al., 2019). While some studies (Crossley & McNamara, 2016; Guo et al., 2013; M. Kim & Crossley, 2018; MacArthur et al., 2019) found a weak positive relation, other studies found a weak negative relation (Crossley

& McNamara, 2010, 2011a). Two studies in particular examined source-based argumentative writing tasks (Guo et al., 2013; M. Kim & Crossley, 2018), the authors of which reported positive relations between local word overlap measures and writing quality. A more consistent pattern has been found for global word overlap measures. Studies have found positive weak to moderate relations between global word overlap measures and writing quality/coherence measures (Crossley et al., 2016; Crossley & McNamara, 2011a; 2016). However, a study using source-based argumentative writing (Perin & Lauterbach, 2018) found that students with lower writing proficiency have a higher value on global argument overlap. In one study with second language (L2) college students, student texts that were revised to be more cohesive (both locally and globally) and not revised to be more elaborated improved their scores on overall writing quality (Crossley & McNamara, 2016). Inconsistent results have been reported for word overlap cohesion measures at the overall level (Guo et al., 2013; MacArthur et al., 2019).

Similar patterns can be seen with semantic similarity cohesion features (Crossley et al., 2019; Crossley & McNamara, 2011a, 2011b; 2016; Guo et al., 2013; MacArthur et al., 2019; McNamara et al., 2013). In addition, the magnitude of the relation between global cohesion features and writing quality/coherence seems to be larger for semantic similarity measures than word overlap measures (Crossley et al., 2019; Guo et al., 2013). In summary, global measures of lexical cohesion have been found to be positively correlated with writing quality/ coherence, and inconsistent results have been found for local and overall lexical cohesion measures.

The use of connectives is one of the critical aspects of overall cohesion. Studies have shown inconsistent results regarding the role of connective use in writing (Crossley & McNamara, 2010, 2011a; McNamara et al., 2010). Findings of several studies indicate that different categories of connectives behave differently in their relations to coherence or writing

quality (Myhill, 2008; Witte & Faigley, 1981; Yang & Sun, 2012). Furthermore, studies that found positive relations between connective usage and writing quality generally concluded that more proficient writers use a wider variety of connectives, including causal, adversatives, and additives (Myhill, 2008; Witte & Faigley, 1981; Yang & Sun, 2012).

The relation between referential cohesion features and writing quality is inconclusive based on the existing literature at this time (Crossley & McNamara, 2010; Crossley & McNamara, 2016). NLP tools can measure several additional aspects of cohesion, including temporality, causality, and intentionality. Ratio of causal articles was found to be negatively correlated with coherence in argumentative writing for undergraduate students (Crossley & McNamara, 2010). Cohesion features related to temporality were found to be positively correlated to writing quality in non-source-based argumentative writing but negatively correlated to writing quality in source-based argumentative writing (Crossley & McNamara, 2011a; Guo et al., 2013). Despite varied relations between cohesion features and writing quality, studies have generally found that cohesion features do not uniquely explain a large amount of variance in writing quality when other linguistic features (lexical and syntactic) are taken into account (Crossley & McNamara, 2011a, 2012; Guo et al., 2013).

Syntactic Features

Among NLP measures, syntactic features include syntactic complexity, measures related to plurality, tense, and aspect (e.g., third-person singular form), and measures related to parts of speech of the words used in texts. Multiple syntactic complexity measures have been created to measure the construct at the phrase, clause, sentence, and other or mixed levels (Jagaiah et al., 2020). Although studies have found that a range of syntactic complexity measures at the phrase, clausal, and sentence level are positively related to writing quality in argumentative writing (e.g.,

Beers & Nagy, 2011; Jagaiah et al., 2020; Kyle, 2016; Lu, 2017; Qin & Uccelli, 2016; Wang et al., 2024), the relation between syntactic features and writing quality is inconclusive for teenagers and adults when other linguistic features are accounted for (Crossley & McNamara, 2011a; Cumming et al., 2005; Gebril & Plakans, 2009; Guo et al., 2013; MacArthur et al., 2019). Several studies did not find that syntactic complexity measures significantly predict writing quality in students' argumentative writing when other linguistic features (e.g., lexical, cohesion, syntax) were accounted for (Crossley & McNamara, 2011a, 2012; Guo et al., 2013; M. Kim & Crossley, 2018). One study found that syntactic complexity (measured by a latent variable consisting of overall syntactic similarity, mean number of words before the main verb, and sentence length) was negatively related to writing quality of argumentative essays by basic college writers (MacArthur et al., 2019). McNamara and colleagues (2010), using corpus data from undergraduate students, found that syntactic complexity (measured by number of words before the main verb) predicted an 11% variance in writing quality when other linguistic features are accounted for. The relation between syntactic complexity and writing quality also tends to differ by measures (Beers & Nagy, 2009; Cumming et al., 2005; Jagaiah et al., 2020). For example, Cumming and colleagues (2005) found that high-quality essays tend to have more words per T-unit but not other syntactic complexity measures tested. Fewer studies have reported a relation between other syntactic features and writing quality. For example, Guo and colleagues (2013) found that EFL learners use more third-person singular forms, and past participle verbs tend to score higher in writing quality.

In addition to lexical, cohesion, and syntactic features, NLP tools can also generate other linguistic features such as readability measures, rhetoric measures (e.g., narrativity score, the degree to which a text exhibits features of a narrative), and n-gram measures. Some of these

60

measures have also been shown to predict writing quality. For instance, Crossley and McNamara (2012) found that narrativity score, conclusion paragraph n-grams, and body paragraphs n-grams all predicted writing quality significantly and independently. In summary, across all linguistic features that have been examined, only clear patterns (positive relations) have been found between syntactic complexity, lexical sophistication, lexical diversity, global word overlap, and global lexical cohesion and writing quality. The relations between other linguistic features and writing quality are still inconclusive.

Dimensions of Argumentative Writing in History

Among all measures of student writing performance, writing quality is widely considered one of the most important outcomes (Y.-S. G. Kim & Graham, 2022). Writing quality is a multidimensional construct (Puranik et al., 2008; Steiss et al., 2022; Wagner et al., 2011; Y.-S. G. Kim & Graham, 2022). According to the dynamic relations hypothesis of DIEW, the relations of skills and knowledge to writing differ as a function of the dimensions of writing (Y.-S. G. Kim & Graham, 2022). One goal of the current study is to explore how various linguistic features relate to various dimensions of writing. In this study, we examine four dimensions using analytic coding: structure, evidence use, historical thinking, and language use.

Both general and discipline-specific knowledge are crucial to historical writing (Monte-Sano, 2010). Two dimensions that have been generally and commonly used in writing rubrics are structure and language use (Culham, 2003), which are included in the current study. The relation between linguistic features and structure has been discussed by previous researchers. Van Dijk (1977) and Bamberg (1983) drew researchers' attention to the distinction between local and global coherence. They noted that while cohesive ties can assist in creating local coherence, global coherence needs to be established by an overall form or structure of that specific genre.

61

Readers have specific schema and expectations of a certain genre regarding its structure and goals. If writers can successfully meet such expectations, the coherence of a text can be better achieved (Bamberg, 1983). The relation between linguistic features and essay structure is thus worth exploring to understand how local or global cohesion features and other linguistic features are associated with the overall structure of the essay. We also examined the relation between linguistic features and the language use dimension of writing quality to understand how NLP linguistic features are related to language use dimension of student writing quality.

In addition to achieving common and general dimensions in writing, writing an argumentative essay in history also requires students' work situated in a disciplinary community (Goldman & Scardamalia, 2013; Goldman et al., 2016; Shanahan & Shanahan, 2008). Evidence use is a relevant construct to examine its relation with linguistic features because integrating evidence is a crucial element in source-based argumentative writing (Monte-Sano, 2010). In addition, language skills are required in the process of reading and comprehending source materials, evaluating and choosing evidence to support arguments, and integrating evidence into the written text with coherence. In the process of choosing evidence and writing up arguments, writing in history communities also requires writers to interpret and critically analyze the sources they are using (Goldman & Scardamalia, 2013), which requires specific historical thinking processes, including sourcing, contextualization, and corroboration (Goldman et al., 2016; Wineburg, 1991). On top of the three processes, since it is common in history for perspectives to be contradicting, we also measured the extent to which students are able to address counterarguments critically as part of the writing prompt (Goldman et al., 2016; Monte-Sano, 2010). By examining which linguistic features are related to these history-specific processes, educators can better understand the linguistic demands of discipline-specific literacy skills.

The relation between linguistic features and different dimensions of writing can vary. For example, linguistic features may show a stronger relation to the language use dimension than other dimensions, such as evidence use. In addition, the relation between language use and content and ideas may also be stronger than the relations between language use and evidence use and historical thinking for two reasons: (1) language skills are crucial for the idea translation process (Graham, 2018; Y.-S. G. Kim & Graham, 2022), (2) students with higher language proficiency (more advanced language use features) have more cognitive resources for other important aspects of writing that may improve the general content of their written compositions (Graham, 2018). Cohesion features may be more strongly related to the structure dimension of writing quality. Van Dijk (1977) and Bamberg (1983) proposed that the structure of essays plays a crucial role in establishing the coherence of texts. Meanwhile, cohesion features can assist writers in clarifying the relations between sentences and paragraphs. Thus, cohesion features of essays with a reasonable and clear structure may differ from those whose structure fails to establish coherence. By exploring how students' linguistic features in writing are related to the analytic coding of writing quality (i.e., structure, evidence use, historical thinking, language use), we aim to expand our understanding of which linguistic features are more crucial for certain aspects of their writing performance, which can inform how to effectively integrate language element in writing instruction. Prior research indicates that functions and purposes of linguistic features should be integrated into writing instruction so that students can associate linguistic features with different meaning-making processes (Jones et al., 2013). Without an understanding of the ways linguistic features are critical to distinct dimensions of writing quality and targeting teaching linguistic features for specific purposes, teaching language may be less effective, given that language skills incorporate a broad range of skills.

The majority of previous studies that examined relations between NLP linguistic features and writing quality used holistic scoring to represent writing quality instead of using analytic scoring of specific dimensions of writing quality (Wang et al., 2024). Golparvar and Abolhasani (2022) examined the relations between syntactic complexity and dimensions of writing quality among Iran EFL adult learners in expository writing, including content, organization, and language use, and found a weak correlation between word per T-unit with the content and language use dimension, but not the organization dimension. Kim and colleagues (2014) found that there was no relation between syntactic complexity and two dimensions of writing quality (i.e., content and organization) for primary school students in the US. As there are few studies that have examined the relations between NLP linguistic features and human ratings of dimensions of writing quality, it is still unclear which linguistic features are associated with different dimensions of human ratings of writing quality. Understanding such relations can inform writing assessment by using NLP measures to assess dimensions of writing quality in student writing.

Bilingual Students' Language Features in Writing

In this section, we review evidence of how linguistic features vary for students of different proficiency levels to understand bilingual writers' writing development and the role of language proficiency in language use in writing. Previous studies have found that bilingual college students with higher language proficiency produced sentences with complex structures, including more passive voice, subordination, and longer T-units (Ferris, 1994; Ortega, 2003). The impact of language proficiency on cohesion features is evident in previous research. Yang and Sun (2012) found that undergraduate students' referential and lexical cohesion features differ by their language proficiency. While bilingual writers with high English proficiency tended to

64

use various referential and lexical cohesion devices, bilingual writers of low language proficiency used repetitive vocabulary to establish cohesion (Ferris, 1994; Yang & Sun, 2012). Unlike L1 writers who move from using explicit cohesion devices like connectives to more implicit cohesion features, bilingual writers in EFL contexts with high proficiency may use connectives more frequently and diversely (Ferris, 1994; Yang & Sun, 2012).

ELs who achieved English proficiency during their annual assessment are classified as Redesignated Fluent English Proficient (RFEP) and do not receive language support from schools. As assessment criteria are different across states, these students' performance in language and literacy can vary. Studies have found that RFEP students sometimes may outperform English Only (EO) students on language and literacy skills (e.g., Ardasheva et al., 2012) in states where the reclassification process has high demands on students (Hwang et al., 2017). Some other studies indicated that RFEP students, especially recent RFEP students, may need sustained support to continue the development of language proficiency and academic development (Hwang et al., 2017; Slama, 2014). In addition, although bilingual students were labeled in terms of their language proficiency, due to their experiences with other languages, their language use may differ from students who have only been exposed to one language. Thus, in the current study, we will examine EL and IFEP/RFEP students as separate groups of bilingual students from EO students to explore whether their linguistic features in written compositions differ. Such understanding can inform educators of the specific language use patterns and the specific demands they may be facing due to language proficiency.

Present study

The present study extends prior work by examining linguistic features in a student population and writing genre that has rarely been examined: source-based argumentative writing

65

in history by bilingual and monolingual middle and high school students in the US. Specifically, we examined the predictive accuracy of machine learning models in using NLP linguistic features to predict writing quality across different dimensions, including structure, evidence use, historical thinking, and language use. The relation between linguistic features and writing quality has been mixed. Many studies have shown that linguistic features using NLP features related to language use can predict writing quality scores with a certain level of accuracy and can explain about 50%-60% variance in writing quality scores (e.g., Crossley et al., 2014; Crossley & McNamara, 2011a, 2012; Guo et al., 2013; Jung et al., 2019; M. Kim & Crossley, 2018). The study aimed to investigate whether machine learning approaches can explain a higher variance in writing quality measures than traditional linear regression models reported in previous research. Machine learning approaches have been found to have higher prediction accuracy than linear regression in predicting features of student writing (Wan et al., 2021), but no study, to our knowledge, has examined the relation between NLP measures and writing quality using machine learning approaches. In addition, by examining students' linguistic features across language proficiency levels, we can understand students' language use patterns in writing and the specific linguistic demands students may face due to language proficiency. Additionally, by investigating the prediction accuracy of various dimensions of writing quality, the study can provide insights into automated feedback for writers and improve the accuracy of prediction of scores in specific aspects of essays written by students (Taghipour, 2017).

The following are specific research questions of the present study:

1. To what extent can linguistic features predict various dimensions of writing (structure, evidence use, historical thinking, and language use)?

2. How do linguistic features that are important in predicting dimensions of writing quality differ for bilingual and monolingual students of different levels of language proficiency, controlling for gender and grade levels?

We hypothesize that lexical features, especially lexical diversity measures, would relate to aspects of analytic coding of writing quality more strongly than other linguistic measures (Crossley & McNamara, 2011a, 2012; Guo et al., 2013; MacArthur et al., 2019). We hypothesize that students designated as proficient in English language skills would show higher lexical sophistication and syntactic complexity in their writing (Crossley & McNamara, 2012; Kyle & Crossley, 2016). Due to a lack of research, we did not have specific hypotheses for differences in other linguistic features.

Method

Participants

Data for this study are from a larger intervention study investigating features of high-quality writing of source-based argument texts in secondary school history class. In the present study, pretest data were used. The data for this study were collected from 27 teachers in 15 schools from two urban districts public schools in the southwest region of the US. Participants are 597 students from Grade 6 to Grade 12 (52.26% female). The participants were drawn from a stratified random sampling procedure where a fourth of the participants were selected. Out of 570 students for whom ethnicity information was available, there were 381 Hispanic students, 79 White, 21 Pacific Islanders, 14 African American students, 55 Asian American students, eight American Indian or Alaska Native, and 12 from other races/ethnicities. Eighteen students were identified with learning disabilities. Out of 534 students for whom parent education information was available, 53 of them had a graduate degree or higher, 77 were college graduates, 90 some

college or associate degrees, 206 were high school graduates, and 108 were not high school graduates. Regarding EL designation, 24.5% were English learners (EL), 33.0% were Redesignated Fluent English Proficiency (RFEP), 8.0% were Initial Fluent English Proficiency (IFEP), and 34.5% were English Only students (EOs).

Measures

Written Composition

Students were randomly assigned to one of the two history source-based argument writing prompts. Students were asked to write an argumentative essay where they wrote about a boycott and the overall claim as an introduction section, reasons for the success with evidence and reasoning in the body paragraph (with a counterargument paragraph), a summary of the main ideas and insights from the event as the conclusion section. Prompt one was about the Montgomery bus boycott, and Prompt two was about the California grape boycott (see source texts in Appendix B). The writing task was completed in two days. Students read the sources on the first day and wrote an essay about why the boycott succeeded on the second day. Each day, students took 50 minutes to finish the tasks independently. The prompts included a section to teach students basic computer skills to be able to write an essay using Microsoft Word. This section introduced students to the procedures to finish the two-day activity, information where the background of the boycott was briefly introduced, four sources of the boycott, the requirements of the essay, and the section for students to write. A read-aloud recording of all sources was attached to the prompt; students could access it freely (with instruction on how to access it). There was no instruction on the reading materials beyond the read-aloud. Teachers directed students through the process. Among the four sources, the first one was an overall timeline of the boycott, while the other sources introduced events that may have led to the

success of the boycott. The four sources for the bus boycott had 198, 369, 269, and 345 words,

respectively. The Flesch-Kincaid grade-level readability scores for four sources were 46.3, 62.4,

66.4, and 46.7. The four sources for the grape boycott had 237, 278, 224, and 343 words,

respectively. The Flesch-Kincaid grade-level readability scores for four sources were 53.5, 77.6,

75.4, and 72.2.

Analytic coding of writing can provide comprehensive information on students'

weaknesses and strengths (Bacha, 2001; Culham, 2003). The analytic coding framework in the

study was developed based on prior studies (Culham, 2003; National Writing Project, 2010),

literature on writing research, and expert perspectives, and a recent study showed its relation to

writing quality measured by a holistic scoring for text-based argumentative writing quality

(Steiss et al., 2022). An additional component, historical thinking, was included in the analytic

coding framework as the writing assignment being analyzed in the current study involves source-

based argumentative writing in a history class. Previous research highlights the significance of

disciplinary literacy (Fang & Coatoam, 2013; Shanahan & Shanahan, 2012). Writing an

argument in history requires not only general argumentation skills (e.g., language skills, writing

claims and supporting them with evidence) but also specific historical skills (e.g.,

contextualizing arguments; Wineburg, 1991). There were 20 items in total in four categories:

structure (three items), evidence use (four items), historical thinking (four items), and language

use (three items). All items were scored based on a scale from 1-7, where one represented "not

evident" and seven represented "highly effective" (see Appendix C for a list of the items). Six

PhD students and one professor in education scored the essays. All scorers underwent rigorous

training where they did multiple rounds of coding, modified the coding framework based on

ongoing discussion, and selected multiple anchor papers for each score. Discrepancies were

discussed and settled before moving on to the next round. Final inter-rater reliabilities within one point for each item were reported in Appendix C (using 28 essays, 13% of the total sample). On average, inter-rater reliabilities within one point for structure items were 96.7%, evidence use items were 96.4%, historical thinking items were 93.9%, and language use items were 88.6%.

Linguistic Features

Tool for the Automatic Analysis of Cohesion (TAACO; Crossley et al., 2019), Tool for the Automatic Analysis of Lexical Diversity (TAALED; Kyle et al., 2021), Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Crossley & Kyle, 2018), and Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016) were used to generate NLP measures. All measures related to lexical, cohesion, and syntactic features were screened to select appropriate measures for regression models. In addition, indices related to source citation and quoting were also included because the current investigation was on source-based argumentative writing (Crossley et al., 2023). Indices whose correlation with writing quality outcome was smaller than .2 were dropped before running machine learning models. By selecting variables relevant to the outcome variables, machine learning model accuracy can be improved, and the potential overfitting issue can be alleviated.

Procedures

Directions on how to implement the writing prompts were given to teachers through a professional development session. Digital and paper copies of the directions were provided to all teachers. Students wrote their responses using Google Docs. All essays were processed to align with the requirements of NLP tools before generating the indices: All irrelevant information

(e.g., title, student information) was deleted; all spelling errors were corrected; incorrect use of periods and paragraph breaks were corrected.

Data Analysis Strategy

Three machine learning approaches, random forest, gradient boosting, and support vector machine (SVM) were employed to train regression models for the first research question predicting various dimensions of writing. Traditional statistical methods are rule-based approaches, operate based on assumptions, and have a strong focus on inference. For instance, some traditional statistical models assume the relations between data are linear, polynomial, etc. By contrast, some machine learning methods are not assumption-dependent and are not confined by explicit rules; they set models up to learn and adapt automatically from experience (Ongsulee, 2017). Therefore, machine learning models can analyze data with fewer restrictions and are robust towards data with diverse features due to their strong adaptability. Based on a recent finding using the same analytic coding framework (Steiss et al., 2024), we used the following four dimensions of writing quality: structure, evidence use, historical thinking, and language use. Composite scores of these dimensions were used.

Random Forest is an ensemble machine-learning algorithm that combines a number of tree predictors. In regression tasks, random forest outputs the mean prediction score of all individual trees. This approach is robust against the overfitting of data. Overfitting is a common issue in machine learning models where a model learns the training data too well and becomes too complex to the extent that noise is also picked up. As a result, the trained model cannot generalize well to the test set. In random forest models, each tree is trained independently on a different sub-training dataset. This bootstrapping process can reduce the variance of the model and decrease potential biases. Because of the relative independence of decision trees and each

tree only uses a subset of the data, it is less likely for trees to pick up noise in the data. Thus, random forest is robust in dealing with the potential of overfitting of data. Additionally, the random forest approach can analyze a large number of features without the need to preselect relevant features. Random forest classifier has a built-in feature selection mechanism where a random subset of features is selected at each node to determine the split. This further renders the trees independent from each other and captures distinct patterns in the data without overlearning specific features in the dataset.

Gradient boosting is also a tree ensemble algorithm like random forest. Unlike traditional decision tree models (including random forest), which build trees sequentially and independently, gradient boosting builds trees in a way where each tree corrects the errors made by the previous one. This iterative process allows gradient boosting to gradually improve the model's predictive performance by focusing on the residuals from the previous iterations. By combining multiple weak learners, typically shallow decision trees, into a strong predictive model, gradient boosting can capture complex relations in the data and achieve high predictive accuracy. Moreover, gradient boosting is robust against overfitting due to its regularization techniques, such as shrinkage and tree depth limitation. These techniques made gradient boosting particularly well-suited for handling noisy data and datasets with a large number of features. With its ability to handle both linear and nonlinear relations, gradient boosting is flexible in fitting data with various features.

SVM is also a machine learning model widely used for regression tasks. SVM aims to find the optimal hyperplane that best fits the data points while maximizing the margin. In this way, SVM models can ensure robustness and generalizability. SVM addresses the overfitting issue by seeking a balance between model complexity and generalization. By employing a kernel

function, SVM can map input features into a higher-dimensional space, allowing it to capture nonlinear relations between the features and the target variable. This flexibility enables SVM to handle complex regression tasks. Additionally, SVM offers parameters like the regularization parameter C and kernel parameters, which allow users to exert control over the model's complexity and flexibility to adapt to different features of the data and modeling.

For all models, the k-fold cross-validation method was used to evaluate the performance and stability of the models. Data were randomly split into five folds. Then, four folds of the data were used to train the model, while one fold was used as the test set. The process was repeated iteratively five times, where each fold was used as a test set. Within each iteration, 70% data was used as the training set, and 30% data was used as the test set. Before running machine learning models, linguistic features whose correlation with writing quality scores was smaller than .2 were removed to avoid putting a large number of variables that may potentially be irrelevant into models. Feature scaling was performed. The models' hyperparameters were tuned using the RandomizedSearchCV function, which is a function in scikit-learn that tests a range of hyperparameters and identifies the set of hyperparameters that generated the best results. R-square, mean absolute error (MAE), and standard deviation of MAE were reported. MAE calculates the absolute difference between predicted values and actual values. MAE across five folds were averaged. To understand the stability of the model, standard deviations of MAE were reported for all models. To understand which features are more useful in making predictions, importance values were reported. Importance values show the extent to which features contribute to the prediction of writing quality scores by calculating how often the feature is used to split the data across all trees in the model. Scikit-learn package (version 1.2.2) in Python (version 3.10.12) was used for data analysis.

To answer the second research question (differences in linguistic features as a function of EL status), Bonferroni-adjusted MANCOVA was used with EL status as the group indicator and gender, grade level, and word count as control variables. Word count was added as a covariate because some of the linguistic measures (especially global cohesion and the inflectional morphological complexity) are impacted by text length. Students were examined by three categories: EL, RFEP/IFEP, and EO. SPSS version 28.0 (IBM Corp, 2021) was used for the analysis of this research question. Because multiple measures from the same category (categories include global cohesion, lexical sophistication, lexical diversity, inflectional morphological complexity, and word count) have multicollinearity issues within the same category, measures correlations were examined, and those with a correlation larger than .8 between two linguistic features was regarded as multicollinearity. Among the measures with multicollinearity issues, the linguistic feature with the highest mean correlation with analytic coding scores was retained for the MANCOVA analysis. Other linguistic features with multicollinearity (within the same category) were removed.

Results

RQ1: Linguistic Features Predicting Dimensions of Writing Quality

Model fit information can be seen in Table 2.1. Figures 2.1-2.4 show predicted results and actual results of test sets of the best-fitting model. All models' parameters were tuned. When predicting the structure, evidence use, and historical thinking dimensions of writing quality, random forest regression was the best model. When predicting the language use dimension, gradient boosting's performance was the best. When predicting the structure dimension, the R^2 of the random forest model was .63, indicating that 63% of the variance in the outcome variable was explained by the predictors. MAE of the model was 2.40 ($SD = .14$), indicating the average

of the absolute differences between the predicted values and the actual values was 2.30, and the

standard deviation of MAE showed that the models were relatively stable and performed

consistently. For the evidence use dimension, the model explained 71% variance in the outcome

with an MAE of 2.58 (SD = .19). The best-fitted model for historical thinking fitted similarly

with an explained variance of 70%, and MAE was lower than the previous models (M = 2.13; SD

= .19). Among all models, the language use dimension explained the most variance (75%) and

had the lowest MAE (M = 2.13; SD = .19).

Table 2.1

R-squared, Mean Absolute Error (Standard Deviation in Parentheses) for Regression-Based
Machine Learning Models Predicting Writing Outcomes

Model	Gradient boosting		Random forest		SVM	
	R^2	*MAE* (*SD*)	R^2	*MAE* (*SD*)	R^2	*MAE* (*SD*)
Structure	.62	2.16 (.30)	**.63**	**2.40 (.14)**	.51	2.53 (.09)
Evidence use	.68	2.46 (.62)	**.71**	**2.58 (.19)**	.54	3.14 (.08)
Historical thinking	.44	1.36 (.27)	**.70**	**2.13 (.21)**	.48	2.48 (.03)
Language Use	**.75**	**1.28 (.13)**	.58	1.36 (.29)	.53	2.24 (.04)

Note. Selected models for each writing dimension are bolded.

Figure 2.1 Predicted and Actual Scores of Structure (Random Forest)
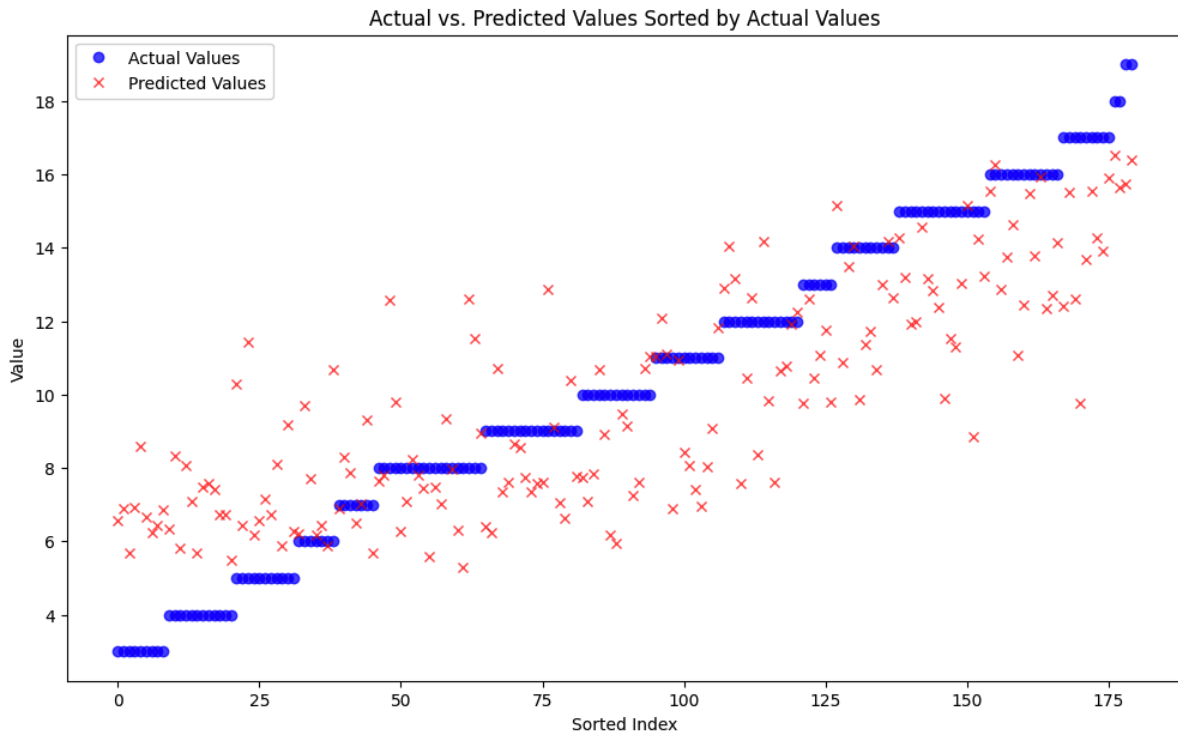
Figure 2.2 Predicted and Actual Scores of Evidence Use (Random Forest)
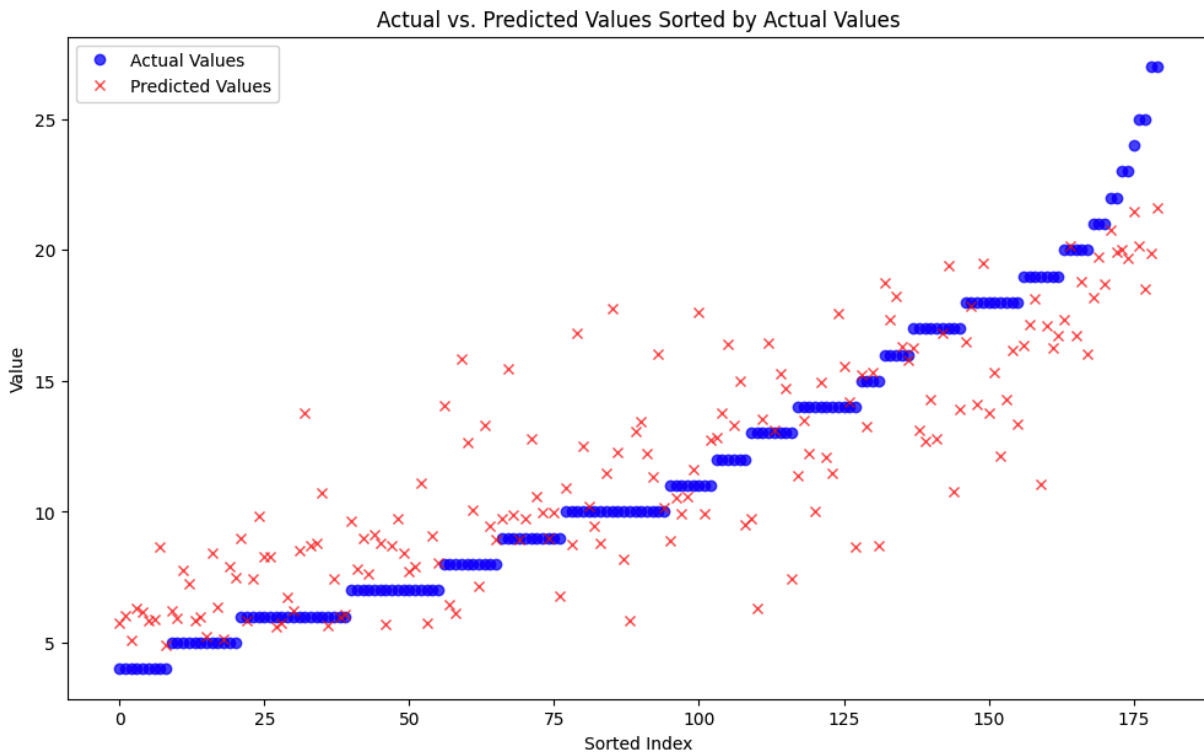
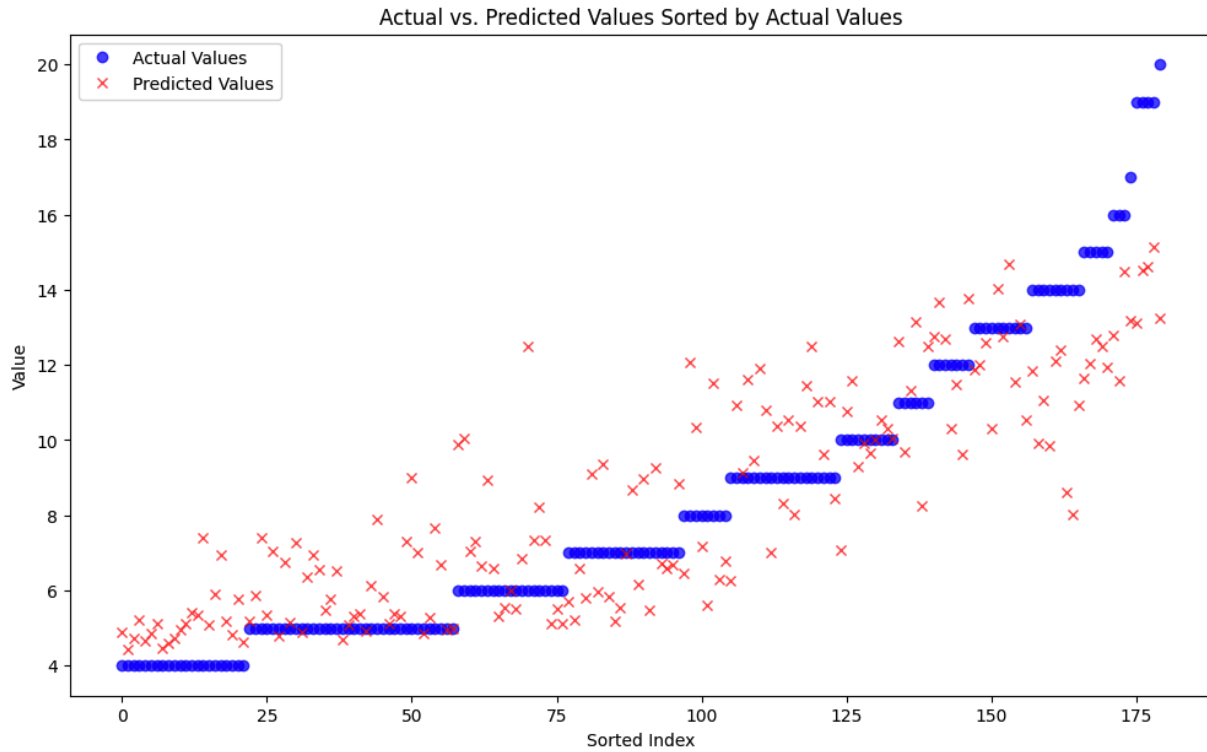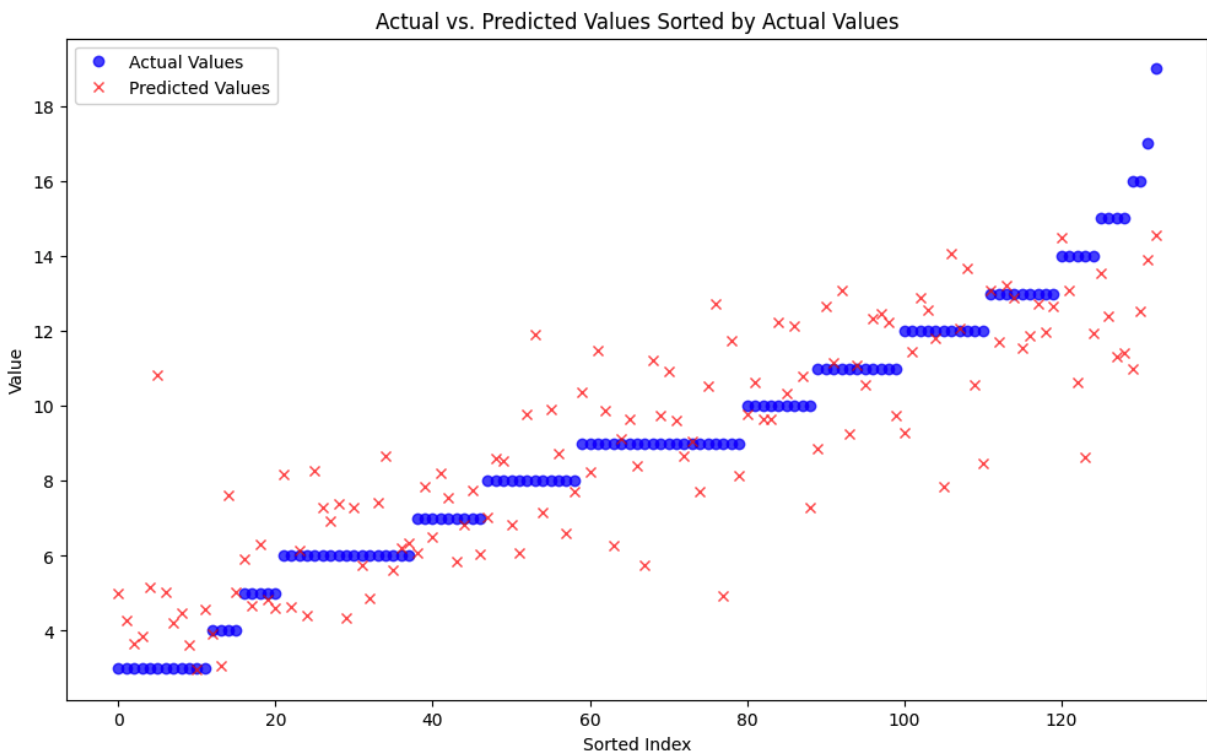Figure 2.3 Predicted and Actual Scores of Historical Thinking (Random Forest)

Figure 2.4 Predicted and Actual Scores of Language Use (Gradient Boosting)

Tables 2-5 present linguistic features with an importance value larger than .01. For importance values, the sum of importance values across all variables was 1. We selected a cutoff of .01 to report the variables that contributed to the prediction the most. However, it is important to note that many other variables still contributed to the modeling to some extent, and we could only represent some of the most important variables. Although random forest and gradient boosting models were not linear, examining the correlations between the linguistic features and outcomes can help us interpret the direction of the relations between linguistic features and writing scores. Thus, correlations were reported in Table 2-5, along with the variable name, feature name, feature category, and the importance value of the feature.

As can be seen in Table 2.2, several global cohesion measures of word overlap were important in predicting the structure dimension in writing quality. In addition to these global cohesion features, a couple of word count features were important in predicting the structure dimension. Other important predictors of the structure dimension included a measure of lexical sophistication and a measure of lexical diversity. All these predictors were positively correlated with the structure dimension of writing quality, except for the lexical diversity measure. The negative correlation was likely caused by the strong impact of text length on type-token ratio (TTR) measures. For the evidence use dimension (see Table 2.3), the pattern was similar: four word count measures, three global cohesion measures that measure word overlap, three lexical sophistication measures, and one lexical diversity measure were important in predicting evidence use scores. All predictors had a positive correlation with the outcome variable, except for a frequency measure of lexical sophistication and a TTR measure.

Important features of the historical thinking (see Table 2.4) dimension were not as similar to the previous two dimensions. When predicting historical thinking, global cohesion played a

less role. Five measures of word count, three measures of lexical sophistication, and one measure of global cohesion were important in predicting historical thinking. Except for two frequency measures, all correlations with the historical thinking score were positive.

For the language use dimension (see Table 2.5), the prediction pattern was distinct. Word count measures played a less important role, whereas many lexical features were crucial in predicting the language use score, and one morphological complexity measure was important. Eight lexical sophistication measures, three word count measures, five lexical diversity measures, one inflectional morphological complexity measure, and one global cohesion measure were important in the prediction. Regarding the correlations with the language use score, TTR, the inflectional morphological complexity measure, and the majority of frequency measures had negative correlations. Two lexical sophistication measures did not have a correlation with the language use score. The rest of the measures all had positive relations to the language score.

RQ2: Language Use Patterns by English Language Proficiency Designations

Based on the results from research question one, fifteen linguistic measures were included in the MANCOVA after the elimination of measures with multicollinearity issues (see Table 2.6): one word count measure, eight lexical sophistication measures, two lexical diversity measures, two global cohesion measures, and one inflectional morphological complexity measure. MANCOVA revealed significant differences across different EL designations in writing performance controlling for gender, grade levels, and text length (Pillai's trace = .09, $F(26, 1140) = 2.13$, $p < .001$, $\eta_p^2 = .05$). Table 2.6 shows mean scores, standard deviations, F value, p value, η_p^2, and results of post-hoc analysis for students across EL designations. Based on values of partial eta-squared, all differences are of small effect sizes (.06 as medium effect size). For lexical sophistication measures, out of the eight measures included, three of them

Table 2.2.

Linguistic Features Predicting Structure

Variable Name	Features	Feature category	Correlation with the structure score	Importance value
adjacent_overlap_2_all_para_div	Adjacent two-paragraph overlap paragraph normed (all lemmas)	Global cohesion (word overlap)	.652***	.182
basic_ntokens	Number of tokens (all words)	Word count	.642***	.161
basic_nfunction_tokens	Number of tokens (function words)	Word count	.635***	.075
adjacent_overlap_2_noun_para_div	Adjacent two-paragraph overlap paragraph normed (noun lemmas)	Global cohesion (word overlap)	.636***	.019
adjacent_overlap_2_cw_para_div_s	Adjacent two-paragraph overlap paragraph normed (content lemmas)	Global cohesion (word overlap)	.642***	.013
kuperman_aoa_aw	Age of Acquisition (all words)	Lexical sophistication	.282***	.012
bigram_lemma_ttr	Bigram lemma TTR	Lexical diversity	-.491***	.011
adjacent_overlap_2_argument_para_div_seg	Adjacent two-paragraph overlap paragraph normed (noun and pronoun lemmas)	Global cohesion (word overlap)	.641***	.010

Note. Only features with importance values larger than .01 were included. *** $p < .001$.

Table 2.3.

Linguistic Features Predicting Evidence Use

Variable name	Features	Feature category	Correlation with the evidence use score	Importance value
basic_ntokens	Number of tokens (all words)	Word count	.712***	.215
basic_nfunction_tokens	Number of tokens (function words)	Word count	.697***	.125
basic_ncontent_tokens	Number of tokens (content words)	Word count	.717***	.065
adjacent_overlap_2_fw_para_div_s	Adjacent two-paragraph overlap function lemmas paragraph normed (function lemmas)	Global cohesion (word overlap)	.650***	.060
adjacent_overlap_2_all_para_div	Adjacent two-paragraph overlap paragraph normed (all lemmas)	Global cohesion (word overlap)	.658***	.039
basic_ntypes	Number of types (all words)	Word count	.725***	.027
adjacent_overlap_fw_para_div_seg	Adjacent paragraph overlap paragraph normed (function lemmas)	Global cohesion (word overlap)	.636***	.017
awl_sublist_1_normed	Number of words in Academic Word List Sublist 1	Lexical sophistication	.249***	.015
oldf_cw	Average log HAL frequency of closest orthographic neighbors (content words)	Lexical sophistication	-.038	.013
coca_academic_lemma_frequency_aw	COCA Academic Frequency (all words)	Lexical sophistication	-.661***	.012
simple_ttr_fw	Type-token ratio (function words)	Lexical diversity	-.641***	.011

Note. Only features with importance values larger than .01 were included. *** $p < .001$.

Table 2.4.

Linguistic Features Predicting Historical Thinking

Variable name	Features	Feature category	Correlation with the historical thinking score	Importance value
basic_nfunction_tokens	Number of tokens (function words)	Word count	.672***	.186
basic_ntokens	Number of tokens (all words)	Word count	.676***	.099
basic_nfunction_types	Number of types (function words)	Word count	.665***	.061
coca_academic_lemma_frequency_fw	COCA Academic Frequency (function words)	Lexical sophistication	-.586***	.039
awl_sublist_1_normed	Number of words in Academic Word List Sublist 1	Lexical sophistication	.268***	.039
basic_ntypes	Number of types (all words)	Word count	.671***	.029
basic_ncontent_tokens	Number of tokens (content words)	Word count	.667***	.015
coca_academic_lemma_frequency_aw	COCA Academic Frequency (all words)	Lexical sophistication	-.588***	.012
adjacent_overlap_fw_para_div_seg	Adjacent paragraph overlap paragraph normed (function lemmas)	Global cohesion (word overlap)	.587***	.012

Note. Only features with importance values larger than .01 were included. *** $p < .001$.

Table 2.5.

Linguistic Features Predicting Language Use

Variable name	Features	Feature category	Correlation with the language use score	Importance value
coca_academic_lemma_frequency_fw	COCA Academic Frequency (function words)	Lexical sophistication	-.698***	.125
basic_ntokens	Number of tokens (all words)	Word count	.750***	.115
simple_ttr_aw	Type token ratio (all words)	Lexical diversity	-.630***	.049
basic_ntypes	Number of types (all words)	Word count	.807***	.047
simple_ttr_fw	Type token ratio (function words)	Lexical diversity	-.722***	.046
coca_academic_lemma_frequency_aw	COCA Academic Frequency (all words)	Lexical sophistication	-.781***	.036
basic_nfunction_tokens	Number of tokens (function words)	Word count	.724***	.030
coca_academic_lemma_frequency_cw	COCA Academic Frequency (content words)	Lexical sophistication	-.637***	.025
nns_raw_range_high_fw_log	Log-transformed TOEFL11 High Proficiency Range (function words)	Lexical sophistication	-.071	.021
nns_raw_freq_wc_fw	TOEFL11 Frequency (function words)	Lexical sophistication	.080	.019

kuperman_aoa_aw	Age of Acquisition (all words)	Lexical sophistication	.315***	.016
inflectionalmci10	Inflectional mean complexity index	Inflectional morphological complexity	-.664***	.016
subtlexus_range_fw_log	SUBTLEXus Range Logarithm (function words)	Lexical sophistication	-.265***	.016
adjacent_overlap_all_para	Adjacent paragraph overlap (all lemmas)	Global cohesion (word overlap)	.440***	.015
mtld_original_cw	MTLD (content word)	Lexical diversity	.443***	.014
aoe_index_above_threshold_40	LDA Age of Exposure with .40 cosine threshold	Lexical sophistication	.236***	.013
noun_ttr	Type token ratio (noun)	Lexical diversity	-.498***	.012
hdd42_fw	Lexical diversity based on Hypergeometric distribution (function words)	Lexical diversity	.533***	.012

Note. Only features with importance values larger than .01 were included. *** $p < .001$.

Table 2.6.

Mean, Standard Deviations, and Multivariate Analysis of Covariance on Selected Linguistic Features Across English Designation Groups

Linguistic Features	Linguistic Category	EL (n = 140)		RFEP&IFEP (n = 242)		EO (n = 204)		$F(2, 581)$	p	η_p^2	Pairwise comparison
		M	SD	M	SD	M	SD				
LDA Age of Exposure with .40 cosine threshold	Lexical sophistication	1.89	.57	2.06	.54	2.03	.52	1.97	.14	.007	n.s.
Number of words in Academic Word List Sublist 1	Lexical sophistication	.01	.01	.01	.01	.01	.01	.69	.50	.002	n.s.
COCA Academic Frequency (all words)	Lexical sophistication	3993.57	1432.02	3223.93	1311.89	3296.31	1432.57	8.08	<.001	.027	EL > EO = RFEP/IFEP
Age of Acquisition (all words)	Lexical sophistication	5.25	.37	5.39	.31	5.37	.31	3.86	.02	.013	RFEP/IFEP > EL
TOEFL11 Frequency (function words)	Lexical sophistication	21210.63	2991.73	21333.47	2575.38	21471.36	2698.71	.34	.71	.001	n.s.
Log-transformed TOEFL11 High Proficiency Range (function words)	Lexical sophistication	1.88	.04	1.87	.03	1.87	.04	1.25	.29	.004	n.s.
Average log HAL frequency of closest	Lexical sophistication	7.75	.19	7.72	.17	7.73	.18	1.24	.29	.004	n.s.

orthographic neighbors
(content words)

Measure	Construct										
SUBTLEXus Range Logarithm (function words)	Lexical sophistication	3.91	.01	3.90	.01	3.91	.01	5.69	.004	.019	EL > RFEP/IFEP
Lexical diversity based on Hypergeometric distribution (function words)	Lexical diversity	.43	.19	.48	.15	.47	.17	1.79	.17	.006	n.s.
MTLD (content word)	Lexical diversity	62.28	42.04	72.37	37.90	73.32	39.56	.65	.52	.002	n.s.
Adjacent two-paragraph overlap (all lemmas)	Global cohesion (word overlap)	.20	.18	.26	.16	.26	.16	13.03	<.001	.043	EO = RFEP/IFEP > EL
Adjacent paragraph overlap (all lemmas)	Global cohesion (word overlap)	9.72	13.02	18.79	16.54	20.30	16.31	3.40	.03	.012	EO > EL
Inflectional mean complexity index	Inflectional Morphological complexity	-.33	.49	-.50	.43	-.50	.42	3.71	.03	.013	EL > EO = RFEP/IFEP

Note. EO = English-only, IFEP = Initial Fluent English Proficiency, RFEP = Redesignated Fluent English Proficiency, EL = English Learners. Bonferroni correction was used for pairwise comparisons. n.s. = no significant difference.

showed differences across groups. One measure that showed difference is COCA academic frequency of all words. For this measure, students designated as EL used words of higher frequency than RFEP/IFEP ($d = .41$; $p < .001$) and EO students ($d = .37$; $p = .003$). One measure that also showed a difference was age of acquisition of all words: students designated as RFEP/IFEP used words that tend to be acquired later than students designated as EL ($d = .22$; $p = .02$). Another lexical sophistication measure, SUBTLEXus range logarithm of function words, which describes the number of documents that a word occurs in, describes the word range information. For this measure, students designated as EL used words that appeared in more documents than RFEP/IFEP students ($d = .21$; $p = .003$). Both of the two global cohesion measures that measure word overlap also showed significant group differences. For adjacent two-paragraph overlap of all lemmas, it measures the number of lemma types that occur at least once in the next two paragraphs. Students designated as EL had fewer words that overlapped in the next two paragraphs than RFEP/IFEP students ($d = .44$; $p < .001$) and EO students ($d = .55$; $p < .001$). For the other global cohesion measure, adjacent paragraph overlap of all lemmas, students designated as EL had fewer words that overlapped in the adjacent paragraph compared to EO students ($d = .26$; $p = .02$). Inflectional morphological complexity is defined as mean average diversity of inflectional morphemes. Students designated as EL had a higher diversity of inflectional morphemes than RFEP/IFEP students ($d = .26$; $p = .04$) and EO students ($d = .27$; $p = .04$). No significant differences were found for other linguistic measures.

Discussion

This study examined the predictive accuracy of machine learning models in using NLP linguistic features to predict writing quality across different dimensions, including structure, evidence use, historical thinking, and language use, with middle and high school students.

Utilizing machine learning approaches, our analysis revealed that linguistic features were significant predictors of various dimensions of writing quality. We found that the machine learning approach can explain a higher amount of variance in writing quality than traditional linear regression models, indicating that machine learning models have the potential to predict multiple dimensions of writing quality scores more accurately. The explained variance of the dimensions ranged from 63% to 75%. This highlighted the potential of using NLP linguistic features to predict students' writing. Furthermore, we investigated how linguistic features varied among bilingual and monolingual students of different English proficiency levels. The results underscored differences in linguistic features between bilingual students designated as EL and monolingual students designated as proficient in English (including RFEP, IFEP, and EO students), highlighting the profound role of language proficiency in writing performance.

The results of this study provide empirical support for multiple theoretical models of writing that highlight the significance of language skills and linguistic features in writing proficiency. Consistent with the DIEW model (Y.-S. G. Kim & Graham, 2022; Y.-S. G. Kim & Park, 2019) and the Writer(s)-Within-Community Model of Writing (Graham, 2018), our findings highlight the pivotal role of language use in writing quality. These findings are in line with previous research (e.g., Crossley et al., 2019; Kyle & Crossley, 2016; Wang et al., 2024) that has identified relations between linguistic features and the evaluation of writing quality, further solidifying the argument that linguistic proficiency is a foundational component of successful writing.

Building on previous studies that employed traditional linear regression models to study the extent to which linguistic features can predict writing quality (e.g., Crossley & McNamara, 2011a, 2012; Crossley et al., 2014; Guo et al., 2013; June et al.,2019; M. Kim & Crossley, 2018;

Lee et al., 2021), the current study found that machine learning models can explain a higher variance in writing quality scores than previous studies that examined relations of linguistic features in writing with writing quality using linear regression models. Despite the large number of variables used in the current study, these variables were generated from several NLP software. Unlike traditional linear regression models that require a large amount of preprocessing to avoid collinearity issues, machine learning models have built-in features to deal with a large number of variables and high correlations between some of the measures. Overall, machine learning models are efficient, allow us to deal with a large number of variables, and can explain a higher amount of variance in predictors. In addition, the explained variance for evidence use and historical thinking dimensions was around 70%, which further indicates that linguistic features are able to explain a large amount of variance in writing quality dimensions related to discipline and genre features.

In addition, the majority of previous studies used holistic scores to operationalize writing quality. The current study showed the potential of using machine learning models and NLP linguistic features to predict genre- and discipline-specific writing quality features. On the one hand, this finding suggests the important role language use plays in displaying genre- and discipline-specific writing skills. On the other hand, the finding can inform writing assessment -- using already existing NLP linguistic features to assess students' writing quality in multiple dimensions, including those that are specific to genres and disciplines.

While global cohesion measures that measure word overlap are more predictive of structure and evidence use dimensions, lexical features (lexical sophistication and lexical diversity) played a more crucial role in predicting the historical thinking and language use dimensions. One reason for global cohesion measures to be predictive of structure and evidence

use dimensions may be the importance of cohesion in these two aspects. Argumentative writing with higher global cohesion is more likely to discuss similar and critical topics that contribute to the main claim of the essay. Another reason for global cohesion measures to be more predictive of the structure and evidence use dimensions may be due to global cohesion measures' ability to capture elaborateness in text and paragraph breaks. For students who did not divide paragraphs appropriately, their global cohesion would be deemed lower due to a lack of paragraph breaks. Global cohesion measures that predicted writing scores in this study are those that measure word overlap across paragraphs, which also capture the extent to which students were able to elaborate on the content within paragraphs. Students who were able to elaborate may have a higher chance of having overlapped words across paragraphs.

Another interesting pattern is the predictive role of words used in an academic context and words of high proficiency. Measures related to more sophisticated and academic words are more important in predicting evidence use, historical thinking, and language use dimensions. This indicates the importance of academic vocabulary in performing discipline- and genre-related goals in the writing process (Durrant, 2014). The role of inflectional morphological complexity played in predicting language use scores indicates that the measure may be able to capture grammatical accuracy, which was captured in the language use scores. Overall, the results of our study indicated that lexical features (sophistication and diversity) and global cohesion features are stronger predictors of writing quality than syntactic features and local cohesion features, which aligns with previous research (Crossley et al., 2016; Crossley & McNamara, 2011a, 2012, 2016; Cumming et al., 2006; Guo et al., 2013; Kyle & Crossley, 2016; MacArthur et al., 2019; McNamara et al., 2010).

Our findings also indicate that bilingual students with limited language proficiency (ELs) used fewer sophisticated words than bilingual students proficient in English and monolingual students (including students designated as RFEP, IFEP, and EO). It is worth noting that no difference in lexical sophistication was found between bilingual students designated as proficient in English and monolingual writers. This indicates that bilingual students with proficient English and experiences in multiple languages perform equivalently with their monolingual peers in linguistic abilities (in our case, lexical sophistication features in writing). It is also worth noting that out of the three lexical sophistication measures, two of them showed differences between the EL and RFEP/IFEP, not between EL and EO groups. This aligns with prior literature; students who are classified as EL have lower English vocabulary size, which results in difficulties in acquiring sophisticated and advanced disciplinary vocabulary compared to students with adequate English proficiency (Carlo et al., 2004). The smaller English lexicon in EL students results in difficulties in using contextual strategies in vocabulary learning and applying complex words (Stoller & Grabe, 2000). Although no statistically significant difference was found in lexical sophistication between bilingual proficient in English and monolingual students in the current study, future research is warranted in further investigating this topic.

In addition, ELs had lower global cohesion of word overlap across paragraphs than students proficient in English. Global cohesion measures whether the content and overall topics across paragraphs are coherent. This could be related to the lack of elaboration within paragraphs because ELs had shorter text lengths or lack of coherence across paragraphs. Writing instruction that supports students to elaborate more on their arguments or to connect their content better in supporting a central claim can improve global cohesion.

Another measure in which ELs performed differently was inflectional morphological complexity, which measures the mean diversity of inflectional morphemes. ELs had significantly higher values than bilingual students proficient in English and monolingual students. This could be due to the nature of the writing task. The task is an argumentative task contextualized in a historical event. Thus, the argument writing should be consistently in the past tense, and the diversity of inflectional morphemes may indicate that ELs did not use past tense consistently, which led to grammatical errors. This is evidenced by the strong negative correlation of inflectional morphological complexity with language use score. This is an interesting pattern and indicates specific patterns of linguistic features in specific disciplines and topics. Meanwhile, the results of our study may indicate that ELs may need support with consistent tense use in their essay, especially for writing tasks that are contextualized in an event already happened. However, because the current study is correlational, implications related to writing instruction are only suggestive.

Limitations and Future Directions

One limitation is the relatively small sample size, given that machine learning models were used for data analysis. Future studies should replicate the study using larger sample sizes to examine whether predictive accuracy can improved.

The current study showed that machine learning models can explain larger variances in writing quality than the traditional linear regression approach used in previous studies. Machine learning models can also deal with a large number of variables more efficiently, which is more suitable for conducting analyses with NLP linguistic features. Further research is warranted to test the application of machine learning models in analyzing student writing and linguistic features. In addition, the findings of the study indicate that lexical and global cohesion features

are more predictive of writing quality than syntactic features. This indicates that complex syntactic structures may not be a necessity in achieving high-quality writing (Crossley et al., 2014). Such findings can inform automatic essay scoring on how to select linguistic features in predicting students' writing performance. It is especially critical in selecting appropriate linguistic features for different aspects of writing outcomes, with the consideration of genre- and discipline-specific features of the writing task. In addition, although our study does not examine the causal relation between linguistic features and writing quality, writing instruction that prepares students with sufficient lexical and cohesion knowledge in language may potentially lead students to be better prepared for the linguistic demands they encounter in writing.

Our study also addresses a critical gap in understanding bilingual students' language use in writing. The results indicated ELs need to learn to use more sophisticated vocabulary and make their written texts more coherent. Furthermore, bilingual students designated as proficient in English performed on par with monolingual students on all linguistic features examined. The study highlights the need to examine bilingual students as a heterogeneous population and to understand language development of different groups of bilingual learners. Our results further highlight the importance of high-quality instruction focusing on language use for students with limited language proficiency in improving writing performance.

References

Ardasheva, Y., Tretter, T. R., & Kinny, M. (2012). English language learners and academic achievement: Revisiting the threshold hypothesis. *Language Learning*, 62(3), 769-812. https://doi.org/10.1111/j.1467-9922.2011.00652.x

Bacha, N. (2001). Writing evaluation: What can analytic versus holistic essay scoring tell us?. *System*, *29*(3), 371-383. https://doi.org/10.1016/S0346-251X(01)00025-2

Bamberg, B. (1983). What makes a text coherent?. *College composition and communication*, 34(4), 417-429. https://doi.org/10.58680/ccc198315261

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing*, 24(2), 183-202. https://doi.org/10.1007/s11145-010-9264-9

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of learning disabilities*, 35(1), 39-56. https://doi.org/10.1177/002221940203500104

Berninger, V. W., & Winn, W. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. *Handbook of Writing Research*, 96-114.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., ... & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*(2), 188-215.

Crossley, S., Heintz, A., Choi, J. S., Batchelor, J., Karimi, M., & Malatinszky, A. (2023). A large-scaled corpus for assessing text readability. *Behavior Research Methods*, *55*(2), 491-507. https://doi.org/10.3758/s13428-022-01802-x

Crossley, S. A., & Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing*, 38, 46-50. https://doi.org/10.1016/j.asw.2018.06.004

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods*, 51(1), 14-27. https://doi.org/10.3758/s13428-018-1142-4

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. https://doi.org/10.3758/s13428-015-0651-7

Crossley, S., & McNamara, D. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).

Crossley, S., & McNamara, D. (2011a). Text coherence and judgments of essay quality: Models of quality and coherence. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Crossley, S., & McNamara, D. (2011b). Text coherence and judgments of essay quality: Models of quality and coherence. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

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. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., & McNamara, D. S. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, 7(3), 351-370. https://doi.org/10.17239/jowr-2016.07.03.02

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. https://doi.org/10.1016/j.jslw.2014.09.006

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2011). Predicting human scores of essay quality using computational indices of linguistic and textual features. In *International Conference on Artificial Intelligence in education* (pp. 438-440). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21869-9_62

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

Culham, R. (2003). *Six Plus One Traits of Writing*. Scholastic Inc..

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & Jamse, M. (2005). Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL®. *ETS Research Report Series*, 2005(1), i-77. https://doi.org/10.1002/j.2333-8504.2005.tb01990.x

Cumming, A., Lai, C., & Cho, H. (2016). Students' writing from sources for academic purposes: A synthesis of recent research. *Journal of English for Academic Purposes*, 23, 47-58. https://doi.org/10.1016/j.jeap.2016.06.002

Dockrell, J. E., Lindsay, G., & Connelly, V. (2009). The impact of specific language impairment on adolescents' written text. *Exceptional children*, 75(4), 427-446. https://doi.org/10.1177/001440290907500403

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied linguistics*, 35(3), 328-356. https://doi.org/10.1093/applin/amt016

Fang, Z., & Coatoam, S. (2013). Disciplinary literacy: What you want to know about it. *Journal of Adolescent & Adult Literacy*, 56(8), 627-632. https://doi.org/10.1002/JAAL.190

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, 28(2), 414-420. https://doi.org/10.2307/3587446

Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests (Vol. 7, No. 1, pp. 47-84). *Spaan Fellow Working Papers in Second or Foreign Language Assessment*.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C., Shanahan, C., & Project READi. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219-246.

Goldman, S. R., & Scardamalia, M. (2013). Managing, understanding, applying, and creating knowledge in the information age: Next-generation challenges and opportunities. *Cognition and Instruction*, 31(2), 255-269.

Graham, S. (2018). A revised writer (s)-within-community model of writing. *Educational*

    Psychologist, 53(4), 258-279. https://doi.org/10.1080/00461520.2018.1481406

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.

    https://doi.org/10.1016/j.asw.2013.05.002

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

Hwang, J. K., Lawrence, J. F., Collins, P., & Snow, C. (2017). Vocabulary and reading

    performances of redesignated fluent English proficient students. *Tesol Quarterly*, 51(4),

    757-786. https://doi.org/10.1002/tesq.346

Hyland, K. (2006). The 'other'English: thoughts on EAP and academic writing. *The European*

    English Messenger, *15*(2), 34-38.

Jagaiah, T., Olinghouse, N. G., & Kearns, D. M. (2020). Syntactic complexity measures:

    variation by genre, grade-level, students' writing abilities, and writing quality. *Reading*

    and Writing, 33(10), 2577-2638. https://doi.org/10.1007/s11145-020-10057-x

Jones, S., Myhill, D., & Bailey, T. (2013). Grammar for writing? An investigation of the effects

    of contextualised grammar teaching on students' writing. *Reading and Writing*, *26*, 1241-

    1263.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of

    children in first and second grade. *Journal of Educational Psychology*, 78(4), 243.

    https://doi.org/10.1037/0022-0663.78.4.243

Jung, Y., Crossley, S., & McNamara, D. (2019). Predicting second language writing proficiency in learner texts using computational tools. *Journal of Asia TEFL*, *16*(1), 37. https://doi.org/10.18823/asiatefl.2019.16.1.3.37

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. https://doi.org/10.1016/j.asw.2018.03.002

Kim, Y. -S. G., & Graham, S. (2022). Expanding the Direct and Indirect Effects Model of Writing (DIEW): Reading–writing relations, and dynamic relations as a function of measurement/dimensions of written composition. *Journal of Educational Psychology*, 114(2), 215. https://doi.org/10.1037/edu0000564

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. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12-24. https://doi.org/10.1016/j.jslw.2016.10.003

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, 18(2), 154-170. https://doi.org/10.1080/15434303.2020.1844205

Lee, C., Ge, H., & Chung, E. (2021). What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong. *System*, *97*, 102461. https://doi.org/10.1016/j.system.2021.102461

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing

    research and implications for writing assessment. *Language Testing*, 34(4), 493-511.

    https://doi.org/10.1177/0265532217710675

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(6), 1553-1574.

    https://doi.org/10.1007/s11145-018-9853-6

McCulley, G. A. (1985). Writing quality, coherence, and cohesion. *Research in the Teaching of*

    *Englis*h, 269-282. https://doi.org/10.58680/rte198515640

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing

    quality. *Written Communication*, 27(1), 57-86.

    https://doi.org/10.1177/0741088309351547

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an

    intelligent writing strategy tutoring system. *Behavior Research Methods*, 45(2), 499-515.

    https://doi.org/10.3758/s13428-012-0258-1

Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of

    adolescents' writing. *The Journal of the Learning Sciences*, *19*(4), 539-568.

Myhill, D. (2008*).* Towards a linguistic model of sentence development in writing. *Language*

    and Education, 22(5), 271-288. https://doi.org/10.1080/09500780802152655

National Writing Project. (2010). *The Analytic Writing Continuum: A comprehensive writing*

    assessment system.

Ongsulee, P. (2017, November). Artificial intelligence, machine learning and deep learning. In 2017 *15th international conference on ICT and knowledge engineering (ICT&KE)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICTKE.2017.8259629

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518. https://doi.org/10.1093/applin/24.4.492

Perin, D., & Lauterbach, M. (2018). Assessing text-based writing of low-skilled college students. *International Journal of Artificial Intelligence in Education*, 28(1), 56-78. https://doi.org/10.1007/s40593-016-0122-z

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. (2008). Assessing the Microstructure of Written Language Using a Retelling Paradigm. *American Journal of Speech-Language Pathology*, 17(2), 107-120. https://doi.org/10.1044/1058-0360(2008/012)

Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing*, 33, 3-17. https://doi.org/10.1016/j.jslw.2016.06.001

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

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter?. *Topics in Language Disorders*, 32(1), 7-18. https://doi.org/10.1097/TLD.0b013e318244557a

Slama, R. B. (2014). Investigating whether and when English learners are reclassified into mainstream classrooms in the United States: A discrete-time survival analysis. *American*

Educational Research Journal, 51(2), 220-252.

https://doi.org/10.3102/0002831214528277

Steiss, J., Krishnan, J., Kim, Y. S. G., & Olson, C. B. (2022). Dimensions of text-based analytical writing of secondary students. *Assessing Writing*, 51, 100600. https://doi.org/10.1016/j.asw.2021.100600

Steiss, J., Wang, J., Kim, Y.-S. G, Olson. C. B. (2024). Challenges and Dimensions of Secondary Students' Source-Based Argument Writing in History. *Written Communication*.

Stoller, F., & Grabe, W. (2000). Implications for L2 vocabulary acquisition and instructionfrom L1 vocabulary research. *Second Language Reading and Vocabulary Learning. Eds. T. Huckin, M. Haynes & J. Coady.(24-45). Norwood: Ablex Publishing*.

Taghipour, K. (2017). *Robust trait-specific essay scoring using neural networks and density estimators* (Doctoral dissertation, National University of Singapore (Singapore)).

Van Dijk, T. A. (1977). Semantic macro-structures and knowledge frames in discourse comprehension. *Cognitive Processes in Comprehension*, 332, 3-31.

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing*, 24(2), 203-220. https://doi.org/10.1007/s11145-010-9266-7

Wan, Q., Crossley, S., Banawan, M., Balyan, R., Tian, Y., McNamara, D., & Allen, L. (2021). Automated Claim Identification Using NLP Features in Student Argumentative Essays. *International Educational Data Mining Society*.

Wang, J., Kim, Y. S. G., & Cho, M. (2024). Linguistic features in narrative and opinion genres and their relations to writing quality in fourth grade writing. *Journal of Research in Reading*, *47*(2), 220-239.

Williams, C., & Lowrance-Faulhaber, E. (2018). Writing in young bilingual children: Review of research. *Journal of Second Language Writing*, *42*, 58-69. https://doi.org/10.1016/j.jslw.2018.10.012

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73. https://doi.org/10.1037/0022-0663.83.1.73

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, 32(2), 189-204. https://doi.org/10.2307/356693

Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*, 23(1), 31-48. https://doi.org/10.1016/j.linged.2011.09.00

CHAPTER 3

Writing Self-Efficacy of Linguistically Diverse Students and its Relations with Dimensions of Writing Quality

Self-efficacy refers to one's perception of their ability to "organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). Self-efficacy in writing (writing self-efficacy hereafter) is crucial in coordinating the writing process (Graham, 2018; Kim & Graham, 2022; Kim & Park, 2019). Studies have illustrated that students with higher writing self-efficacy are more likely to feel motivated towards writing, display more self-regulated behaviors (Ekholm et al., 2015), and spend more effort in writing (Garcia-Sanchez & Fidalgo-Redondo, 2006; Pajares et al., 2007).

Previous studies have confirmed that writing self-efficacy is a multidimensional construct, which includes ideation, self-regulation, writing conventions, writing tasks, and revision (Bruning et al., 2013; Sun et al., 2022; Wang et al., in preparation). Although previous studies have confirmed that there are dissociable aspects in writing self-efficacy, one gap in the research is the lack of understanding of how students vary in different aspects of writing self-efficacy. With an understanding of which aspects students are more or less confident in, educators can target specific aspects of writing instruction. A related gap in the literature is whether the dimension(s) of writing self-efficacy is similarly or differentially related to various dimension(s) of writing quality. The lack of this knowledge makes it challenging to understand which dimension(s) of writing self-efficacy is more crucial in students' writing performance. Because writing self-efficacy is both an antecedent and consequence of writing success (Bruning et al., 2013; Graham et al., 2024; Wang et al., in preparation), understanding the nuanced relation

between writing self-efficacy and writing quality can inform writing interventions and instructions.

Another gap in the literature is writing self-efficacy in linguistical minority (LM) students (students who are exposed to languages other than English at home). Students' language and literacy skills differ due to their different experiences in both school and home contexts (de Jong, 2004; Williams & Lowrance-Faulhaber, 2018). Investigations of the writing skills of LM students tend to focus on viewing them as a homogeneous group and use a strict cutoff to determine individual students' language proficiency as either proficient or not. In the current study, we distinguish between students that are originally fluent (i.e., initially fluent English proficient, IFEP) when they entered school and students designated as Reclassified Fluent English Proficient (RFEP), who did not reach the proficiency level when they first entered school but met the criteria later to opt out of English language (EL) service. By doing so, we recognized diversity within the LM population.

The current paper addresses three critical gaps in the literature: (1) differences in dimensions of writing self-efficacy among secondary students, (2) the relation between dimensions of writing self-efficacy and dimensions of argumentative writing quality in history, and (3) differences in writing self-efficacy across LM and English Only (EO) students with different English language designations in the US context.

Dimensions of Writing Self-efficacy

Based on self-efficacy theory and the theory of social cognitive learning (Bandura, 1986), self-efficacy influences how much effort one puts into certain tasks and how people navigate the self-regulation process. Because self-regulation is critical in the writing process, where writers

need to constantly and strategically manage their cognitive resources, thoughts and ideas, behaviors, and the environment to accomplish specific writing goals, writers' self-efficacy and self-regulation are important so that writers can be motivated enough to navigate the complex writing process (Bereiter & Scardamalia, 1987). During the writing process, writers need to plan based on the requirements of the writing task, monitor what they have written, diagnose areas of improvement, and revise their writing accordingly (Flower & Hayes, 1986). Failure to allocate limited cognitive resources during the writing process can adversely impact students' writing performance (Graham, 2018). Writing models have hypothesized that writing self-efficacy is critical to writing performance. The Direct and Indirect Effects Model of Writing (DIEW; Kim & Graham, 2022; Kim & Park, 2019) recognizes the roles of socio-emotional aspects in writing, including writing self-efficacy, writing self-concept, attitude toward writing, and motivation, positing a bidirectional relation between socio-emotional aspects and writing development. Furthermore, the Writer(s)-Within-Community Model of writing (WWC; Graham, 2018) emphasizes the important role of long-term memory, which consists of knowledge and beliefs that are tapped into during the writing process or play a key role in learners' writing development. According to the WWC model, writers' beliefs impact and determine how they interact with the texts they write: writers' beliefs about their ability to accomplish writing tasks impact the effort they invest in the writing process, the tools and resources they apply, and the extent to which they interact with their peers and teachers to improve their written compositions. Writing self-efficacy is one of the key beliefs that motivate writers to write and overcome challenges in developing writing skills.

Various aspects of writing self-efficacy have been examined in previous research. One aspect that has received much attention is self-efficacy in writing conventions and language

skills, which has been examined in early and current research (Bruning et al., 2013; McCarthy et al., 1985; Pajares, 2007; Pajares & Johnson, 1994; Shell et al., 1989; Sun et al., 2022). Another aspect that received much attention is the writing task aspect, which refers to students' ability to write across multiple tasks and genres (Pajares & Johnson, 1994; Shell et al., 1989; Sun et al., 2022). Ideation and self-regulation aspects have been examined more frequently using Bruning and colleagues' (2013) framework. This framework, unlike previous frameworks to assess writing self-efficacy, systematically reflects writing theory in its factor construction. As Bruning and colleagues (2013) have mentioned, idea translation and language skills have been emphasized in writing models and empirical studies that investigate the relations between language and writing outcomes (Bereiter & Scardamalia, 1987; Flower & Hayes, 1981; Graham, 2018; Kim & Graham, 2022; Schleppegrell, 2007; Wang et al., 2024). Additionally, given that strategy-related writing interventions have been shown to be the most effective among various populations (e.g., Gillespie & Graham, 2014; Graham et al., 2003), the ability to self-regulate and use strategies is crucial to success in writing. Due to the current writing theories and empirical studies, three factors were included in this framework, and this was reported as the most commonly used self-efficacy scale in writing research (Graham et al., 2024).

Furthermore, only recently, research started to examine revision (Chen & Zhang, 2019; Chung et al., 2021; Wang et al., in preparation), which is a key component in the writing process given the problem-solving nature of the writing process (Bereiter & Scardamalia, 1987). In addition, revision is not only an error correction and editing process but a contextualized meaning-making activity that is central to the writing process and writing outcomes (Czerniewska, 1992; Hayes, 2005; Holliway & McCutchen, 2005). However, secondary students tend to only make a few substantive revisions during their writing process, and the majority of

revisions in writing are error corrections (Fitzgerald, 1987; MacArthur & Graham, 2016). Thus, in the current study, we adopt a revised framework based on Bruning and colleagues' (2013) framework that also assesses students' writing self-efficacy in revision (Wang et al., in preparation). Although examinations of multiple dimensions of writing self-efficacy are common, we do not know of any studies that explicitly compare which dimension(s) present more challenges for students.

Relations Between Writing Self-efficacy and Writing Quality

In accordance with theoretical models of writing noted in the previous section, empirical studies have found that writing self-efficacy shows a positive relation with writing quality (e.g., Bruning et al., 2013; Graham et al., 2019; Klassen, 2002; Pajares, 2003; Pajares & Johnson, 1994) and writing productivity (Graham et al., 2017). Many studies have examined the relations between dimensions of writing self-efficacy and writing performance. For instance, Bruning and colleagues (2013) found that the writing conventions dimension of self-efficacy of US high school students had a relatively higher correlation with self-reported writing performance than the ideation and self-regulation dimensions. Erkan and Saban (2011) found that self-efficacy in the design-unity and accuracy dimensions have a relatively higher correlation with writing achievement than writing self-efficacy in content and punctuation for a group of students learning English in a foreign language (EFL) context. Sanders-Reio and colleagues (2014) worked with undergraduate students in the US and found that dimensions of writing self-efficacy (self-regulation, mechanical, and substantive) were related to writing performance similarly ($r =$ .15-.23), with self-regulation having the weakest relation. Teng and Wang (2018), working with college students in an EFL context, examined five dimensions of writing self-efficacy (including information organization, linguistic knowledge, rehearsal, and memory, self-regulatory, and

writing performance efficacy). Using correlation and linear regression analysis, the authors found that these dimensions were related to writing quality similarly, with information organization having a relatively higher association ($r = .60$). Overall, findings are mixed regarding which dimension(s) of writing self-efficacy is more predictive of student writing. In addition, the majority of the studies examining the relations between dimensions of writing self-efficacy and writing performance used correlation or linear regression analysis and did not use structural equation modeling (SEM). Using SEM allows the dimensions to be more accurately represented with smaller measurement errors and multiple dimensions to be examined in one model at once. On top of the mixed findings and limited analysis approaches used in previous research, various dimensions of writing self-efficacy have been examined in previous studies, and the naming/categorization of the items is also inconsistent across studies. As a result, it remains unclear which dimension(s) of writing self-efficacy would be most predictive of writing performance when using a validated writing self-efficacy scale.

Multiple writing models have emphasized the importance of understanding genre- and discipline-specific features in writing proficiency (Graham et al., 2018; Kim & Graham, 2022). For example, the WWC model (Graham et al., 2018) conceptualizes writers, the writing process, and writing skills within the context of writing communities. Writers belong to multiple communities, where communities can be defined as groups of people who share the same goals and assumptions in achieving certain purposes in their writing. Certain purposes, strategies, tools, and actions of writing are thus key to being a successful writer in a particular writing community. The task examined in the current study is a source-based argumentative writing task in history class. On top of the complex writing process, writing an argumentative text in history poses genre- and disciplinary-specific demands on students. The writing community of history

discipline requires students to comprehend and integrate multiple perspectives (Cho et al., 2022), establish strong relations between claims and evidence, and critically integrate those perspectives into their writing (Goldman et al., 2016; Rowan & White, 2022).

Because of the complexity associated with writing in this task, writing quality is a multidimensional construct consisting of multiple associated but distinct aspects (e.g., macrostructure, spelling and writing conventions, productivity; Kim et al., 2014; Kim & Graham, 2022; Puranik et al., 2009; Steiss et al., 2023). Evaluating writing quality using an analytic approach instead of a holistic approach can depict a more accurate picture of students' written composition. Previous research indicates that holistic scoring tends to draw scorers' attention to the strengths of the texts, and the weaknesses are usually masked using this approach (Wiseman, 2012). While holistic scoring has the benefit of being economical, it cannot capture the nuanced features of writing development. Using an analytic coding approach, Bacha (2001) found that although correlations between the dimensions (content, organization, vocabulary, language, and mechanics) were high, there were significant differences in students' performance across different aspects for EFL freshman students, indicating the value of examining multiple dimensions of writing quality. In another study, Kim and colleagues (2014) found that content, organization, vocabulary, and sentence fluency are best described as a single construct that is related but dissociable from writing conventions, writing productivity, and writing syntactic complexity for children in Grade 1. Common analytic features examined by extant research include ideas/content, organization, aspects of language use (e.g., voice, vocabulary, syntax), and conventions (Bacha, 2001; Coe et al., 2011). However, when it comes to writing at the secondary level and across content areas, it is critical to examine genre-specific and discipline-specific writing features as well (Goldman et al., 2016; Shanahan & Shanahan, 2012).

110

Using evidence to support arguments is crucial in source-based argumentative writing (Monte-Sano, 2010). However, using evidence in an appropriate way presents challenges to writers because writing in history is knowledge-transformation: writers need to actively and critically interpret and analyze the sources (De La Paz, 2012; Goldman & Scardamalia, 2013; Monte-Sano, 2010). In this process, writers need to consider potential biases from the authors and sources and to evaluate the reliability of the sources. Historical writing also demands that writers interpret and critically evaluate their sources (Goldman & Scardamalia, 2013). This involves specific historical thinking skills such as sourcing, contextualization, and corroboration (Goldman et al., 2016; Wineburg, 1991). Sourcing entails assessing potential biases from the authors, contextualization involves placing evidence and arguments within historical context and environment, and corroboration means comparing multiple sources to verify and strengthen arguments. Thus, we incorporated a historical thinking dimension in addition to evidence use to capture these essential processes unique to argumentative writing in history.

Steiss and colleagues (2024) found that writing quality of source-based argumentative writing in history of secondary students was best depicted by a bifactor model with a general factor and four specific factors (structure, evidence use, historical thinking, and language use) for students in secondary grade levels. The current study uses the same analytic coding framework as Steiss and colleagues (2024). In the current study, we aim to comprehensively examine four aspects of writing quality using analytic coding: structure, evidence use, historical thinking, and language use in an argumentative writing task in history. The analytic evaluation approach allows researchers to investigate various dimensions of writing quality specific to the genre (Bacha, 2001; Coe et al., 2011; Culham, 2003).

The majority of previous studies investigating the relation between writing self-efficacy and writing quality did not examine multiple dimensions of writing self-efficacy or writing quality. One exception is Zabihi (2017); by using data from EFL college students, they found that writing self-efficacy predicted complexity, accuracy, and fluency of a narrative writing task. The current study aims to explore the relations between dimensions of writing self-efficacy and dimensions of writing quality (including genre- and discipline-specific dimensions) in argumentative writing in history to understand which dimensions of writing quality can be predicted by writing self-efficacy.

Writing Self-efficacy in LM and EO Students

Due to the complexity of writing, ELs, by definition, have not reached proficiency in English language skills and struggle with writing tasks (U.S. Department of Education, n.d.). Language skills play a critical role when translating ideas into words (Graham & Kim, 2022) and making revisions (Chen & Zhang, 2019; Graham & Kim, 2022). Thus, the writing process may pose a greater demand on self-efficacy for some LM writers (Chen & Zhang, 2019; Garcia-Sanchez & Fidalgo-Redondo, 2006). A meta-analysis found that self-efficacy and students' language proficiency are positively correlated (Wang & Sun, 2020), indicating the potential need to pay attention to writing self-efficacy of students with limited language proficiency in schools.

Despite the importance of writing self-efficacy, to our knowledge, no study has investigated writing self-efficacy across different English language designations in the US context. Several studies conducted in Australia and Malaysia have examined bilingual students' writing self-efficacy in an English as a Second Language (ESL) context (e.g., Rayner et al., 2016; Shah et al., 2011). Rayner and colleagues (2016) found that bilingual Australian and Malaysian undergraduate students had higher levels of writing self-efficacy than monolinguals in

112

writing summaries and literature reviews and asking and refining leading questions. Shah and colleagues (2011) found that bilinguals with limited language proficiency in secondary school had a moderate level of writing self-efficacy. In an EFL context, results showed that bilingual undergraduate students also tended to have a moderate level of writing self-efficacy (Sun & Wang, 2020). In addition, the study found that these students were more efficacious in developing ideas and constructing paragraphs but had relatively lower self-efficacy for specific writing tasks. In summary, previous studies examining self-efficacy of bilingual students of various language proficiency levels have yielded mixed findings regarding their writing self-efficacy; there are potential differences as a function of language proficiency and the specific aspect of writing self-efficacy examined.

In the current study, we examine writing self-efficacy of LM and EO students, including students designated as RFEP and EL. Understanding students' writing self-efficacy of LM students across different English language designations, a common way to label learners by their language proficiency in the US, is absent. LM students entering school with limited language proficiency are classified as either proficient or not proficient in English language skills, and only those classified as ELs will continue to receive additional support in English language development. However, recent studies (Hwang et al., 2017) pointed out that students classified as Reclassified Fluent English Proficient (RFEP) may constitute a heterogeneous group, where some may have unique needs that could benefit from sustained support. Previous studies found that the academic performance of students designated as RFEP may differ from EO/IFEP peers (Kim & Herman, 2009), and this is likely due to their previous experience as ELs and varying redesignation standards and students designated as RFEP may still need further support in literacy skills after they are reclassified (Hwang et al., 2017). However, few studies have

explicitly examined literacy performance of students designated as RFEP (with a few exceptions: Hwang et al., 2017; Steiss et al., 2024). In addition, it remains unclear whether this group of students may over- or underperform in writing self-efficacy, given the different classification processes across states and the lack of research isolating this group of learners.

Present Study

Although writing self-efficacy is associated with writing performance (Camacho et al., 2020), writing self-efficacy research rarely focuses on secondary students in the US context. It remains unclear which dimension(s) of writing self-efficacy students need support with. In the current study, we aim to examine this by investigating which dimension(s) of writing students feel less confident in and which dimension(s) are more predictive of writing performance. In addition, the current study aims to address a critical gap in investigating writing self-efficacy in LM students with various language designations. By examining EL and RFEP as separate groups from other students, we aim to understand the specific needs of various LM learners in writing self-efficacy.

The following research questions guided the present study:

To what extent do dimensions of writing self-efficacy (ideation, self-regulation, and revision) vary among secondary students, controlling for English language designation, gender, SES, race and ethnicity, and grade levels?

To what extent do dimensions of writing self-efficacy relate to dimensions of analytic coding of writing quality (structure, evidence use, historical thinking, and language use), controlling for English language designation, gender, SES, race and ethnicity, and grade levels?

To what extent do students of different English language designations (EL, RFEP, IFEP/ EO)

perform similarly or differently on writing self-efficacy, controlling for gender, SES, race

and ethnicity, and grade levels?

As for the relation between writing self-efficacy and dimensions of writing quality, we

hypothesized that writing self-efficacy would predict multiple aspects of analytic coding of

writing quality significantly and independently over and above student demographics (gender,

English language designation, SES, race; Bruning et al., 2013; Graham et al., 2019; Pajares,

2003; Sun et al., 2022). Depending on various reclassification criteria of different states, students

designated as RFEP may perform lower or higher than their EO/IFEP peers. We did not have a

specific hypothesis regarding how writing self-efficacy would vary for students designated as

RFEP, EL, and EO/IFEP because no previous study has investigated this question to our

knowledge.

Method

Participants

Participants in the study were from a large intervention study investigating instructional

practices that improve secondary students' performance in source-based argumentative writing in

history (Author et al., in preparation). Only pretest data were used in the present study. The

sample included 360 students from Grades 6 to 11. The sample was selected using a stratified

random sampling procedure. The larger sample was blocked by biological sex and English

language status to ensure that the sample was balanced in biological sex and English language

designations. However, 36 students did not consent to take the writing self-efficacy survey, and

11 students responded to only 0-2 items out of 20 items in the writing self-efficacy survey. In

addition, nine students gave the same response for all items in the survey. These students were deleted from the dataset. Thus, the final dataset included 304 participants.

Fifty-seven percent of students were females (with 11 students having missing data on biological sex). Students were from Grade 6 to 11, with 54 students in Grade 6, 84 in Grade 7, 37 in Grade 8, 48 in Grade 10, and 81 in Grade 11. Students were from 39 classrooms in three school districts' public schools in California. According to laws in California, reclassification criteria include the following aspects: (1) assessment of English language proficiency, (2) teacher evaluations, (3) parent consultation, and (4) basic skills relative to English proficient students (California Department of Education, 2024). The criteria require evaluating students' language skills from multiple sources across multiple contexts and thus may represent students' language skills more accurately. Based on the district record, 87 were designated as EL, 106 were designated as RFEP, and 111 were designated as IFEP/EO. Of the students whose parent education information was available (n = 270), 37 reported having a graduate degree or higher, 49 had a college degree, 43 had some college education or an associate degree, 86 were high school graduates, and 55 did not graduate from high school.

According to a survey of history teachers participating in the study, teachers rarely assigned extended writing tasks to students and assigned writing tasks to students several times per semester. The majority of the writing tasks involved filling out worksheets. Most teachers tended to teach writing strategies a few times per semester. Many teachers focused on several strategies: reading strategies, vocabulary related to specific topics, how to write an introduction and conclusion paragraph, how to revise essays, how to write a claim, and how to establish the point of view in writing. Most teachers indicated they are knowledgeable of some teaching strategies but not others: text organization, tap prior knowledge, how to write introduction and

conclusion sections, and genre-related vocabulary. The survey also indicated that they are less knowledgeable about teaching students to set goals in writing, self-assessing in writing, and providing feedback for peers.

Measures

Writing Self-Efficacy

Writing self-efficacy was assessed by a survey modified from the Self-efficacy for Writing Scale (SEWS; Bruning et al., 2013; see Appendix D). The survey included the same self-efficacy items (five items in ideation and five in self-regulation; one self-regulation item was removed because of high correlation and similar content with another item) from SEWS. The language and conventions aspect was tested to be a distinct dimension of writing self-efficacy from ideation and self-regulation (Bruning et al., 2013; Sun et al., 2022). We did not administer items in this aspect. Because source-based argument tasks were used, where students may borrow language from the sources, compared to other aspects of writing self-efficacy, we decided language and conventions was a less relevant factor for this genre.

Another aspect of writing self-efficacy that was included is revision, which has been previously tested to be a separate aspect of writing self-efficacy (Wang et al., in preparation). Construct validity and reliability of revision items in writing self-efficacy have been established in previous studies as well (Chen & Zhang, 2019; Wang et al., in preparation). Four items related to writing revision were created in consultation with writing research experts. The four items evaluated students' perceived capability in identifying areas of improvement and strengths and revising the introduction, body, and conclusion to improve the essay. All items used a five-point scale, ranging from "strongly disagree" as one to "strongly agree" as five. The items all had good

reliability. The Cronbach's alpha for ideation, self-regulation, and revision subscales and the writing self-efficacy scale were .78, .72, .78, and .88, respectively.

Written Composition

Two history source-based argumentative writing prompts, one about the Montgomery bus boycott and the other about the Delano grape boycott, were randomly administered among students in a two-day activity. On the first day, students read the sources and the prompt and filled out graphic organizers provided right next to the sources for 50 minutes. On the second day, they wrote an essay about why the boycott succeeded for an additional 50 minutes. The prompts included instructions on basic computer skills, procedures for the activity, background information about the boycott, four sources related to the boycott, graphic organizers that helped students make notes about what they had read, requirements for the essay, and space for students to write. The four sources included an overall timeline of the boycott and three sources that introduced events that may have led to the success of the boycott. The four sources for the bus boycott prompt had 198, 369, 269, and 345 words, respectively. The Flesch-Kincaid grade-level readability scores for four sources were 46.3, 62.4, 66.4, and 46.7. The four sources for the grape boycott prompt had 237, 278, 224, and 343 words, respectively. The Flesch-Kincaid grade-level readability scores of four sources were 53.5, 77.6, 75.4, and 72.2. Students had access to a read-aloud of the sources but received no additional instruction. They were asked to write an argument essay with an introduction, body paragraphs with evidence and reasoning (including a counterargument paragraph), and a summarizing their ideas and what can be learned from this historical event.

The study used an analytic coding framework developed based on previous research, existing writing rubrics, and expert opinions. This coding framework has been tested and

validated in recent research (Steiss et al., 2022; Steiss et al., 2024). An extra aspect, historical thinking, was added to the analytic coding framework because the writing task examined in the current study was source-based argumentative writing in history class. Extant research has indicated the importance of disciplinary literacy (Goldman et al., 2016; Shanahan & Shanahan, 2012). Writing an argument in history requires general argumentative skills (e.g., writing a claim, using evidence to support claims) and disciplinary-specific skills in history (e.g., contextualizing the arguments; Wineburg, 1991). The framework included 14 items (Steiss et al., 2024). Appendix C shows all items that belong to four categories: structure (three items), evidence use (four items), historical thinking (four items), and language use. The items were scored on a scale from one to seven, with one representing "not evident" and seven representing "highly effective". Three PhD students with expertise in language and literacy as well as three junior and senior undergraduate students with research experience in education scored the essays. All scorers underwent rigorous training where they did multiple rounds of coding, discussed discrepancies in coding, modified and provided more details on the coding framework, and selected multiple anchor papers for each score. Final inter-rater reliabilities within one point were reported in Appendix C using 20% of the total sample. Inter-rater reliability (within one point) ranged from .87 to 1.00.

Procedures

Detailed directions to administer the writing prompts and the survey were delivered to teachers through a professional development session; digital and printed copies of the directions were also distributed to all teachers. Students wrote digitally using Google Documents. The writing self-efficacy survey was given to students through Qualtrics. Item orders were randomized within surveys to prevent fatigue effects.

Data Analysis Strategy

There was no missing data in students' written composition evaluated using analytic coding. Among the 14 writing self-efficacy survey items, 11 of them had several missing responses (see Table 3.1). Little's MCAR test was conducted using all variables involved in the study. The result was not significant, indicating that completely missing at random cannot be rejected ($\chi 2$ = 339.18; df = 345, p = .58).

Table 3.1

Descriptive Statistics

	EL		RFEP		EO/IFEP		
Item	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	Missing data
Writing Self-Efficacy							
Ideation1	3.49	.91	3.50	1.03	3.62	.94	4
Ideation2	3.48	.97	3.27	1.02	3.53	.89	2
Ideation3	3.91	.92	3.78	.85	3.93	.91	2
Ideation4	3.34	.87	3.25	.87	3.33	.88	1
Ideation5	3.36	.81	3.26	.81	3.67	.88	1
Self-regulation1	3.08	1.00	3.24	1.11	3.36	1.00	4
Self-regulation2	3.11	1.02	3.16	1.04	3.41	1.10	0
Self-regulation3	3.10	1.05	2.99	1.10	2.83	1.04	0
Self-regulation4	3.05	.99	3.18	1.04	3.35	1.03	0
Self-regulation5	3.28	.98	3.42	.89	3.68	.70	1
Revision1	3.52	.89	3.62	.97	3.84	.85	1
Revision2	3.58	.94	3.71	.94	3.92	.84	2
Revision3	3.44	.90	3.74	.90	3.98	.77	1
Revision4	3.59	.87	3.70	.91	3.85	.96	4
Writing Quality							
Claim	2.83	1.69	3.91	1.61	4.40	1.58	0

Introduction	2.60	1.04	3.44	1.12	3.77	1.51	0
Conclusion	1.94	1.85	3.16	2.27	4.06	2.37	0
Body	2.55	1.25	3.57	1.43	4.03	1.72	0
Evidence	1.95	1.16	2.85	1.57	3.59	1.74	0
Commentary	1.45	.74	2.42	1.43	3.15	1.71	0
Attribution	2.37	1.32	3.27	1.69	3.99	1.89	0
Sourcing	1.47	.75	1.89	1.00	2.24	.98	0
Contextualization	1.85	.97	2.89	1.30	3.41	1.54	0
Corroboration	1.25	.53	1.80	1.13	2.37	1.24	0
Counterargument	1.37	.73	2.17	1.49	2.61	1.59	0
Sentence fluency	2.51	1.19	3.54	1.27	4.01	1.56	0
Diction	2.67	1.16	3.73	1.44	4.24	1.95	0
Conventions	2.39	1.07	3.33	1.32	3.81	1.54	0

Note. SE: writing self-efficacy scale. EL: English Learners; RFEP: Reclassified Fluent English Proficiency; IFEP: Initially Fluent English Proficiency.

The first research question was answered by running within-subject repeated measure ANOVA, with EL status, gender, SES, race and ethnicity, and grade levels as covariates. Repeated measures were aspects of writing self-efficacy (ideation, self-regulation, and revision). The analysis was adjusted using the Bonferroni approach to account for increased Type-I error caused by multiple comparisons. Because different numbers of items were included for each aspect, the average score of each aspect was used.

The second research question was addressed using structural equation modeling (SEM) with dimensions of writing self-efficacy predicting dimensions of writing quality. Before fitting SEM models, confirmatory factor analyses were conducted to confirm the factor structures of writing self-efficacy and writing quality. Biological sex, grade levels, race and ethnicity, EL status, and SES as operationalized as parent education were included as control variables in the SEM model. Mplus version 8.3 (Muthén & Muthén, 2017) was used for data analysis.

To answer the third research question, we ran a multivariate analysis of covariance (MANCOVA) models on the three dimensions of writing self-efficacy measures by three groups (EL, RFEP, and EO/ IFEP), controlling for biological sex, SES, race, and ethnicity, and grade levels. Post-hoc analysis was conducted. Bonferroni was used to correct the Type-I error rate for multiple comparisons for the first and third research questions. SPSS 29.0 was used for the data analysis of these two research questions.

Results

Descriptive Statistics and Preliminary Analysis

Descriptive statistics (mean, standard deviation, and number of missing data) of writing self-efficacy and analytic coding scores of writing quality can be seen in Table 3.1. Students designated as EL and RFEP scored lower on all writing self-efficacy items than those designated as IFEP/EO. Els scored higher on ideation than those designated as RFEP for four out of five items, whereas students designated as Els scored lower on self-regulation (four out of five items) and revision (all four items) than students designated as RFEP. Kurtosis and skewness of all items were checked and were within normal range. In addition, across all groups of students, students overall had lower writing self-efficacy in self-regulation than ideation and revision.

The correlation table of writing self-efficacy survey items and analytic coding scores is shown in Table 3.2. All analytic coding items were moderately or strongly correlated with each other. Most correlations between writing self-efficacy items were moderate, and the rest were small.

Research Question 1: Within-Subject Differences in Dimensions of Writing Self-efficacy

Within-subject repeated ANOVA revealed significant differences in students' scores across dimensions of writing self-efficacy after controlling for gender, EL status, gender, SES, race and ethnicity, and grade levels ($F(2,248) = 5.51$, $p =.005$, partial $\eta2 = .02$). The Mauchly's Test of Sphericity was significant (Mauchly's $W = .94$, $p < .001$), so the Greenhouse-Geiser adjusted results were reported.

The results of pairwise comparisons can be seen in Table 3.3. All dimensions of writing self-efficacy differed significantly (ps < .001). On average, students' scores on the revision dimension were .48 higher than the self-regulation dimension and .18 higher than the ideation dimension. Students scored .30 higher on the ideation dimension than on the self-regulation dimension. The difference between ideation and self-regulation and the difference between self-regulation and revision were of medium effect sizes ($d = .51$; $d = .74$ respectively), whereas the difference between ideation and revision was of small effect size ($d = .42$).

Research Question 2: Relation of Writing Self-efficacy to Dimensions of Writing Quality

Regarding the dimensionality of the writing self-efficacy measure, writing self-efficacy was best represented as a three-factor structure, including ideation, self-regulation, and revision (see Figure 3.1 for the three-factor model; model fit information is presented in Table 3.4): $\chi^2 =$ 109.35 ($df = 74$, $p = .005$); RMSEA = .04, 90%CI[.02, .06], CFI = .97, TLI = .96, SRMR = .04. Regarding the dimensions of writing quality, we found that a three-factor model (structure, evidence use/historical thinking, and language use and conventions) represented the data the best for this dataset: $\chi^2 = 402.86$ ($df = 74$, $p < .001$), RMSEA = .12, 90%CI[.11, .13], CFI = .92, TLI = .90, SRMR = .04 (see Appendix E and F for confirmatory factor analysis models and model fit information, respectively).

Table 3.2

Correlation Table

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Presenting claim	--												
2 Introduction	.66**	--											
3 Conclusion	.60**	.44**	--										
4 Body	.66**	.52**	.63**	--									
5 Evidence	.66**	.53**	.62**	.82**	--								
6 Commentary	.62**	.53**	.60**	.73**	.80**	--							
7 Attribution	.53**	.45**	.50**	.75**	.77**	.66**	--						
8 Sourcing	.52**	.40**	.44**	.67**	.73**	.65**	.82**	--					
9 Contextualization	.70**	.57**	.60**	.75**	.74**	.78**	.64**	.61**	--				
10 Corroboration	.51**	.40**	.50**	.63**	.69**	.59**	.66**	.65**	.62**	--			
11 Counterargument	.56**	.49**	.61**	.62**	.69**	.68**	.55**	.52**	.66**	.61**	--		
12 Sentence fluency	.61**	.64**	.68**	.75**	.72**	.69**	.59**	.57**	.72**	.57**	.60**	--	
13 Diction	.57**	.63**	.63**	.71**	.66**	.71**	.59**	.57**	.69**	.53**	.58**	.85**	--
14 Conventions	.56**	.59**	.61**	.70**	.66**	.65**	.58**	.57**	.67**	.52**	.54**	.89**	.88**
15 SE-ideation1	.01	-.03	-.05	-.03	-.01	-.01	-.01	.02	.01	-.06	-.04	.00	-.03
16 SE-ideation2	0.1	-.03	.05	.07	.10	.06	.06	.10	.08	.04	.05	.04	-.01
17 SE-ideation3	-.02	-.05	.01	.04	.02	-.03	.04	.07	.00	-.03	-.04	-.02	.00
18 SE-ideation4	.03	.01	.04	.04	.05	.02	.06	.06	.12*	.02	.08	.04	.06
19 SE-ideation5	.06	-.06	.04	.03	.07	.00	.03	.04	.06	.05	.05	.00	.01

LANGUAGE, SELF-EFFICACY, & WRITING

20 SE-regulation1	.15*	.05	.09	.20**	.20**	.15**	.17**	.11	.20**	.16**	.10	.10	.13*
21 SE-regulation2	.17**	.13*	.15**	.22**	.22**	.23**	.21**	.23**	.26**	.13*	.22**	.24**	.25**
22 SE-regulation3	.11	-.01	-.01	.11	.10	.03	.11*	.09	.12*	.08	.05	.07	.00
23 SE-regulation4	.10	.06	.10	.10	.12*	.10	.06	.04	.14*	.05	.12*	.08	.02
24 SE-regulation5	.17**	.13*	.11	.19**	.17**	.17**	.20**	.24**	.22**	.16**	.11	.16**	.19**
25 SE-revision1	.14*	.11	.08	.12*	.13*	.11	.10	.16**	.18**	.11	.07	.09	.09
26 SE-revision2	.17**	.15*	.12	.19**	.23**	.15**	.16**	.20**	.22**	.14	.12*	.14*	.12*
27 SE-revision3	.21**	.22**	.15**	.26**	.26**	.22**	.19**	.19**	.25**	.19	.15**	.22**	.15**
28 SE-revision4	.01	-.03	.08	.09	.10	.10	.11	.17**	.18**	.13	.06	.09	.05

	16	17	18	19	20	21	22	23	24	25	26	27	28
16 SE-ideation2	--												
17 SE-ideation3	-.02	--											
18 SE-ideation4	.04	.45**	--										
19 SE-ideation5	.01	.41**	.47**	--									
20 SE-regulation1	.09	.45**	.33**	.41**	--								
21 SE-regulation2	.02	.42**	.29**	.38**	.26**	--							
22 SE-regulation3	.11*	.25**	.28**	.27**	.23**	.19**	--						
23 SE-regulation4	.25**	.32**	.39**	.36**	.36**	.16**	.18**	--					
24 SE-regulation5	.06	.30**	.35**	.36**	.27**	.15**	.28**	.34**	--				
25 SE-revision1	.05	.41**	.47**	.40**	.36**	.31**	.34**	.40**	.47**	--			
26 SE-revision2	.15**	.33**	.32**	.39**	.28**	.25**	.26**	.30**	.28**	.42**	--		

27 SE-revision3	.11	.33**	.36**	.35**	.26**	.22**	.21**	.26**	.29**	.36**	.41**	--
28 SE-revision4	.14*	.27**	.28**	.32**	.23**	.19**	.24**	.24**	.26**	.32**	.31**	.52**	--

Note. SE: writing self-efficacy scale; 1-14 are writing quality items. *$p < .05$; **$p < .01$.

Table 3.3

Summary of Pairwise Comparisons in the Within-Subject Repeated Measure ANOVA on Dimensions of Writing Self-efficacy

Pairwise comparison	Mean difference	*SE*	Cohen's *d*	*p*
ideation - self-regulation	.30	.03	.51	< .001
ideation - revision	-.18	.04	.42	< .001
self-regulation -revision	-.48	.04	.74	< .001

SEM model of writing self-efficacy and various dimensions of writing quality can be seen in Figure 3.2. Information about covariates can be seen in Table 3.5. Model fit indices were acceptable: χ^2 = 1178.27 (*df* = 660; *p* < .001), RMSEA =.06, 90%CI [.05, .06], CFI = .89, TLI = .88, SRMR = .06. Based on this model, controlling for students' demographic information (including biological sex, grade, race, EL status, and parent education), the self-regulation dimension of writing self-efficacy was positively associated with two dimensions of writing quality: structure (*β* = .56, *p* = .05) and evidence use/historical thinking (*β* = .66, *p* = .02). Other associations were not statistically significant.

Research Question 3: Self-efficacy of Students from Different English Language Designations

MANCOVA results are shown in Table 3.6. The main effect was statistically significant (Pillai's trace = .076, *F*(2, 232) = 3.04, *p* = .006). The post-hoc analysis found that students designated as RFEP scored significantly lower on the ideation and revision dimension than those designated as IFEP/EO, with a mean difference of 1.74 (*p* = .003). For the revision dimension, students designated as EL scored 1.34 lower than those designated as IFEP/EO (*p* = .03). Students designated as RFEP also scored 1.14 lower on the revision dimension than those designated as IFEP/EO (*p* =.04). No differences were found for the self-regulation dimensions.

Discussion

In the present study, we investigated which dimension(s) of writing self-efficacy students scored lower, which dimension(s) of writing self-efficacy is crucial for writing performance, and the differences in writing self-efficacy among LM students with different levels of language proficiency. Results indicated that students had the lowest self-efficacy in the self-regulation
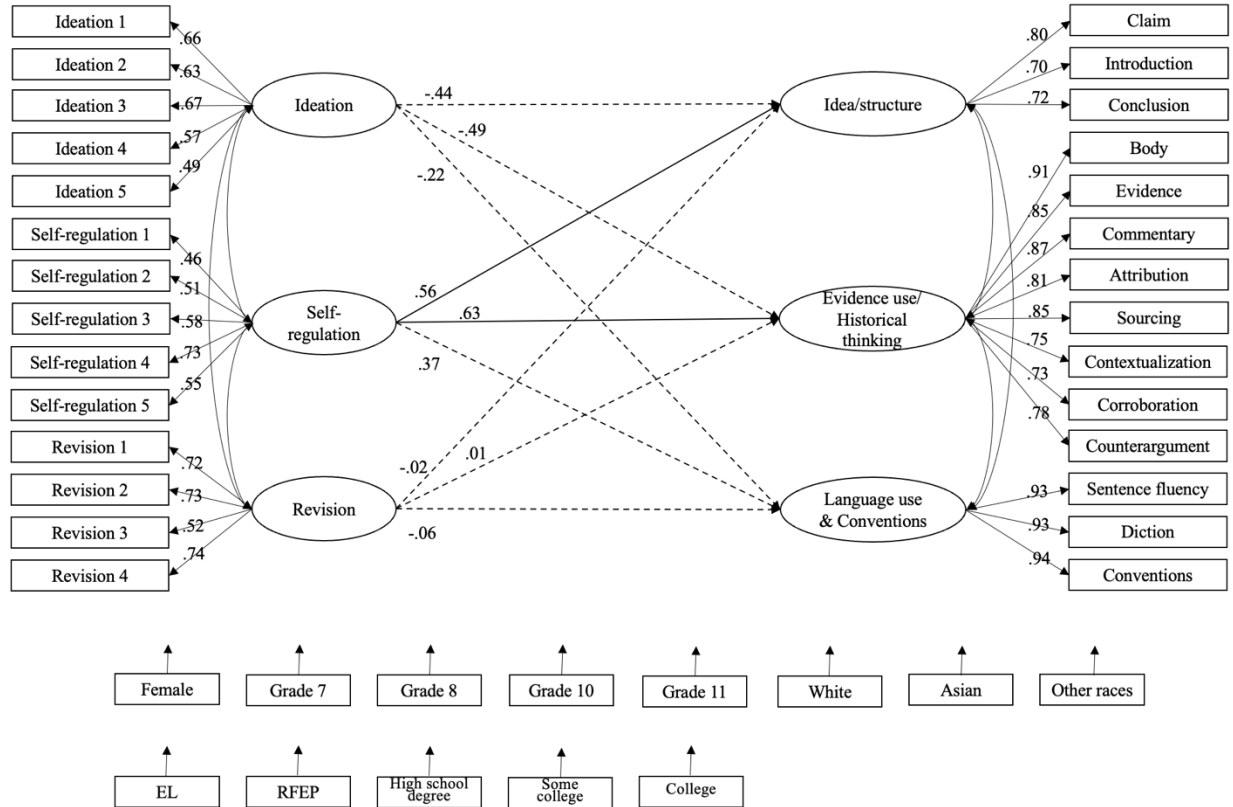
Figure 3.1

Three-factor Model for Writing Self-efficacy (Standardized Results)
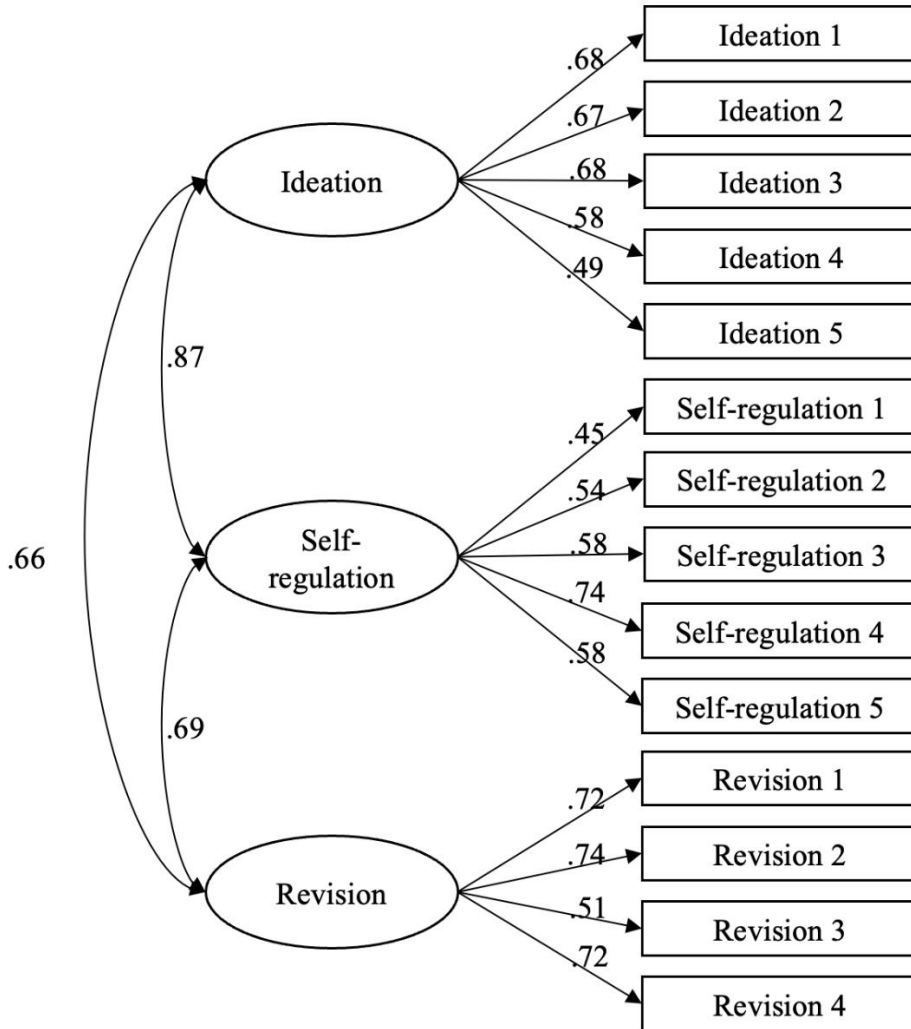
Figure 3.2

SEM Model (Standardized Coefficients)

Note. EL: English Learners; RFEP: Reclassified Fluent English Proficiency. Solid lines represent significant coefficients; dashed lines represent insignificant coefficients. The reference group is Grade 6 for grade levels, Hispanic for races and ethnicities, EO/IFEP for EL status, and no high school degree as the reference group for parent education.

Table 3.4: Model Fit Information of Confirmatory Factor Analysis of Writing Self-efficacy

	Model	χ^2	p	df	CFI	TLI	RMSEA	RMSEA 90%CI	SRMR	Model Comparison			
										model	χ^2	df	p
a	One-factor	223.64	<.001	77	.88	.85	.08	[.07, .09]	.06				
b1	Two-factor: ideation & self-regulation as one dimension	124.73	<.001	76	.96	.95	.05	[.03, .06]	.04	a	98.91	1	<.001
b2	Two-factor: ideation & revision as one dimension	191.29	<.001	76	.90	.88	.07	[.06, .08]	.05				
b3	Two -factor: self-regulation & revision as one dimension	198.44	<.001	76	.91	.89	.07	[.06, .08]	.05				
c	Three-factor	109.35	.005	74	.97	.96	.04	[.04, .06]	.04	b	15.38	1	<.001
d	Second order three-factor model	109.35	.005	74	.97	.96	.04	[.04, .06]	.04				

Note. Both restricted and covariate bifactor models did not converge.

Table 3.5

Standardized Coefficients, Standard Errors, and P-values for Covariates in the SEM Models

	Structure		Evidence use/Historical thinking		Language use	
	β	SE	β	SE	β	SE
Female	.06	.06	.05	.05	.06	.05
Grade 7	.04	.08	.16*	.07	.15	.07
Grade 8	.14*	.07	.20**	.06	.18**	.06
Grade 10	.20**	.08	.24***	.06	.29***	.06
Grade 11	.31***	.08	.39***	.07	.44***	.07
White	-.04	.07	.02***	.05	.08	.06
Asian	.12	.08	.20**	.06	.17**	.06
Other races	-.13*	.06	-.07	.05	-.06	.05
EL	-.36***	.07	-.29***	.06	-.25***	.06
RFEP	-.14	.07	-.13*	.06	-.05	.06
High school degree	.03	.08	.08	.07	.05	.07
Some college	.13	.07	.09	.06	.04	.06
College or higher	.19	.10	.20*	.08	.19*	.08

Note. $*p < .05$; $**p < .01$; $***p < .001$. EL: English Learners; RFEP: Reclassified Fluent English Proficiency. Solid lines represent significant coefficients; dashed lines represent insignificant coefficients. The reference group is Grade 6 for grade levels, Hispanic for races and ethnicities, EO/IFEP for EL status, and no high school degree as the reference group for parent education.

Table 3.6

Comparisons of Writing Self-efficacy Among Students With Different English Language Designations

Testing items	Group	Adjusted mean	SE	*F(2, 232)*	*η2*	post-hoc
Writing self-efficacy -- Ideation						
	EL	17.74	.42	5.75	.05	EO/IFEP>RFEP**
	EO/IFEP	18.59	.36			
	RFEP	16.85	.34			
Writing self-efficacy -- Self-regulation						
	EL	15.85	.46	1.17	.01	
	EO/IFEP	16.69	.39			
	RFEP	15.91	.36			
Writing self-efficacy -- Revision						
	EL	14.31	.37	4.29	.04	EO/IFEP>EL*, RFEP*
	EO/IFEP	15.65	.31			
	RFEP	14.51	.29			

Note. *p < .05; **p < .01; ***p < .001. Post-hoc analysis was adjusted using Bonferroni

dimension, which was the dimension that is most predictive of writing quality. Students who did not enter school with proficient language skills, including students designated as EL and RFEP, had lower writing self-efficacy in certain dimensions of writing self-efficacy and equivalent writing self-efficacy in self-regulation compared to students who were initially fluent in English when they entered school.

Theoretically, writing self-efficacy is crucial for writers to navigate through the complete writing process, as it influences the extent to which students invest in the writing process (Graham, 2018; Kim & Graham, 2022). Our study adds to a growing body of evidence showing the importance of writing self-efficacy in writing development (Brunning et al., 2013; Ekholm et al., 2015; Garcia-Sanchez & Fidalgo-Redondo, 2006; Pajares, 2003; Pajares et al., 2007). Aligned with previous research (Wang et al., in preparation), writing self-efficacy is a multi-dimensional construct consisting of ideation, self-regulation, and revision dimensions. We found that students scored their writing self-efficacy in self-regulation the lowest, but the self-regulation dimension was uniquely predictive of students' writing quality. These findings indicate the need to focus on students' self-regulation skills to improve their writing performance. Writing is a complex task where writers need to constantly coordinate and self-regulate various resources available to them: cognitive and linguistic resources, tools they can employ during the writing process, collaborators, and mentors who are part of the writing communities (Graham, 2018). Self-regulation is critical to the writing process because writing is not linear and demands students to direct their attention and effort to deliberately consider the overall structure, audience, or rhetorical goals of their writing (Scardamalia & Bereiter, 1987).

Despite the importance of self-regulation in the writing process, students perceived their ability to self-regulate as the lowest among all dimensions, which indicates that writing

134

instruction should more purposefully improve self-regulation skills. One such instruction approach is the Self-Regulated Strategy Development (SRSD) approach. SRSD writing instruction focuses on explicitly teaching students to use various writing strategies to plan, draft, and revise their writing (Graham & Harris, 2018) and has been found to be effective in improving writing quality of students from various backgrounds and developmental phases (e.g., Gillespie & Graham, 2014). Given that difficulties with self-regulation are seen in all aspects of problem-solving of unskilled writers (Ferretti & Lewis, 2013), if students can learn to coordinate their writing process in a more effective way where they adjust their writing based on the critical rhetorical, communicative, and pragmatic goals of their writing, they are more likely to be successful writers (Harris & Graham, 2013).

In addition, the current study also found that only certain dimensions of writing quality were predicted by writing self-efficacy: structure and evidence use/historical thinking dimensions (not language use and dimension). This finding aligned with the dynamic relations hypothesis of DIEW (Kim & Graham, 2022), which states differential relations of component skills to writing as a function of measurement (i.e., dimensions of writing). The current study highlighted that different dimensions of writing self-efficacy did not relate to different dimensions of writing quality in the same way. We found that students' perception of their writing self-efficacy is more strongly related to the genre- and discipline-specific features and structure in written compositions than language use. This could be due to higher demands and challenges posed to students in structure and distinct genre and discipline writing features when writing in this specific genre/discipline. This finding underlines the importance of writing instruction in improving students' ability to structure their essays effectively and to align their writing with genre- and discipline-specific requirements.

Differences in writing self-efficacy were found across linguistically diverse students with different English language designations. Students classified as RFEP have lower writing self-efficacy in ideation than their EO/IFEP peers. In addition, we found that controlling for demographic factors, students designated as EL and RFEP had lower writing self-efficacy in revision. This could be due to these students' lower perceived language proficiency or their lower perceived ability to revise their essays, given their experiences of receiving additional support in language development. However, it is surprising that students classified as RFEP had lower writing self-efficacy in ideation and revision, given the relatively strict reclassification criteria in California, where the study was conducted. Based on criteria to reclassify students designated as EL to RFEP (California Department of Education, 2024), multiple sources of information should be considered when evaluating students' language skills, which should generate a relatively accurate understanding of students' language skills. The fact that students designated as RFEP rate their writing self-efficacy lower may indicate multiple possibilities. First, it is likely that although students designated as RFEP are equally capable with their peers, but they may feel less confident in their writing skills due to their previous experience designated as EL. Students designated as EL may face limited access to mainstream curriculum, social activities, and so on, which may lead to a lower level of self-efficacy in this group of students. Second, it is also possible that students designated as RFEP may need sustained support in certain aspects of their language and literacy development (Hwang et al., 2017). Overall, the results suggest the importance of catering to linguistically diverse students' needs in writing self-efficacy in ideation and revision. However, it is also crucial to note that students designated as EL have equivalent writing self-efficacy in ideation and self-regulation as their EO/IFEP peers. These findings indicate that they perceive equally self-efficacious in developing ideas and

sustaining/regulating their writing process. This finding reminds educators to draw on the strengths, skills and resources LM students already possess when designing instructional practices to improve their literacy development (Rodriguez, 2013). In addition, Camping and colleagues (2020, 2023) found that bilinguals had higher writing motivation than their monolingual peers on the majority of motivational constructs they assessed. Camping and colleagues (2003) found that bilingual elementary students had higher self-regulatory writing motives than their monolingual peers. The authors point out that it is possible that bilingual students may interpret items on motivation scales differently from monolingual students. For example, Camping and colleagues (2023) argued that bilingual students may conceptualize writing in a different way: while bilinguals might have a broader definition of writing, which includes texting and other social written communication, monolingual students may not. Similarly, it is likely that bilingual students may have interpreted writing self-efficacy items differently than monolingual students. Future research is warranted to examine whether different populations interpret writing self-efficacy items in similar ways to guarantee the assessment is equivalent across groups.

Limitations and Future Directions

One limitation is that we did not measure grammar and conventions as one dimension of writing self-efficacy, which was indicated as one dimension of writing self-efficacy (Bruning et al., 2013; Sun et al., 2022; Sun & Wang, 2020). Future studies should examine this dimension and the revision dimension, which was tested as a separate dimension of writing self-efficacy in the current study. In addition, due to the small number of IFEP students in the current study, we were unable to examine this group as an independent group of students from EO students. Bilingual students with linguistic resources in multiple languages may have different language

and literacy performances from monolinguals. Future studies should examine linguistically diverse students' writing self-efficacy and investigate writing self-efficacy of bilingual students who already had proficient English language skills when entering school. Additionally, because the current study is a correlational study, implications are only suggestive. Future studies should replicate the analysis using a longitudinal design.

The study, to our knowledge, is the first to examine writing self-efficacy of LM students by their English language designations and explicitly and comprehensively explore which dimension of writing self-efficacy students need more support with, and the relation between writing self-efficacy and dimensions of writing quality in source-based argumentative writing in history with consideration of genre- and discipline-specific features in writing. Future studies should further explore these topics to better understand the needs of heterogeneity in linguistically diverse students and specific demands in various writing tasks posed to students' writing and writing self-efficacy. Understanding sources of lower self-efficacy in certain aspects for students designated as EL and RFEP can help identify instructional approaches to support their needs more effectively.

References

Bacha, N. (2001). Writing evaluation: What can analytic versus holistic essay scoring tell us?. *System*, *29*(3), 371-383.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The Exercise of Control*.

Bereiter, C., & Scardamalia, M. (1987). An attainable version of high literacy: Approaches to teaching higher-order skills in reading and writing. *Curriculum Inquiry*, *17*(1), 9-30.

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology*, *105*(1), 25.

California Department of Education. (2024). Reclassification. https://www.cde.ca.gov/sp/el/rd/

Camacho, A., Alves, R. A., & Boscolo, P. (2020). Writing motivation in school: A systematic review of empirical research in the early twenty-first century. *Educational Psychology Review*, *33*(1). https://doi.org/10.1007/s10648-020-09530-4

Camping, A., Graham, S., Ng, C., Aitken, A., Wilson, J. M., & Wdowin, J. (2020). Writing motivational incentives of middle school emergent bilingual students. *Reading and Writing*, *33*, 2361-2390.

Chen, J., & Zhang, L. J. (2019). Assessing student-writers' self-efficacy beliefs about text revision in EFL writing. *Assessing Writing*, *40*, 27-41.

Cho, M., Kim, Y. S. G., & Wang, J. (2023). Perspective taking and language features in secondary students' text-based analytical writing. *Scientific Studies of Reading*, *27*(3), 199-214.

Chung, H. Q., Chen, V., & Olson, C. B. (2021). The impact of self-assessment, planning and goal setting, and reflection before and after revision on student self-efficacy and writing performance. *Reading and Writing*, *34*(7), 1885-1913.

Coe, M., Hanita, M., Nishioka, V., & Smiley, R. (2011). An Investigation of the Impact of the 6+ 1 Trait Writing Model on Grade 5 Student Writing Achievement. Final Report. NCEE 2012-4010. *National Center for Education Evaluation and Regional Assistance*.

Culham, R. (2003). *6+ 1 traits of writing: The complete guide grades 3 and up*. New York, NY: Scholastic Professional Books.

De Jong, E. (2004). L2 proficiency development in a two-way and a developmental bilingual program. *NABE Journal of Research and Practice*, *2*(1), 77-108.

De La Paz, S., Ferretti, R., Wissinger, D., Yee, L., & MacArthur, C. (2012). Adolescents' disciplinary use of evidence, argumentative strategies, and organizational structure in writing about historical controversies. *Written Communication*, *29*(4), 412-454.

Ekholm, E., Zumbrunn, S., & Conklin, S. (2015). The relation of college student self-efficacy toward writing and writing self-regulation aptitude: Writing feedback perceptions as a mediating variable. *Teaching in Higher Education*, *20*(2), 197-207.

Erkan, D. Y., & Saban, A. İ. (2011). Writing performance relative to writing apprehension, self-efficacy in writing, and attitudes towards writing: A correlational study in Turkish tertiary-level EFL. *The Asian EFL Journal Quarterly*, *13*(1), 164-192.

Ferretti, R. P., & Lewis, W. E. (2018). Argumentative writing. In S. Graham, C. A. MacArthur, & M. Hebert. *Best practices in writing instruction*. The Guilford Press.

Garcia-Sanchez, J. N., & Fidalgo-Redondo, R. (2006). Effects of two types of self-regulatory instruction programs on students with learning disabilities in writing products, processes, and self-efficacy. *Learning Disability Quarterly*, *29*(3), 181-211.

Gillespie, A., & Graham, S. (2014). A meta-analysis of writing interventions for students with learning disabilities. *Exceptional Children*, *80*(4), 454-473.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C., Shanahan, C., & Project READi. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, *51*(2), 219-246.

Graham, S. (2018). A revised writer (s)-within-community model of writing. *Educational Psychologist*, *53*(4), 258-279.

Graham, S., Harris, K. R., Kiuhara, S. A., & Fishman, E. J. (2017). The relationship among strategic writing behavior, writing motivation, and writing performance with young, developing writers. *The Elementary School Journal*, *118*(1), 82-104.

Graham, S., Ng, C., Hebert, M., Santangelo, T., Aitken, A., Camping, A., Nusrat, A. (2024). *Can teaching writing enhance students' writing self-efficacy?* Manuscript submitted for

publication.

Graham, S., Wijekumar, K., Harris, K.R., Lei, P., Fishman, E., Ray, A., & Houston, J. (2019). Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction. *Elementary School Journal, 119,* 487-510.

Harris, K. R., & Graham, S. (2013). "An adjective is a word hanging down from a noun": Learning to write and students with learning disabilities. *Annals of Dyslexia*, *63*, 65-79.

Hwang, J. K., Lawrence, J. F., Collins, P., & Snow, C. (2017). Vocabulary and reading performances of redesignated fluent English proficient students. *TESOL Quarterly*, *51*(4), 757-786.

Kim, J., & Herman, J. L. (2009). A three-state study of English learner progress. Educational *Assessment*, *14*(3-4), 212-231.

Kim, Y. S. G., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranike, C. (2014). Evaluating the Dimensionality of First-Grade Written Composition. *Journal of Speech, Language, and Hearing Research*, *57*, 199-211.

Kim, Y. S. G., & Graham, S. (2022). Expanding the Direct and Indirect Effects Model of Writing (DIEW): Reading–writing relations, and dynamic relations as a function of measurement/dimensions of written composition. *Journal of Educational Psychology*, *114*(2), 215.

Kim, Y. S. G., & Park, S. H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing*, *32*(5), 1319-1343.

Klassen, R. (2002). Writing in early adolescence: A review of the role of self-efficacy beliefs. *Educational Psychology Review*, *14*, 173-203.

Lee, S. Y. (2005). Facilitating and inhibiting factors in English as a foreign language writing performance: A model testing with structural equation modeling. *Language Learning*, *55*(2), 335-374.

McCarthy, P., Meier, S., & Rinderer, R. (1985). Self-efficacy and writing: A different view of self-evaluation. *College Composition and Communication*, *36*(4), 465-471.

Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *The Journal of the Learning Sciences*, *19*(4), 539-568.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide (version 8)*. Los Angeles, CA: Muthén & Muthén.

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, *19*(2), 139-158.

Pajares, F., & Johnson, M. J. (1994). Confidence and competence in writing: The role of self-efficacy, outcome expectancy, and apprehension. *Research in the Teaching of English*, 313-331.

Pajares, F., Johnson, M. J., & Usher, E. L. (2007). Sources of writing self-efficacy beliefs of elementary, middle, and high school students. *Research in the Teaching of English*, 104-120.

Rayner, G., Papakonstantinou, T., & Gleadow, R. (2016). Comparing the self-efficacy and writing-related abilities of native and non-native English-speaking students. *Cogent Education*, *3*(1), 1179164.

Rodriguez, G. M. (2013). Power and agency in education: Exploring the pedagogical dimensions of funds of knowledge. *Review of Research in Education*, *37*(1), 87-120.

Rowan, B., & White, M. (2022). The Common Core State Standards initiative as an innovation network. *American Educational Research Journal, 59*(1), 73-111.

Sanders-Reio, J., Alexander, P. A., Reio Jr, T. G., & Newman, I. (2014). Do students' beliefs about writing relate to their writing self-efficacy, apprehension, and performance?. *Learning and Instruction*, *33*, 1-11.

Shah, P. M., Mahmud, W. H., Din, R., Yusof, A., & Pardi, K. M. (2011). Self-efficacy in the writing of Malaysian ESL learners. *World Applied Sciences Journal*, *15*(11), 8-11.

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter?. *Topics in Language Disorders*, *32*(1), 7-18.

Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, *81*(1), 91.

Steiss, J., Olson, C. B., Baker, T., Graham, S., Kim, Y.-S. G., Tate, T., Krishnan., J., Chung, H., Collins, P., Gilbertson, N. F., Moon, Y., & Wang, J (under preparation). *Effects of a Cognitive Strategies Intervention on Secondary Students' Argument Writing in History*. Manuscript under preparation.

Author (2024).

Sun, T., & Wang, C. (2020). College students' writing self-efficacy and writing self-regulated learning strategies in learning English as a foreign language. *System*, *90*, 102221.

Sun, T., Wang, C., & Kim, S. Y. (2022). Psychometric properties of an english writing self-efficacy scale: aspects of construct validity. *Reading and Writing*, *35*(3), 743-766.

Teng, M. F., & Wang, C. (2023). Assessing academic writing self-efficacy belief and writing performance in a foreign language context. *Foreign Language Annals*, *56*(1), 144-169.

U.S. Department of Education. n.d. Institute of Education Sciences, National Center for Education Statistics, *National Assessment of Educational Progress (NAEP), 2011 Writing Assessment.*

Wang, J., Graham, G., Kim, Y. S. G., & Steiss, J. (in preparation). Testing the Dimensionality and Generality Hypothesis Using an Argumentative and History Specific Writing Self-efficacy Scale. Manuscript submitted for publication.

Wang, C., & Sun, T. (2020). Relationship between self-efficacy and language proficiency: A meta-analysis. System, 95, 102366.

Williams, C., & Lowrance-Faulhaber, E. (2018). Writing in young bilingual children: Review of research. *Journal of Second Language Writing*, *42*, 58-69.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the

    evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*,

    83(1), *73*. https://doi.org/10.1037/0022-0663.83.1.73

Wiseman, C. S. (2012). A comparison of the performance of analytic vs. holistic scoring rubrics

    to assess L2 writing. *International Journal of Language Testing*, *2*(1), 59-92.

Zabihi, R. (2018). The role of cognitive and affective factors in measures of L2 writing. *Written*

    Communication, *35*(1), 32-57.

CHAPTER 4

Summary and Conclusions

This dissertation highlights the role of language and self-efficacy in students' writing performance, which aligns with theoretical models (Berninger et al., 2002; Graham, 2018; Kim & Graham, 2022). DIEW model (Kim & Graham, 2022) explicitly recognizes the role of various language skills and posits differential relations between language and writing performance.

First, the dissertation considers the role of measurement and whether the relations may differ as a function of measurements according to dynamic relations as a function of measurement in the DIEW model (Kim & Graham, 2022). Across all studies, we found differential relations as a function of measurement features, especially regarding which dimension of the target outcome is measured. Depending on the specific writing outcomes (writing productivity, writing fluency, and various dimensions of writing quality), the relations of language use features in writing and writing self-efficacy and writing performance may differ. Specifically, we found that the relations between language use in writing and writing performance vary as a function of writing genre, writing outcomes, and whether the writing task is text-based. For instance, we found that the relation between noun phrase complexity and writing performance was stronger for the expository genre than the narrative genre. In addition, we found that NLP linguistic features contribute to dimensions of writing quality differently: global cohesion measures and word count measures are more important in predicting the structure dimension and the evidence use dimension, whereas the lexical features are more important in predicting historical thinking and language use dimensions. These results suggest that it is important to consider genre- and discipline-specific features of the writing task and their specific linguistic demands on students. Writing assessments should also take the different

147

linguistic demands into consideration, especially when evaluating written compositions from various aspects or when examining a specific aspect of writing outcome.

Second, the dissertation focuses on linguistically diverse populations' writing development and the needs of those with limited language proficiency (ELs). Across all studies, whether students speak another language and their proficiency levels (operationalized as EL designation) were considered. Bilingual students with limited language proficiency were found to have specific linguistic demands in writing. For example, students with limited language proficiency may especially benefit from mastering complex syntactic structures like noun phrases during their writing development. In addition, they also had lower lexical sophistication and global cohesion in their written compositions. The findings highlighted the significant role of language proficiency in writing performance, showing that bilingual students who were not proficient and proficient in English displayed different linguistic features, which may be associated with their writing quality (Crossley et al., 2016; Kyle & Crossley, 2016). The findings underscore the necessity for instructional strategies that address the unique linguistic needs of bilinguals with limited language proficiency.

In addition to linguistic demands, this dissertation also investigated writing self-efficacy of linguistically diverse students. As the WWC model suggests, students with a stronger sense of efficacy are likely to value writing more because their perceived competence leads them to develop a deeper interest in writing, resulting in greater commitment and effort when composing. The dissertation focused on writing self-efficacy among linguistically diverse students, noting that EL and RFEP students may exhibit lower self-efficacy in certain dimensions of writing (i.e., ideation and revision). The results of the finding and previous findings with linguistic features in writing indicate the need to understand linguistically diverse students' various aspects of writing

development. For RFEP specifically, their language use in writing is on par with their monolingual peers. However, they had lower writing self-efficacy in ideation and revision, which may indicate their lower perceived ability due to their language experiences in school. As writing self-efficacy is associated with academic achievement, writing instruction should target specific needs in writing for linguistically diverse students so that they can be efficacious and motivated during the writing process.

Taken together, the dissertation illustrates the essential roles that language use in writing and writing self-efficacy play in writing and writing development in linguistically diverse students. Writing instruction that focuses on specific linguistic features, addresses the needs of different populations, and considers various dimensions of writing self-efficacy is promising in improving students' writing performance. Future research should further investigate relevant topics in linguistically diverse student populations due to the lack of research on writing development of linguistically diverse students.

Appendix A: Operationalization of Writing Performance and Syntactic Features

Category	Construct	Operationalization	Measures included
Writing performance	Writing quality	Commonly examined analytic aspects include content, organization, coherence, and language use; writing conventions were not counted toward writing quality.	
Writing performance	Writing productivity	Measures that quantify overall text length	Number of words, number of clauses, number of T-units, and number of sentences
Writing performance	Writing fluency	Productivity within a certain amount of time	Writing productivity measures of writing tasks with a time limit of 5 minutes or less were coded as writing fluency; measures that calculate speed of student writing.
Syntactic accuracy	Syntactic accuracy	Measures the extent to which the text is free of grammatical errors (excluding measures that tap into both grammar and writing conventions)	Ratio of error-free T-unit, number of grammatical errors, grammatical error per 100 words, percentage of grammatical sentences, grammaticality score, number of grammar errors, number of fragments per T-unit, run-on unit per T-unit, percentage of omission errors.
Syntactic complexity	Noun phrase complexity	Measures that depict complexity within the noun phrase level	Word/noun phrase, complex nominal phrase/T-unit, complex nominal phrase/clause, incidence of preposition, and number of modifiers/noun phrase
Syntactic complexity	Words per unit	Measures that depict length of clause/T-unit	Words per clause, words per T-unit.
Syntactic complexity	Subordination	Measures that depict ratio or frequency of the number of embedded clauses	Dependent clauses per clause, dependent clauses per T-unit, number of clauses per T-unit, subordinating conjunctions, and percentage of sentences containing embedded clauses.
Syntactic complexity	Words per sentence	NA	NA

Syntactic complexity	Left embeddedness	Number of words before the main verb	NA

Appendix B: Included Articles.

Acuna, D. B. (1986). *English as a second language (ESL) Development: an investigation into the relationship of the students' reading comprehension and writing ability at the college level in Puero Rico*. [Doctoral dissertation, New York University]. ProQuest Dissertations and Theses Global.

Arthur, S. V. (1981). *The effects of two writing treatments on the reading and writing of third graders*. [Doctoral dissertation, University of Georgia]. ProQuest Dissertations and Theses Global.

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?. *Reading and Writing*, *22*, 185-200. https://doi.org/10.1007/s11145-007-9107-5

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing*, *24*, 183-202. https://doi.org/10.1007/s11145-010-9264-9

Belanger, J., & Martin, R. G. (1984). The influence of improved reading skill on writing skill. *Alberta Journal of Educational Research, 30*(3), 194–212.

Benzehaf, B. (2017). Comparing learners' general proficiency levels with their writing productive ability: How correlated are they?. *Eurasian Journal of Applied Linguistics*, *3*(2), 43-58. https://doi.org/10.32601/ejal.460961

Bertrand, C. V. (1985). *Factors in Holistic Rating of Children's Writing*. [Doctoral dissertation, Rutgers University]. ProQuest Dissertations and Theses Global.

Black, E. M. (1997). *A text analysis of the argumentative writing in English by Spanish and English bilingual college students and by English monolingual college students in the Southwest United States.* [Doctoral dissertation, Indiana University of Pennsylvania]. ProQuest Dissertations and Theses Global.

Bromley, D. B. (1991). Aspects of written language production over adult life. *Psychology and Aging*, *6*(2), 296. https://doi.org/10.1037/0882-7974.6.2.296

Brown, A. J. (2012). *The factor structure of curriculum-based writing indices at grades 3, 7, and 10.* [Doctoral dissertation, Arizona State University]. ProQuest Dissertations and Theses Global.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42-65. https://doi.org/10.1016/j.jslw.2014.09.005

Carlisle, R. S. (1986). *The writing of Anglo and Hispanic fourth and sixth graders in regular, submersion, and bilingual programs.* [Doctoral dissertation, University of Illinois at Urbana-Champaign]. ProQuest Dissertations and Theses Global. https://doi.org/10.1017/S0272263100008135

Carlisle, R. S. (1989). The writing of Anglo and Hispanic elementary school students in bilingual, submersion, and regular programs. *Studies in Second Language Acquisition*, *11*(3), 257-280.

Chen, H., Xu, J., & He, B. (2014). Automated essay scoring by capturing relative writing quality. *The Computer Journal*, *57*(9), 1318-1330. https://doi.org/10.1093/comjnl/bxt117

Chew, C. R. (1978). *A study to determine the relationship between indexes of maturity and a quality grade in a written examination*. ERIC Document Reproduction Service No. ED 178 914

Chou, F. H. (1982). *Variables in College Composition*.

Collins, P., Tate, T. P., Won Lee, J., Krishnan, J. A., & Warschauer, M. (2021). A multi-dimensional examination of adolescent writing: considering the writer, genre and task demands. *Reading and Writing*, *34*(8), 2151-2173. https://doi.org/10.1007/s11145-021-10140-x

Connolly, S. K. (2006). *Peer-to-peer dialogue journal writing by Japanese junior high school EFL students*. [Doctoral dissertation, Temple University]. ProQuest Dissertations and Theses Global.

Couch, L. L. (1989). *An examination of the effects of sentence imitation practice upon the syntactic features of college freshman writers*. [Doctoral dissertation, University of Missouri]. ProQuest Dissertations and Theses Global.

Crossley, S. A., Allen, L. K., Snow, E. L., & McNamara, D. S. (2016). Incorporating learning characteristics into automatic essay scoring models: What individual differences and linguistic features tell us about writing quality. *Journal of Educational Data Mining*, *8*(2), 1-19.

Crossley, S. A., Kyle, K., Allen, L. K., Guo, 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).

Dizon, G. (2016). A comparative study of Facebook vs. paper-and-pencil writing to improve L2 writing skills. *Computer Assisted Language Learning*, *29*(8), 1249-1258. https://doi.org/10.1080/09588221.2016.1266369

Dobbs, C. L. (2014). Signaling organization and stance: Academic language use in middle grade persuasive writing. *Reading and Writing*, *27*, 1327-1352. https://doi.org/10.1007/s11145-013-9489-5

Donahue, B. J. (1982). *Indicators of quality in natural language composition*. Montana State University. [Doctoral dissertation, Montana State University]. ProQuest Dissertations and Theses Global.

Drijbooms, E., Groen, M. A., & Verhoeven, L. (2015). The contribution of executive functions to narrative writing in fourth grade children. *Reading and Writing*, 28, 989-1011. https://doi.org/10.1007/s11145-015-9558-z

Drijbooms, E., Groen, M. A., & Verhoeven, L. (2017). How executive functions predict development in syntactic complexity of narrative writing in the upper elementary grades. *Reading and Writing*, *30*, 209-231. https://doi.org/10.1007/s11145-016-9670-8

Eck, A. L. F. (1988). *A multivariant study of linguistic background characteristics and writing proficiency in Spanish*. [Doctoral dissertation, Temple University]. ProQuest Dissertations and Theses Global.

Fischer, O. H. (1981). *The relationship between the reflection-impulsivity dimension of cognitive style and selected qualitative and syntactic aspects of time-bound, first draft, expository compositions*. [Doctoral dissertation, Vanderbilt University]. ProQuest Dissertations and Theses Global. https://doi.org/10.1016/S0166-4115(02)80017-6

Flahive, D. & Snow, B. (1980). Measures of syntactic complexity in evaluating ESL

    compositions. In J. W. Oller, Jr.& K. Perkins (Eds.), *Research in language testing* (pp.

    171–176). Rowley, MA: Newbury House.

Francis, W. S., Romo, L. F., & Gelman, R. (2002). 14 Syntactic structure, grammatical accuracy,

    and content in second-language writing: An analysis of skill learning and on-line

    processing. *Advances in Psychology*, *134*, 317-337. https://doi.org/10.1016/S0166-

    4115(02)80017-6

Fukuta, J., Tamura, Y., & Kawaguchi, Y. (2019). Written languaging with indirect feedback in

    writing revision: is feedback always effective?. *Language Awareness*, *28*(1), 1-14.

    https://doi.org/10.1080/09658416.2019.1567742

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C.,

    Shanahan, C., & Project READi. (2016). Disciplinary literacies and learning to read for

    understanding: A conceptual framework for disciplinary literacy. *Educational

    Psychologist*, *51*(2), 219-246.

Goldman, S. R., & Scardamalia, M. (2013). Managing, understanding, applying, and creating

    knowledge in the information age: Next-generation challenges and opportunities.

    Cognition and Instruction, *31*(2), 255-269.

Golparvar, S. E., & Rashidi, F. (2021). The effect of task complexity on integrated writing

    performance: The case of multiple-text source-based writing. *System*, *99*, 102524.

    https://doi.org/10.1016/j.system.2021.102524

Greenberg, K. L. (1981). *The effects of variations in essay questions on the writing performance of college freshmen*. [Doctoral dissertation, New York University]. ProQuest Dissertations and Theses Global.

Guan, C. Q., Ye, F., Wagner, R. K., & Meng, W. (2013). Developmental and individual differences in Chinese writing. *Reading and Writing*, *26*, 1031-1056. https://doi.org/10.1007/s11145-012-9422-3

Gunderson, L., & Murphy, S. (1981). *Developmental characteristic of the writing of urban students at grades 2, 5, 8, and 11*.

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. https://doi.org/10.1016/j.asw.2013.05.002

Hartman, D. M. (1984). *An investigation into the predictive relationship of ten writing assessment variables to reading comprehension*. [Doctoral dissertation, University of Connecticut]. ProQuest Dissertations and Theses Global.

Hartnett, C. G. (1978). *Measuring Writing Skills* [Doctoral dissertation, University of Connecticut]. Educational Resources Information Center. https://files.eric.ed.gov/fulltext/ED170014.pdf

Hoffman, S. J. (1980). *Cognitive style and language complexity among gifted children*. [Doctoral dissertation, The George Washington University]. ProQuest Dissertations and Theses Global.

Holtz, D. J. (1987). *Interrelationships between the reading comprehension and writing achievement of college freshmen and their abilities to reconstruct scrambled expository paragraphs*. [Doctoral dissertation, University of Northern Colorado]. ProQuest Dissertations and Theses Global.

Humes, A. (1983). *Textual features in the descriptions and fictional narratives of young writers*.

Hwang, H. J. (1991). *The development of English writing proficiency among Korean students from ninth-grade to college juniors*. [Doctoral dissertation, University of California, Santa Barbara]. ProQuest Dissertations and Theses Global.

Irvine, A. (2007). *Extensive reading and L2 development: a study of Hong Kong secondary learners of English*. Edinburgh Research Archive. https://era.ed.ac.uk/handle/1842/29171

Jo, C. W. (2021). Short vs. extended adolescent academic writing: A cross-genre analysis of writing skills in written definitions and persuasive essays. *Journal of English for Academic Purposes*, *53*, 101014. https://doi.org/10.1016/j.jeap.2021.101014

Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, *21*(3), 264-282. https://doi.org/10.1016/j.jslw.2012.05.011

Kim, Y. -S. G., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranike, C. (2014). Evaluating the Dimensionality of First-Grade Written Composition. *Journal of Speech, Language, and Hearing Research*, *57*, 199-211. https://doi.org/10.1044/1092-4388(2013/12-0152)

Koutsoftas, A. D. (2018). Writing-process products of fourth-and sixth-grade children: A descriptive study. *The Elementary School Journal*, *118*(4), 632-653. https://doi.org/10.1086/697510

Koutsoftas, A. D., & Gray, S. (2013). A structural equation model of the writing process in typically-developing sixth grade children. *Reading and Writing*, *26*, 941-966. https://doi.org/10.1007/s11145-012-9399-y

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, *34*(4), 513-535. https://doi.org/10.1177/0265532217712554

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, *102*(2), 333-349. https://doi.org/10.1111/modl.12468

Langer, J. A. (1984). The effects of available information on responses to school writing tasks. *Research in the Teaching of English, 18,* 27-44.

Larkin, G., & Shook, R. (1978). *Interlanguage, the Monitor, and Sentence Combining*. Paper presented at the Los Angeles Second Language Research Forum.

Lee, C., Ge, H., & Chung, E. (2021). What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong. *System*, *97*, 102461. https://doi.org/10.1016/j.system.2021.102461

Li, Y. (1998). *Using task-based e-mail activities in developing academic writing skills in English as a Second Language*. [Doctoral dissertation, University of Arizona]. ProQuest Dissertations and Theses Global.

Li, Y. (2000). Assessing second language writing: The relationship between computerized

    analysis and rater evaluation. *International Journal of Applied Linguistics*, *127*(1), 37-51.

    h ttps://doi.org/10.1075/itl.127-128.02li

Longobardi, E., Spataro, P., & Pizzicannella, E. (2018). Handwriting, spelling, and narrative

    competence in the fictional stories of Italian primary-school children. *European Journal*

    of Psychology of Education, *33*, 277-293. https://doi.org/10.1007/s10212-017-0328-y

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.

    https://doi.org/10.1016/j.jslw.2021.100822

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.

    https://doi.org/10.1007/s11145-018-9853-6

Mackie, C., & Dockrell, J. E. (2004). The nature of written language deficits in children with

    SLI. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1469-1484.

    https://doi.org/10.1044/1092-4388(2004/109)

Mackie, C. J., Dockrell, J., & Lindsay, G. (2013). An evaluation of the written texts of children

    with SLI: The contributions of oral language, reading and phonological short-term

    memory. *Reading and Writing*, *26*, 865-888. https://doi.org/10.1007/s11145-012-9396-1

Markwood, R. A. (1982). *An Experimental study of the influence of freshman English on*

    students' writing skills at Western State College. [Doctoral dissertation, University of

    Colorado at Boulder]. ProQuest Dissertations and Theses Global.

Marlowe, J. P. (2019). *The effects of sentence-combining on the longitudinal development of syntactic complexity in L2 writing*. [Doctoral dissertation, Temple University]. ProQuest Dissertations and Theses Global.

Mason, L. H., Davison, M. D., Hammer, C. S., Miller, C. A., & Glutting, J. J. (2013). Knowledge, writing, and language outcomes for a reading comprehension and writing intervention. *Reading and Writing*, *26*, 1133-1158. https://doi.org/10.1007/s11145-012-9409-0

Matthews, J., & Wijeyewardene, I. (2018). Exploring relationships between automated and human evaluations of L2 texts. *Language Learning & Technology, 22*(3), 143–158.

Meskin, L. K. (1983). *An investigation into the interrelationship between the poetic and transactional modes of language function: a quasi-experimental study*. [Doctoral dissertation, Rutgers University]. ProQuest Dissertations and Theses Global.

McNeil, B. R. (2005). *A comparative statistical assessment of different types of EFL writing by Japanese college students*. [Doctoral dissertation, University of Birmingham]. ProQuest Dissertations and Theses Global.

Merrill, S. M. (1986). *Audience Adaptiveness in Job Application Letters Written by College Students.* [Doctoral dissertation, Purdue University]. ProQuest Dissertations and Theses Global.

Miller, N. L. (1985). *Syntactic and semantic characteristics of the written and oral language of reading disabled and normally reading school-aged children*. [Doctoral dissertation, University of Denver]. ProQuest Dissertations and Theses Global.

Nippold, M. A., LaFavre, S., & Shinham, K. (2020). How adolescents interpret the moral

    messages of fables: Examining the development of critical thinking. *Journal of Speech,*

    Language, and Hearing Research, *63*(4), 1212-1226.

    https://doi.org/10.1044/2019_JSLHR-19-00168

Nold, E. W., & Freedman, S. W. (1977). An analysis of readers' responses to essays. *Research in*

    the Teaching of English, *11*(2), 164-174.

Park, Y. M. (1987). *The influence of the task upon writing performance*. [Doctoral dissertation,

    University of Illinois at Urbana-Champaign]. ProQuest Dissertations and Theses Global.

Phillips Galloway, E., Qin, W., Uccelli, P., & Barr, C. D. (2020). The role of cross-disciplinary

    academic language skills in disciplinary, source-based writing: Investigating the role of

    core academic language skills in science summarization for middle grade writers.

    Reading and Writing, *33*(1), 13-44. https://doi.org/10.1007/s11145-019-09942-x

Pomplun, M., Sundbye, N., & Kelley, J. (1999). Exploring gender differences in answers to

    constructed-response reading items. *Scientific Studies of Reading*, *3*(1), 55-67.

    https://doi.org/10.1207/s1532799xssr0301_3

Pongsiriwet, C. (2001). *Relationships among grammatical accuracy, discourse features, and the*

    quality of second language writing: The case of Thai EFL learners. [Doctoral

    dissertation, West Virginia University]. ProQuest Dissertations and Theses Global.

Puranik, C., Duncan, M., Li, H., & Ying, G. (2020). Exploring the dimensionality of

    kindergarten written composition. *Reading and Writing, 33*, 2481-2510.

    https://doi.org/10.1007/s11145-020-10053-1

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. (2008). Assessing the Microstructure of

    Written Language Using a Retelling Paradigm. *American Journal of Speech-Language*

    Pathology, 17(2), 107-120. https://doi.org/10.1044/1058-0360(2008/012)

Roos, M. E. (1981). *Syntactic Maturity and Grading: A Correlational Study*. Education

    Resources Information Center. https://files.eric.ed.gov/fulltext/ED207071.pdf

Qian, L., Yang, Y., & Zhao, Y. (2021). Syntactic complexity revisited: Sensitivity of China's

    AES-generated scores to syntactic measures, effects of discourse-mode and topic.

    Reading and Writing, *34*(3), 681-704. https://doi.org/10.1007/s11145-020-10087-5

Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of

    Chinese EFL learners' writing performance. *Journal of Second Language Writing*, *33*, 3-

    17. https://doi.org/10.1016/j.jslw.2016.06.001

Ransdell, S. E., & Levy, C. M. (1994). Writing as process and product: The impact of tool,

    genre, audience knowledge, and writer expertise. *Computers in Human Behavior*, *10*(4),

    511-527. https://doi.org/10.1016/0747-5632(94)90044-2

Reynolds, D. W. (2001). Language in the balance: Lexical repetition as a function of topic,

    cultural background, and writing development. *Language Learning*, *51*(3), 437-476.

Reynolds, D. W. (2002). Linguistic and cognitive development in the writing of middle-grade

    English language learners. *Southwest Journal of Linguistics*, *21*(2), 101-120.

    https://doi.org/10.1111/0023-8333.00161

Saddler, B. T. (2002). *An analysis of the effects of sentence combining practice on the writing of students with above-and below-average sentence combining skills*. [Doctoral dissertation, University of Maryland, College Park]. ProQuest Dissertations and Theses Global.

Salas, N., Llauradó, A., Castillo, C., Taulé, M., & Martí, M. A. (2016). Linguistic correlates of text quality from childhood to adulthood. *Written and Spoken Language Development Across the Lifespan: Essays in Honour of Liliana Tolchinsky*, *11*, 307-326. https://doi.org/10.1007/978-3-319-21136-7_18

Samosir, S. (1991). *Holistic quality rating and selected linguistic text features in regular and learning-disabled fifth-graders' written compositions*. [Doctoral dissertation, State University of New York at Albany]. ProQuest Dissertations and Theses Global.

Schuler, P. (1992). *Assessment of English proficiency: A computer-assisted approach* [Doctoral dissertation, University of San Francisco]. ProQuest Dissertations and Theses Global.

Shoaf, J. (1986). *Measuring gain in writing proficiency qualitatively and quantitatively*. [Doctoral dissertation, The University of North Carolina at Chapel Hill]. ProQuest Dissertations and Theses Global.

Song, M. (2006). *A correlational study of the holistic measure with the index measure of accuracy and complexity in international English-as-a-second-language (ESL) student writings*. [Doctoral dissertation, University of Mississippi]. ProQuest Dissertations and Theses Global.

Spilton, R. (1986). *The effects of individualized language arts, sentence-combining, and traditional grammar on the syntactic maturity and quality of writing of a select group of*

eighth graders. [Doctoral dissertation, Georgia State University]. ProQuest Dissertations and Theses Global.

Stewart, M. F., & Grobe, C. H. (1979). Syntactic maturity, mechanics of writing, and teachers' quality ratings. *Research in the Teaching of English*, *13*(3), 207-215.

Stewart, M. F., & Leaman, H. L. (1983). Teachers' writing assessments across the high school curriculum. *Research in the Teaching of English*, *17*(2), 113-125.

Stilley, M. J. (1982). *An investigation of syntactic maturity in reading comprehension and writing ability.* [Doctoral dissertation, University of Pittsburgh]. ProQuest Dissertations and Theses Global.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, *14*(3), 153-173. https://doi.org/10.1016/j.jslw.2005.05.002

Swiggett, B. M. (1981). *A study of the relation of syntactic complexity to mode of discourse in the writing of eighth and twelfth graders*. [Doctoral dissertation, New York University]. ProQuest Dissertations and Theses Global.

Thomas, D., & Donlan, D. (1980). *Correlations between Holistic and Quantitative Methods of Evaluating Student Writing, Grades 4-12*.

Troia, G. A., Shen, M., & Brandon, D. L. (2019). Multidimensional levels of language writing measures in grades four to six. *Written Communication*, *36*(2), 231-266. https://doi.org/10.1177/0741088318819473

van Rijt, J. H., van den Broek, B., & De Maeyer, S. (2021). Syntactic predictors for text quality in Dutch upper-secondary school students' L1 argumentative writing. *Reading and Writing*, *34*(2), 449-465. https://doi.org/10.1007/s11145-020-10079-5

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing*, *24*, 203-220. https://doi.org/10.1007/s11145-010-9266-7

Wang, D. X. (2014). *Beyond CTT and IRT: Using an interactional measurement model to investigate the decision making process of EPT essay raters*. [Doctoral dissertation, University of Illinois at Urbana-Champaign]. ProQuest Dissertations and Theses Global.

West, G. K. (1988). *The construct validity of the holistic writing score: An analysis of the essay subtest of the College-Level Academic Skills Test*. [Doctoral dissertation, University of Florida]. ProQuest Dissertations and Theses Global.

Whitt, L. M. (1986). *The effects of sentence combining instruction on the growth of syntactic fluency in adult basic writers (18—25)*. [Doctoral dissertation, University of Virginia]. ProQuest Dissertations and Theses Global.

Wong, L. E. (1986). *Syntactical measures of cognitive complexity with written samples varying in cultural theme*. [Doctoral dissertation, Washington State University]. ProQuest Dissertations and Theses Global.

Xu, Y. (2020). *Second Language Writing Complexity in Academic Legal Discourse: Development and Assessment Under a Curricular Lens*. [Doctoral dissertation, Georgetown University]. ProQuest Dissertations and Theses Global.

Yang, W. (2014). *Mapping the relationships among the cognitive complexity of independent writing tasks, L2 writing quality, and complexity, accuracy and fluency of L2 writing*. [Doctoral dissertation, Georgia State University]. ProQuest Dissertations and Theses Global.

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, *28*, 53-67. https://doi.org/10.1016/j.jslw.2015.02.002

Yi, B., & Luo, S. (2013). Working memory and lexical knowledge in L2 argumentative writing. *Asian Journal of English Language Teaching*, *23*(1), 83-102.

Yoshimura, T. (2002). *Formal instruction of rhetorical patterns and the effectiveness of using the L1 in argumentative writing in an EFL setting*. [Doctoral dissertation, Temple University]. ProQuest Dissertations and Theses Global.

Zhang, L., & Liu, H. (2021). Genre effect on L2 syntactic complexity and holistic rating for writing quality of intermediate EFL learners. *Chinese Journal of Applied Linguistics*, *44*(4), 451-469. https://doi.org/10.1515/CJAL-2021-0029

Appendix C: *Analytic Coding Items and Reliability Information*

Category	Item	Description and criteria for score 7	IRR (one-point difference)
Structure	Claim	How well does the writing present a clear and compelling claim?	.96
	Introduction	How well does the writing advance a strong introduction with context, direction, and a clear claim?	.90
	Conclusion	How well does the conclusion relate to claims made throughout and give the writing a sense of completeness?	.94
	Body	How well does the body present a structure that enhances the central argument?	.93
Evidence use/ Historical thinking	Evidence	How well does the student use source material as evidence?	.96
	Commentary	How well does commentary interpret and use the textual evidence (to support a claim)?	.87
	Attribution	How well does the writing typically attribute evidence to sources?	.86
	Sourcing	To what extent does the writing use source material for sourcing (identifying and understanding the opinions, positioning, and bias of the author of a particular document; assessing credibility)?	.97
	Contextualization	To what extent does the writing use contextualization (locating actors and actions from source material in their time and place and their social and historical contexts?	.93
	Corroboration	To what extent does the writing use corroboration (checking sources against each other to determine the validity of a claim)?	.94
	Counterargument	To what extent does the writing present and address alternative viewpoints/ opposing perspectives?	.95

	Sentence Fluency	How well does the essay demonstrate sentence fluency and sentence flow?	1.00
Language use and conventions	Diction	How well does the essay demonstrate command of diction and word choice?	.91
	Conventions	How well does the essay demonstrate control of language and standard grammar conventions including spelling, capitalization, and punctuation?	1.00

Appendix D: Writing Self-efficacy Survey Items

Ideation:

1. I can think of many ideas for my writing.

2. I can think of many words to describe my ideas.

3. I can put my ideas into writing.

4. I know exactly where to place my ideas in my writing.

5. I can think of a lot of original ideas.

Self-regulation:

1. I can control my frustration when I write.

2. I can focus on my writing for at least one hour.

3. I can keep writing even when it's difficult.

4. I can think of my writing goals before I write.

5. I can focus on my writing and avoid distractions while I write.

Revision:

1. I can revise my paper so that my conclusion is stronger.

2. I can revise my paper so that I provide better support for my claim/thesis.

3. I can identify what I did well in my paper and what to improve upon.

4. I can revise my paper so that my introduction is stronger.

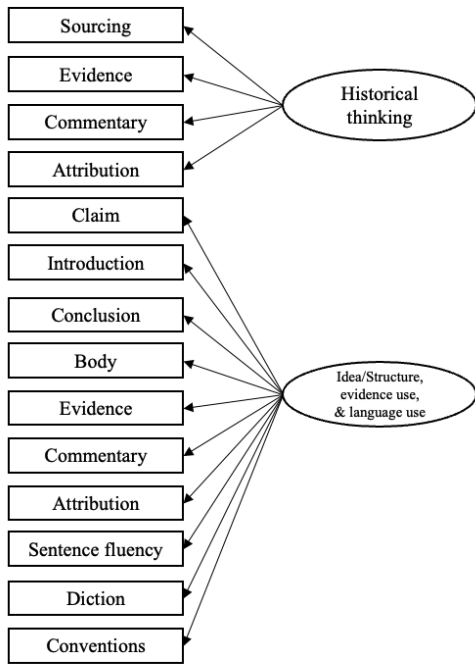**Appendix E: Figures of Confirmatory Factor Analysis of Dimensions of Writing Quality**

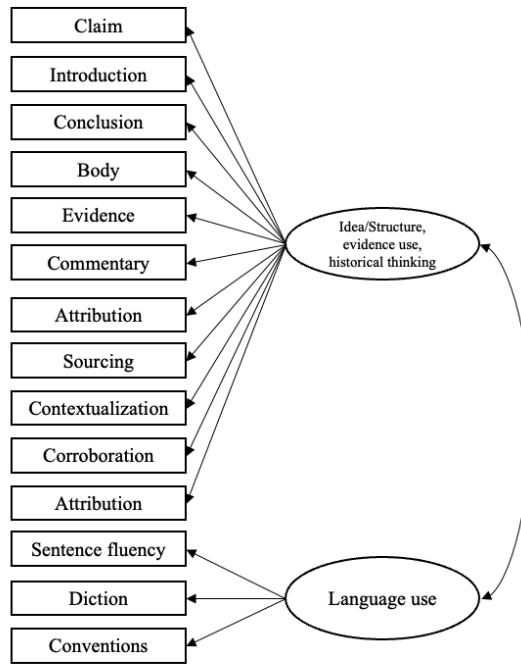(a)   (b)   (c)
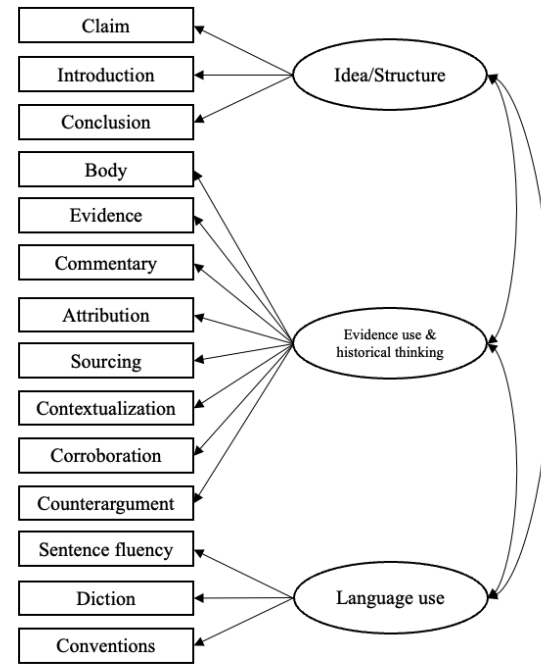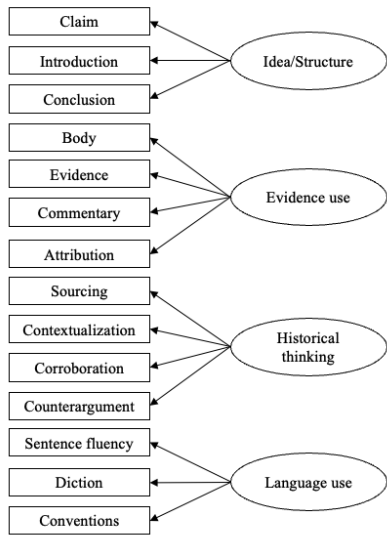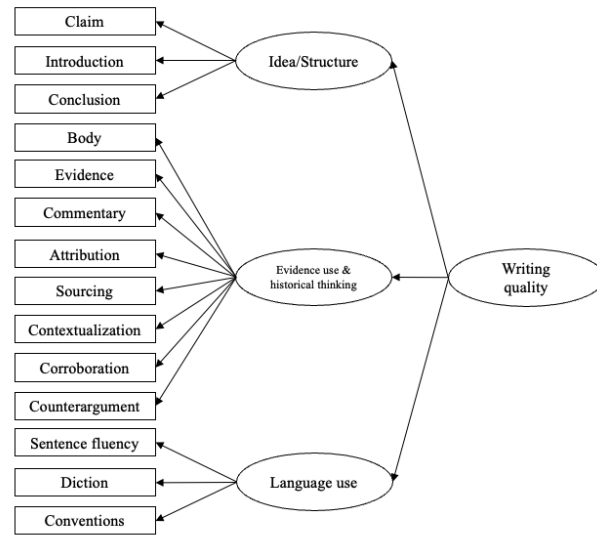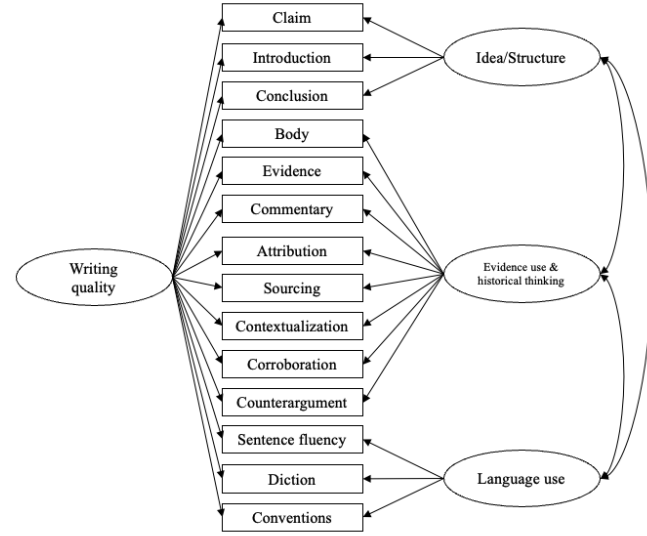
(d)

(e)

(f)
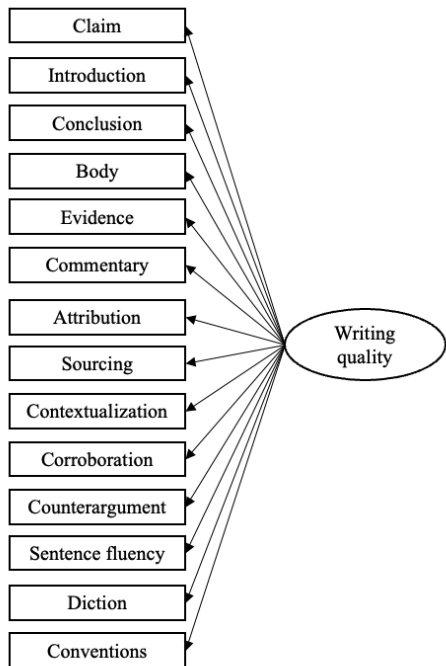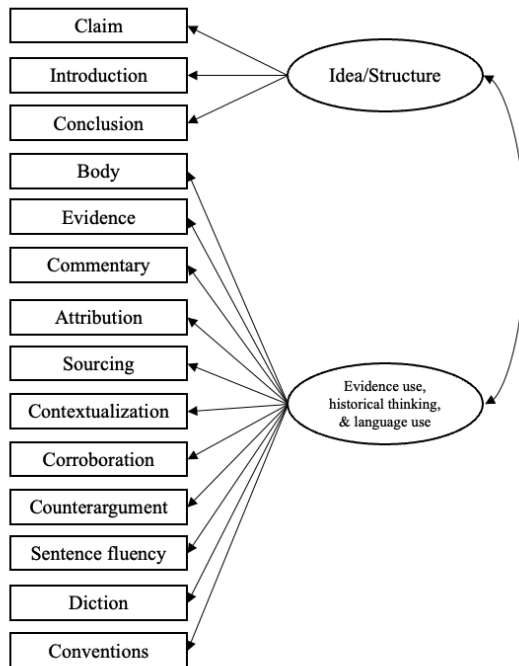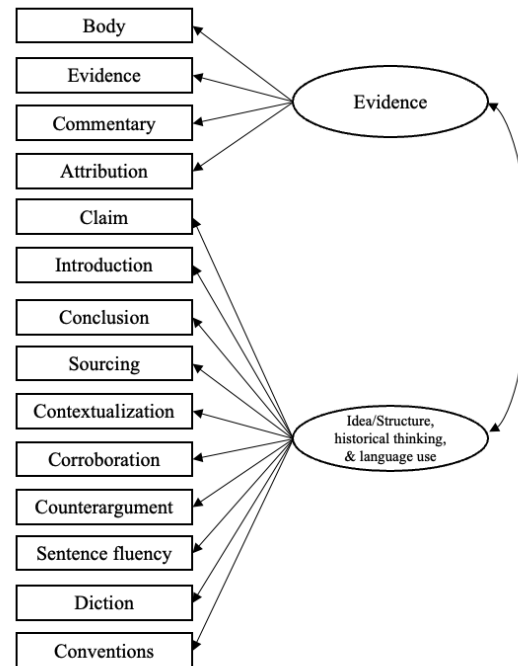
(g)

(h)

(i)

Appendix F: Model Fit Information of Confirmatory Factor Analysis of Dimensions of Writing Quality

	Model	χ^2	p	df	CFI	TLI	RMSEA	RMSEA 90%CI	SRMR	model comparison Model	χ^2	df	p
a	One-factor	803.96	<.001	77	.83	.79	.18	[.17, .19]	.06				
b	Two-factor: ideas/structure & others	781.66	<.001	76	.83	.80	.18	[.16, .19]	.06				
c	Two-factor: evidence use & others	730.49	<.001	76	.84	.81	.17	[.16, .18]	.06				
d	Two-factor: historical thinking & others	795.69	<.001	76	.83	.79	.18	[.17, .19]	.06				
e	Two-factor: language use & others	457.42	<.001	76	.91	.89	.13	[.12, .14]	.05	a	346.54	1	<.001
f	Three-factor	402.86	<.001	74	.92	.90	.12	[.11, .13]	.04	e	54.56	2	<.001
g	Four-factor	399.44	<.001	71	.92	.90	.12	[.11, .14]	.04	f	3.42	3	.33
h	Second order model (three-factor)	402.86	<.001	74	.92	.90	.12	[.11, .13]	.04				
i	Covariate bifactor model (three-factor)	161.69	<.001	60	.98	.96	.08	[.06, .09]	.02				

Note. Restricted bifactor model did not converge. There is Heywood case in model I; thus, no model comparison was conducted.