Developing Novel Approaches to Analyzing Vocabulary, Syntax, and Discourse Structure in Fifth-to-Eighth Grade Argumentative Writing

Citation

Permanent link
https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37370284

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Developing Novel Approaches to
Analyzing Vocabulary, Syntax, and Discourse Structure
in Fifth-to-Eighth Grade Argumentative Writing

Ziyun Deng

Catherine Snow, Co-Chair
Paola Uccelli, Co-Chair
Dana McCoy

A Thesis Presented to the Faculty
of the Graduate School of Education of Harvard University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

2021
Thesis Abstract

This dissertation consists of three studies on adolescents’ argumentative writing. Standards, assessments, and most research in the field of adolescent writing development rely primarily on holistic rubrics to analyze students’ written language. Evaluation of the major linguistic domains that contribute to effective writing, such as vocabulary, syntax, and discourse structure, is often incorporated only implicitly. At the same time, the latest U.S. national writing assessment results reported that as many as 72% of fourth graders and 76% of eighth graders did not reach the proficient level in argumentative writing (NAEP, 2011). Thus, understanding argumentative writing in greater detail is needed to advance theory and to inform instructional approaches that support writing development.

To better describe the language characteristics of adolescents’ written arguments, in this dissertation I present newly developed approaches to measuring three domains, vocabulary (Study 1), syntax (Study 2), and discourse structure (Study 3), using a database of argumentative essays written by a cross-sectional sample of fifth to eighth graders \((N = 512)\) from urban public school districts in the Northeastern and Mid-Atlantic regions of the United States. In Study 1, I generated the Vocabulary in Writing \((VW)\) latent construct from five indicators selected from writing development and corpus linguistics literatures: lexical diversity, lexical density, lexical specificity, lexical rarity, and academic vocabulary. In Study 2, I developed the Diversity of Advanced Syntactic Structures \((DASS)\) index to capture the variability in academic syntactic structures in
adolescents’ essays. In Study 3, I present the Argumentation Complexity Scale (ACS) developed on the basis of a qualitative coding scheme to identify key elements of written argumentative discourse. As evidence for the validity of each new approach, analyses in each study showed that participants’ scores in each measure (i.e., VW, DASS, ASC) were positively associated with grade and were predictive of writing quality (measured following the traditional method of assessing it via a holistic rubric).

The three studies together reveal the importance of examining fine-grained language skills in order to understand developmental trends and individual differences in adolescent writing. The findings provide insightful empirical evidence to inform more specific learning objectives, assessments, and pedagogy for emerging academic writers.
Table of Contents

Acknowledgments............................................................................................................................................v

Thesis Introduction .............................................................................................................................................1

Study 1: *Vocabulary in Writing (VW)* ..........................................................................................................3

Study 2: *Diversity of Advanced Syntactic Structures (DASS)* .................................................................73

Study 3: *Argumentation Complexity Scale (ACS)* ..................................................................................130

Thesis Conclusion.............................................................................................................................................199
To all writers who strive to express and explore their meanings
Acknowledgments

I would like to thank my wonderful committee, Catherine Snow, Paola Uccelli, and Dana McCoy, for their invaluable advice and generous support throughout my research journey.

I want to thank Catherine for transforming my understanding of language learning in her *From Language to Literacy* course which inspired me to step into the world of researchers when I was a master student at HGSE. I had the privilege to work with her on the *Catalyzing Comprehension through Discussion and Debate* research project out of which the inchoate ideas for this dissertation were born. For countless times, Catherine dragged me from rabbit holes to a-ha moments and scaffolded me to become a better reader and writer. I greatly appreciate the liberal atmosphere she created with insightful guiding questions, free discussions, and her iconic witty remarks. I do not know if I will ever become such a language master, sharp thinker, wise scholar, and charming person one day, but I will definitely keep a growth mindset on doing good works and remember to be deadline’s best friend just as I learned from her.

I am immensely grateful to Paola for her guidance, care, and encouragement that help me develop from a junior to advanced doctoral student in her *Language for Learning* research group. I am extremely lucky to have her guidance and to conduct my dissertation analyses within her writing development research project. Paola always made me see the blind spots in my writing and clarify my thinking. I have learned and enjoyed so much when working with her, like the days when we spread the data analysis printouts on the big table in Larsen and I breathed in the highest density of brainstorm oxygen. Paola’s research on academic language has been a topic of inquiry as well as a self-help English learning guide for me. As Paola has said, we are in such a unique research field as we are using language to share our understanding of language. I am beyond words to express my deepest gratitude to Paola for the support and love she showered me with, the precious gift that I will always bring with me in my work and life.

I want to thank Dana for her essential instruction and feedback that made the statistical analyses in this dissertation possible. Her insightful comments have also elevated this dissertation much beyond the methods section, from contents, structures, to writing styles. In Dana’s courses and office hours, I became a more confident and comfortable user of statistical methods for my own research questions and came to see numbers and letters as an integrated system to describe and express thoughts.

Furthermore, I want to express my infinite thanks to Pamela Mason, Robert Selman, Terry Tivnan, James Kim, Meredith Rowe, Gigi Luk, Paul Harris, Todd Rose, Andrew Ho, and so many other wonderful faculty and mentors for inspiring me in my doctoral journey. I want to give big a shout-out to Lisa Hsin, Erqian Xu, Min Hyun Oh, Qihan Chen, Katie Wingert, Dongkeun Han, Christina Strunk, Jia Yang, Zhongyu Wei, Xuanzhou Du, Chao Wang, Jiangcheng Zhu, Tianmeng, Xu,
Liwei Liu, Weike Zhang, Meixuan Li, Yuwei Rong, Yuelin Li, Meng Wang, Jin Ren, Xiaoyu Song, Sihan Yang, Jinyu Tong, Yuyang Gao, Young Chang, Mekides Mezgebu, and all the colleagues and research assistants whose diligent and careful transcribing, coding, and data management work made this dissertation possible. Big hugs to Ashley Lee, Laura Mesite, Tim Matthews, Maung Nyeu, Megan Powell Cuzzolino, Marcus Waldman, Zuhra Faizi, and all my wonderful cohort-mates who fight in the fields shoulder by shoulder with me and cheering me up from the orientation day to the night after defense. I want to thank Kelly Ann Robinson, Dorothy Bisbee, Clara Lau, Eric Zeckman, Joe McIntyre, Dylan Lukes, Gladys Aguilar, Mariam Dahbi, Linda Andreev, Wenjuan Qin, Dongying Li, all the amazing members of the SnowCats group, the L4L team, and the READS lab, as well as the wonderful Harvard peers, alums, and colleagues for giving me the most supportive intellectual family and community that one can ever dream of.

I am forever grateful to my parents, Hong Li and Yusen Deng, who have always encouraged me to explore the world since I was little, and to my grandmother, Yi Yan, who was a fighter at every moment of her life. I want to thank my childhood friend Ran Li for her decades of emotional support. I want to thank Bing Yang, Jun Wang, Fengfeng Gao, Yihong Gao, Rick Jackson, Shuhan Wang, and all my teachers and mentors for helping me grow at school and at work. In the past few years, especially during the COVID-19 pandemic, the anchors of my life are my dear friends Liao Cheng, Lin Du, Yan Zhou, Meimei Zhang, You Wang, Kaiwen Sun, Siyan Guo, Yi’an Xu, Yujia Li, Tiffany Wong, Amanda Manhke, and Sina Caliskan. I look forward to more new chapters and adventures with you.
An argument is “a reasoned, logical way of demonstrating that the writer’s position, belief, or conclusion is valid” (CCSS, 2010). In the last thirty years, more than two-thirds of fourth and eighth graders in the United States have performed below grade-level expectations in argumentative writing (Applebee, 1986; NAEP, 2011; Persky, Daane, & Jin, 2003). This persistent underperformance may reflect ever-increasing demands for writing skills for which assessments, curricula and educators have not been prepared. Traditionally, students were expected to demonstrate full argumentative writing skills only in high school or college (McCann, 1989). The U.S.’s Common Core State Standards, however, set this expectation for upper elementary and middle school students (CCSS, 2010). For example, current educational standards state that fourth graders are expected to state an opinion, provide reasons using facts and details, and link the opinion and reasons using words and phrases (p. 20); eighth graders are expected to distinguish the claim(s) from opposing claims, support the claim(s) with evidence, use words, phrase, and clauses to clarify the relationships among them, and maintain a formal style (p. 42). Given these new demands, more detailed research is needed to understand adolescents’ writing development in order to design more informative assessments and more supportive curricula and instructional approaches. To contribute to this gap in research, in this dissertation, I investigate developmental trends and individual differences in the language skills (i.e., vocabulary, syntax, and discourse structure) involved in argumentative
writing. In the following chapters, I will identify the research gaps in previous studies, describe the alternative approaches I developed to conceptualize and measure each language skill, and provide evidence to support the validity of my new approaches.
Study 1.

Vocabulary in Writing (VW) as a Unitary yet Multifaceted Construct in Fifth-to-Eighth Grade Argumentation

Abstract

Educational standards and national assessments in the U.S. require adolescents to “use precise words” (CCSS, 2010) and “make specific word choices” (NAEP, 2011) in argumentative writing. However, this domain that I call Vocabulary in Writing (VW) (i.e., words and word choices in students’ written production) is usually assessed as one loosely defined element that is inseparable from the holistic essay quality score. Therefore, the goal of the study is to propose and to provide validating evidence for a more refined and comprehensive model to conceptualize and operationalize Vocabulary in Writing (VW). First, I established a latent construct for adolescents’ VW by investigating and selecting candidate indicators from two sources: a) indicators already established in monolingual English-speaking adolescents’ writing research, and b) additional potentially applicable indicators used in research with English-as-a second or foreign-language writers. Second, I examined the validity of the VW latent construct in two ways: a) by testing the hypothesis that the latent construct would be a stronger predictor of overall writing quality when compared to the individual indicators; and b) by testing the hypothesis that students VW scores would be positively and significantly associated with grade. Using a school-context dilemma as the writing
prompt, I analyzed argumentative essays produced by a cross-sectional sample of fifth-to-eighth graders (N = 512) attending urban public school districts in the Northeastern and Mid-Atlantic regions of the United States. Essays were rated using automated language processing tools on five candidate indicators that represented distinct dimensions for VW: lexical diversity, lexical density, lexical rarity, lexical specificity, and academic vocabulary. Essays were also holistically rated by human scorers for writing quality. Results from structural equation modeling confirmed that the five-indicator measurement model for the VW latent construct fits well. Furthermore, the VW latent construct showed significant, positive, and moderate association with writing quality (r = .38); in contrast, the five individual indicators showed either positively significant yet weak associations (r = .12 to .23) or non-significant associations with writing quality. After controlling for students’ socioeconomic status, the VW factor scores of eighth graders were significantly higher than that of sixth and seventh graders, which in turn were significantly higher than those of fifth graders. The current study suggests that Vocabulary in Writing (VW) is a unitary but multifaceted domain jointly indicated by complementary yet distinct dimensions, and that it captures individual differences within and across grades throughout mid-adolescence. The study foregrounds the important role that productive vocabulary plays in writing quality and highlights the utility of comprehensive and fine-grained assessment tools to reveal students’ strengths and needs that can be relevant to inform curricula and instruction.
Introduction

Adolescent students in the United States have long been struggling with writing skills. About two-thirds of fourth and eighth graders in the U.S. have performed consistently below expected levels on writing over the last three decades (Applebee, 1986; Graham et al., 2014; NAEP, 2011; Persky, Daane, & Jin, 2003). The urgency of understanding students’ needs in writing is highlighted in the most recent national assessment results which report that as many as 72% of fourth graders and 76% of eighth graders did not reach the proficient level in argumentative writing (NAEP, 2011). In response to the persistent challenge, expectations of adolescent writing have been described via educational standards and students’ performance have been assessed in order to equip teachers to provide targeted instruction. However, although the standards and assessments have emphasized the general importance of writing, the requirements for the language skills that constitute writing performances, such as vocabulary, are typically described only vaguely. For example, the National Writing Assessment Framework for fourth and eighth graders describes word choice in a high-quality essay as “precise and evaluative” and in a low-quality essay as “often unclear and inappropriate” (NAEP, 2011). In these standards and rubrics, no operational definitions are provided for precise words or appropriate word choices, and the handful of sample essays included cannot sufficiently describe the full variety of students’ vocabulary production. Thus, more explicit expectations can be derived from the analysis of students’ data and from better understanding the extent to
which the vocabulary students produced in writing is related to developmental trends and/or to the quality of their essays.

In short, in order to facilitate students’ academic writing improvement in the upper elementary and middle school grades, research is needed to offer insight on the promising instructional area of *Vocabulary in Writing (VW)*, i.e., the words and word choices of students’ written production. Specifically, in this study, I developed a new model to conceptualize and operationalize *Vocabulary in Writing (VW)* for adolescents’ argumentative writing. In the following sections, I first review prior VW research with developing academic writers. Next, I propose my new VW model that integrates the indicators identified from prior research. Then, I examined the validity of the new VW model by testing its prediction of the essays’ overall writing quality and between-grade differences. Finally, I discuss research and practice implications of my new VW model and suggested directions for future research.

**Vocabulary in Writing: Prior Research**

Among studies focusing on academic writing in upper elementary and middle school, receptive vocabulary knowledge is often measured as a predictor of writing outcomes (e.g., Papadopoulou, 2007; Stæhr, 2008; Trapman et al., 2018), yet less attention has been paid to productive vocabulary. Extant research on developing academic writers’ VW consists of two main lines of investigation: a) research on adolescent English monolingual (henceforth EO) writers; b) research
on older English-as-a second or foreign-language (henceforth ESL/EFL) writers.

In this section, I will review the conceptual dimensions that have been identified to represent VW, as well as the measures that operationalize them, in the two respective lines. I will propose that some dimensions and measures adopted for ESL/EFL writers may also be applicable to EO writers, with the potential benefits of enhancing and deepening our understanding of the latter group. I will subsequently explain the necessity to examine whether the established and proposed dimensions for adolescent EO students’ VW indeed jointly indicate the same domain.

**Vocabulary in Writing for English Monolingual Students in Mid-Adolescence**

Writing research on English monolingual (EO) students in mid-adolescence (i.e., in upper elementary and middle school grades) has identified dimensions of Vocabulary in Writing (VW) such as lexical diversity and lexical density, and lexical sophistication. Their developmental trends and relations with overall writing quality have also been analyzed.

**Lexical Diversity**

Lexical diversity refers to the extent to which writers use a variety of words in a text (Jarvis, 2013). Lexical diversity was originally measured as word types (i.e., the total number of unique words in a text) or the type-token ratio (i.e., the total number of unique words divided by the total word count in a text) (Johnson,
1944). For example, the phrase *one for you and one for me* has a word type count of 5 (i.e., five unique words including *one, for, you, and, me*) and a type-token ratio of 0.71 (i.e. five unique words divided by a total of seven words). Different transformations on the type-token ratio have been introduced (Carroll, 1964; Malvern et al., 2004; McCarthy & Jarvis, 2010) to attempt to alleviate the impact of text length on the index. For example, a widely adopted measure is the index $D$, which adjusts the type-token ratio according to a probabilistic model based on random samples of words selected from the text (Malvern et al., 2004). The higher $D$ value indicates higher lexical diversity. Another widely adopted measure is MLTD (Measure of Textual Lexical Diversity, McCarthy & Jarvis, 2010), which is calculated as the mean length of sequential word strings in a text that maintain a given type-token ratio value. A higher MLTD value also indicates higher lexical diversity.

Lexical diversity, measured as word types, type-token ratio, or type-token ratio transformation, has been extensively investigated in the oral language of young children. It was found to predict quality and growth in early childhood language development (e.g., Pan et al., 2005; Rowe, 2012). Comparatively few studies have been conducted on vocabulary diversity in writing. The extant research has offered a crucial insight that its developmental trends differ by genre (narrative vs. expository). Some research has found that vocabulary diversity in narrative writing does not show noticeable increase in mid-adolescence: for example, Wood et al., (2020) found no significant difference in word types across
fourth to eighth grade; similarly, Chipere et al. (2001) found no significant difference in \( D \) values between fifth and eighth grade. On the contrary, some evidence suggests that vocabulary diversity in expository writing consistently develops at these grade levels. Berman and colleagues have found that the average \( D \) value of seventh grade students was higher than that of fourth grade students; the result was found not only for EO (English monolingual) students but also for students writing expository texts in other native languages such as French, Spanish, or Hebrew (Berman & Nir-Sagiv, 2010; Berman & Ravid, 2009; Berman & Verhoeven, 2002). Correspondingly, it is not surprising to find that the between-genre difference in vocabulary diversity seems to increase in mid-adolescence. For example, no significant difference in MLTD values was found between narrative and persuasive (i.e., one type of expository) writing at fifth grade (Olinghouse & Wilson, 2013); in contrast, higher \( D \) values were noticed in expository than narrative writing for seventh graders (Berman & Verhoeven, 2002).

Investigations on the association between lexical diversity and the overall quality of the text that an adolescent student writes (henceforth \textit{writing quality}) also found different results by genre. For example, Olinghouse and Wilson (2013) found that fifth-grade students’ MTLD values were positively associated with writing quality in their narrative texts, but surprisingly not in their persuasive texts. One possible explanation of the non-significant association is the lack of within-grade variability in adolescents’ expository writing quality. Specifically,
students in mid-adolescence are just starting learning to produce appropriate
global discourse features for the expository genre, such that the vast majority of
them only produced minimal representations of expository discourse at fourth
grade and partial expressions without full genre-typical structure at seventh grade
(Berman & Nir-Sagiv, 2007); in turn, expository writing quality may be at the
floor level or unstably developing during upper elementary and middle school
grades, and therefore not associated with the consistently developing lexical
diversity. Nonetheless, there is also a possibility that both lexical diversity and
expository writing quality are consistently developing during this time period, but
lexical diversity, as merely one of many dimensions for *Vocabulary in Writing*,
cannot sufficiently account for much variability in writing quality. Given the
scarcity of expository writing research on the mid-adolescent age group, especially
on the full range of upper elementary and middle school grade levels, it is unclear
which explanation of this paradox is more plausible.

**Lexical Density**

Lexical density is the extent to which writers use content words in text
(Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009; Halliday, 2004; Johansson,
2009; Read, 2000; Ure, 1971). Content words refer to the words that primarily
convey semantic content, such as nouns, adjectives, lexical verbs, and adverbs, in
contrast to function words which refer to the words that primarily signal
grammatical relations, such as articles, prepositions, conjunctions, pronouns, and
auxiliary verbs. Lexical density has typically been measured as the mean number of content words per clause or the proportion of content words among all words in a text. For example, the sentence *The small black cat jumped quickly into the brown box* includes seven content words (i.e., *small, black, cat, jumped, quickly, brown, box*) among a total of ten words, and therefore has a lexical density of 0.7. In contrast, the sentence *She told him that she saw a cat he liked* includes four content words (i.e., *told, saw, cat, liked*) among a total of ten words, and therefore has a lexical density of 0.4, lower than the previous example. High lexical density is a characteristic of written texts, whereas lower lexical density is more characteristic of oral communications (Biber & Conrad, 2009; Halliday, 2004; Johansson, 2009; Ure, 1971). The proportion of content words among all words is around .45-.55 in English textbooks for beginner to intermediate level learners (To et al., 2013).

As students are required to make transitions from more colloquial to more academic language in their school literacy environment (Snow & Uccelli, 2009), lexical density is expected to increase in mid-adolescence. Indeed, studies have found that lexical density at seventh grade is higher than that at fourth grade for Hebrew-speaking and Swedish-speaking secondary school students (Berman & Ravid, 2009; Strömqvist et al., 2002). Lexical density in English writing was found to be higher in expository than narrative writing at seventh grade and above (Berman & Nir-Sagiv, 2007). However, to my knowledge, no research has examined between-grade difference throughout mid-adolescence (from fifth to
eighth grade) in this dimension on EO writers. More evidence needs to be accumulated on whether lexical density as a dimension of *Vocabulary in Writing (VW)* develops during upper elementary and middle school, and in turn whether it is associated with expository writing skill in this age span.

**Lexical Sophistication**

Lexical sophistication refers to the “selection of low-frequency words that are appropriate to the topic and style of the writing, rather than just general, everyday vocabulary” (Read, 2000, p. 200). In the adolescent academic writing context, it refers to the extent to which a word is abstract, rare, and/or academic. Prior research on EO adolescent writers has operationalized lexical sophistication through word length, word origin, and nominal complexity rating (Bar-Ilan & Berman, 2007; Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009; Ravid, 2006). Word length refers to the number of syllables that a word contains. Polysyllabic words (i.e., words with three syllables or more such as *investigate*, *comprehensive*, or *transformation*) have been shown to be rarer than words with one or two syllables (e.g., *check*, *full*, or *change*); polysyllabic words are also more characteristic of academic texts than of colloquial discourse (Wimmer et al., 1996, as cited in Berman & Nir-Sagiv, 2007). Word origin refers to the historical source of a word. In English, Latinate origin words (e.g. *ancient*, *mystic*) have been shown to occur in more academic contexts with a later acquisition age than Germanic origin words (e.g. *old*, *strange*) (Biber et al., 1998). Nominal
complexity rating refers to researcher-developed scales to distinguish the nouns that occur in students’ writing samples from the lowest (i.e., concrete and frequent) to the highest level (i.e., abstract and rare) (Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009; Ravid, 2006). Accordingly, the indices of lexical sophistication include the proportions of polysyllabic words out of total words, the ratio of Latinate vs. Germanic origin words out of total content words, and the proportion of nouns at the highest level of abstraction out of total nouns.

The three lexical sophistication indices have been found to show developmental trends and genre difference in academic writing during mid-adolescence. All three of these measures have been found to show, on average, higher values at seventh grade than fourth grade, and in expository than in narrative writing (Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009). Additionally, word length and word origin have the advantage that they can be easily and reliably identified, so the lexical sophistication level of a text can be straightforwardly calculated. However, these two measures also have a few limitations. First, both are proxy measures, rather than direct measures of lexical sophistication. In other words, word length and word origin co-occur with abstractness, rarity, or academic register, but they do not measure words’ abstractness, rarity, or usage in academic register directly. Second, the two dichotomous indices are insufficient in capturing the nuances in lexical sophistication. For example, the two words ideological and intelligent are both polysyllabic and of Latinate origin, but the former is less frequent and conveys a
more complex meaning than the latter. Alternative operationalizations are needed to represent lexical sophistication on a continuum.

The nominal complexity rating has partially solved the problems by providing a direct measurement and a hierarchy of words on a four-point or ten-point scale (Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009; Ravid, 2006). Nonetheless, it evaluates only nouns, without considering other content words such as verbs. The human rating process is accurate and reliable, but immensely time consuming. In addition, the list of nouns included in the scales was developed based on the scope of writing samples collected from the original studies (Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009; Ravid, 2006), which was constrained by the writing prompts, students’ linguistic background, and the instruction they received. In turn, expansion, adaptation, and probably re-validation is needed when the scales are applied to other research or educational contexts, which makes the rating process more time-consuming.

Finally, although sophisticated words are considered to be abstract, rare, and academic, prior research on EO adolescents has emphasized the overlap among these three aspects rather than the unique variation of each aspect. In operationalization, each lexical sophistication index aims to address multiple aspects simultaneously; for example, a polysyllabic word is considered to be both more rare and more academic.

Vocabulary in Writing for ESL/EFL Learners
English-as-second or foreign-language (ESL/EFL) writing research has typically been conducted on college or adult students, a group older than the participants in the previously reviewed EO research but have a common status as developing academic writers. Various lexical dimensions, including those also sensitive to EO writers such as vocabulary diversity and lexical density, have been identified and found to predict ESL/EFL writing quality (as reviewed in Crossley, 2020; McNamara et al., 2010). Moreover, this line of research has also identified dimensions, namely Lexical Rarity, Lexical Specificity, and Academic Vocabulary, that I found especially relevant to describe lexical sophistication in students’ writing.

**Lexical Rarity**

I used lexical rarity to refer to a word’s frequency or range in a corpus. A corpus is a representative collection of texts or speech transcripts produced by language users in an environment. For example, a few commonly adopted corpora include British National Corpus (BNC, 2007) and Corpus of Contemporary American English (COCA; Davies, 2010). Frequency refers to the number of times a word occurs in a corpus; range refers to the occurrence of a word across several subsections of a corpus (Davies, 2009; Kyle et al., 2018; Halliday, McIntosh, & Strevens, 1964). A word with lower frequency or range is considered to be less commonly seen and less familiar to language users. For example, in the Corpus of Contemporary American English (COCA) -Spoken Language
Subcorpus (Davies, 2009), the word *begin* has a frequency of 112,407 (i.e., occurs 112,407 times in all speech transcripts) and a range of .26 (i.e., occurs in 26% of the speech transcripts), whereas the word *commence* has a frequency of 1,745 and a range of .001. Lexical rarity measured as frequency or range provides a scale on which all words, as long as they are part of the original corpus, in a text can be located and compared. Recent research on ESL/EFL college and adult learners’ argumentative writing found that the word frequency or range score in their texts predicted essays’ writing quality (Kyle & Crossley, 2016; Kim et al., 2018; Vögelin et al., 2019; Yoon, 2018). Given research that shows that adolescent EO students are also in the process of learning the language of academic texts (Berman, 2004; Uceeli et al., 2013; Uccelli et al., 2015; Uccelli, 2019), it is worth exploring if measures shown to be sensitive to differences in ESL/EFL learners, such as lexical rarity, might also be relevant to describe the vocabulary adolescents use in their argumentative writing.

**Lexical Specificity**

Lexical specificity refers to the degree of precision of word meanings. Textual Linguistics research has suggested that synonyms (i.e., words with similar meanings) can be compared on their precision based on the category they respectively represent (Fellhaum, 1998). For example, for the synonym pair of *mammal-animal*, *mammal* represents a category within *animal*, and therefore is considered to be more semantically specific; similarly, the word *declare* is a
considered to be more precise than the word *say*. By integrating multiple corpora and thesauruses (e.g., Grishman et al., 1993; Urdang, 1985), Fellbaum (1998) constructed a corpus called *WordNet* aiming to maximally encompass content words in English lexicon, in which pairs of synonyms are linked to form a hierarchical semantic framework (e.g., the highest-level word for nouns is *entity*). Utilizing this corpus, Kyle et al. (2018) developed algorithms to quantify how specific a noun or verb is based on its comparative position in the framework. For example, the three nouns *animal*, *mammal*, and *primate* respectively receives a value of 6.0, 9.0, and 9.83; the three verbs *say*, *declare*, and *proclaim* respectively receives a value of 2.82, 3, and 5. The lexical specificity score of a given text was calculated as the average specificity score per noun and/or verb. Research on ESL/EFL writing has found that texts with higher lexical specificity scores showed higher writing quality (Crossley et al., 2009; Guo et al., 2013; as cited in Kyle et al., 2018).

The concept of lexical specificity directly corresponds to the educational standards of “use precise language and domain-specific vocabulary” for upper elementary and middle school grades (CCSS, 2010) and the criterion of “precise word choice” in national writing assessment rubrics (NAEP, 2011); the corpus-based algorithm provides efficient operationalization via automated language processing. To my knowledge, no adolescent writing research has adopted this approach to analyze argumentative writing of fifth to eighth grade students, especially in populations that are representative of public urban schools. It is
worth exploring whether lexical specificity can reflect within- or between-grade variability in the population of students that teachers serve in public schools.

**Academic Vocabulary**

Academic vocabulary refers to the words or word families that are typically found in the academic register, a way of using language characteristic of school texts and texts in academic disciplines (Coxhead, 2000; Gottlieb & Ernst-Slavit, 2014; Nagy & Townsend, 2012). Textual linguistics studies have identified academic vocabulary in English lexicon. For example, Corpus of Contemporary American English - Academic Text Subcorpus (Gardener & Davies, 2014) has been compiled to words or word families that frequently occur in academic journals. Academic Word List (Coxhead, 2000) includes academic words used frequently in texts across disciplines and was been developed with the primary purpose of informing writing instruction in the university setting. The proportion of academic vocabulary (Coxhead, 2000) among all words in adult ESL/EFL students’ argumentative writing has been found to predict their writing quality (Kim et al., 2018). For EO adolescents, academic vocabulary has typically been measured as a receptive skill: researchers selected a small group of words from the abovementioned lists and examined whether students knew the word meanings. For example, the academic vocabulary knowledge of seventh and eighth grade students explained their achievements in standardized assessments across disciplines (Townsend et al., 2012). An intervention that integrated instruction of
cross-discipline academic words enhanced fourth to seventh grade students’ literacy achievement (Jones et al., 2019).

Although the importance of receptive academic vocabulary is widely acknowledged, few studies have investigated EO adolescents’ academic vocabulary production. Extant research suggests that middle-school writers are in the early stages of utilizing academic vocabulary. For example, Olinghouse and Wilson (2013) found that about only 1% of words in fifth graders’ narrative, persuasive, or informative writing were academic words (AWL; Coxhead, 2000). In contrast, about 10% of words in mature writers’ academic writing have been found to belong to this category (Coxhead, 2000). Not surprisingly, academic vocabulary was not found to predict writing quality at fifth grade (Olinghouse & Wilson, 2013). Given the lack of research on higher grade levels, it is unclear whether productive academic vocabulary develops throughout upper elementary and middle school and whether it predicts writing quality at other grade levels.

**Gaps in Research on Vocabulary in Adolescents’ Argumentative Writing**

In summary, research on English monolingual (EO) adolescents has identified several dimensions of *Vocabulary in Writing (VW)*, including lexical diversity, lexical density, and lexical sophistication. Students’ writing performance, as captured by these dimensions, improves across grades in upper elementary and middle school. Measures based on these dimensions have also been found to be sensitive to genre differences, with expository writing displaying,
on average, higher values than narrative writing. However, although lexical
diversity and lexical density have been clearly defined and operationalized, the
extant approach to identify sophisticated words via word length, word origin, and
nominal complexity scale has potentials to be improved on precision and
efficiency. On the other hand, research on college and adult ESL/EFL writers has
offered alternative approaches to define important aspects of lexical sophistication
by identifying three dimensions, i.e., lexical rarity, lexical specificity, and
academic vocabulary. Furthermore, the fine-grained automated measures used to
quantify students’ performance on these three dimensions can be more directly
and efficiently utilized, and cover more parts of speech. Therefore, it is worth
exploring whether they can potentially be applied to analyzing EO adolescents’
writing. To determine whether the potential approaches can be adopted, the
current study examines whether the established dimensions (i.e., lexical diversity
and lexical density) and potentially applicable dimensions (i.e., lexical rarity,
lexical specificity, and academic vocabulary) indeed jointly reflect the same skill
domain of VW.

In addition to expanding the measures used to explore vocabulary in native
language writing, the current study expands prior research in two ways: by
examining not only developmental trends but also individual variability within
grade and its relation to writing quality; and by zooming in into grade-level
differences from upper elementary to middle school. First, most studies on
adolescents’ VW have examined general developmental trends by describing
average performance at one grade level and testing for differences between grade levels (e.g., Berman & Nir-Sagiv, 2010; Berman & Ravid, 2009; Berman & Verhoeven, 2002). Fewer studies have examined individual differences within a grade level or predictions to writing quality. In the small number of studies where the relation was examined, a paradox has emerged that among individual dimensions of VW which have been found to be developing at this age some (e.g., lexical diversity) did not show a significant relation with persuasive writing quality, whereas other dimensions (e.g., word origin) showed significant and positive relations in the same genre (Olinghouse & Wilson, 2013). Given that VW can be conceptualized as encompassing several dimensions, it is possible that each dimension can only account for part of the variability in this skill domain. If so, a latent construct of Vocabulary in Writing (VW) which integrates various complementary dimensions may capture more variability, as well as provide more robust evidence on the relation with writing quality. Therefore, the strength of association between the VW domain, which is jointly indicated by the candidate measures, and the overall writing quality should be examined, in comparison with the strength of association between writing quality and each individual measure.

Last but not the least, the VW development during mid-adolescence have not been described comprehensively. Typically only one or two grade levels have been analyzed in a study, with the majority of the studies focused on the beginning of the upper elementary school (e.g., fourth or fifth grade) and near the end of middle school (e.g., seventh grade). Therefore, more research needs to be
conducted to examines the between-grade differences in detail by including more grade levels in upper elementary and middle school within a study.

To address the gaps in adolescent writing research on *Vocabulary in Writing*, the current study is driven by the following research questions:

**RQ 1:** Can *Vocabulary in Writing* (VW) be conceptualized as a single latent construct indicated by performance in a variety of vocabulary dimensions (i.e., lexical diversity, lexical density, lexical rarity, lexical specificity, and academic vocabulary) in argumentative writing throughout mid-adolescence?

**RQ 2:** Does the latent construct VW (established through addressing RQ 1) predict student essays’ writing quality?

   **RQ 2a:** Is there evidence that VW predicts writing quality?

   **RQ 2b:** Is VW a stronger predictor of writing quality than each of the individual dimensions?

**RQ 3:** Does the latent construct VW (established through addressing RQ 1) reflect students’ developmental trends?

   **RQ 3a:** What are the between-grade difference patterns in VW, controlling for students’ sociodemographic backgrounds?

   **RQ 3b:** Can the between-grade difference patterns also be found via the individual dimensions?
For RQ 1, I hypothesized that a measurement model for *Vocabulary in Writing (VW)* could be built based on five indicators that respectively represent research-based domains of vocabulary proficiency, i.e., lexical diversity, lexical density, lexical rarity, lexical specificity, and academic vocabulary. For RQ 2, I hypothesized that *VW* would display a positive association with student essays’ overall *writing quality*. Given that the *VW* latent construct would incorporate the variability of individual dimensions, I hypothesized that it would show a stronger positive relation to writing quality than each individual dimension. For RQ 3, I hypothesized that the *VW* latent construct would reveal developmental trends, with students in higher grades in general displaying higher performances, controlling for students’ sociodemographic backgrounds. I also hypothesized that *VW* would reflect developmental trends that otherwise would not be detected using only the individual dimensions.

**Methods**

**Participants**

The full sample of the study included 512 fifth-to-eighth graders from Title 1 urban public schools in the Northeastern and Mid-Atlantic regions of the United States. Participating students were part of the control group in a large-scale literacy intervention. Since the current study aims to investigate general developmental patterns and individual differences, rather than a treatment effect,
the treatment group was not included in the current study. Participants’ socio-demographic backgrounds are shown in Table 1. About half of the participants were female; about two-thirds of the participants were eligible for free/reduced-price lunch. The vast majority (97%) were native English speakers. The two largest race/ethnicity sub-groups in the sample were White (41%) and Black (41%), followed by Latinx (13%). The sample consisted of 20% fifth graders, 30% sixth graders, 30% seventh graders, and 20% eighth graders.

**Procedures**

I focused on participants’ responses to one writing prompt administered at the end of spring 2014. The writing prompt was: *Should we allow iPads in our classrooms?* The writing task was developed by the IES-funded *Catalyzing Comprehension through Discussion and Debate (CCDD)* team (Jones et al., 2019; LaRusso et al., 2016; Lawrence et al., 2015; Snow et al., 2009) to assess upper elementary and middle school students’ writing. Participants were given 20 to 25 minutes to write an argumentative essay and were provided with the following scenario: their school principal had decided to stop the school’s policy of providing iPads to students, thus participants were asked to take a position and to write an argumentative essay to be published by their school newspaper. Participants read a brief description of why iPads had been popular and why they were subsequently prohibited. In their essay, students were asked to give reasons to support their position, to try to convince people, to explain the impact on others,
and to discuss potential alternative resolutions to the problem. Participants wrote the essays in the paper-and-pencil format (see full prompt in Appendix 1.1).

**Data Preparation**

Prior to analysis, all the hand-written essays were transcribed using the *Code for the Human Analysis of Transcripts (CHAT)* conventions (MacWhinney, 2000). All spelling errors were corrected in the transcribed essay data in order to assure that human scorers of writing quality were not negatively biased by non-relevant misspellings or other orthographic features. Original files with misspellings were also preserved.

**Measures**

*Writing Quality Measure: Dimension Scores*

Students’ responses were scored using a holistic rubric developed by a team of language and writing researchers and informed by the NAEP (2011) Writing Framework. The rubric includes four dimensions: (1) Position: the number of sides that the essay considers; (2) Organization: the extent to which the essay is coherently structured. (3) Development of Ideas: the degree of depth, complexity, elaboration, and connectedness of ideas provided; (4) Clarity: the extent to which the essay conveys information in a precise and unambiguous manner. Each dimension was scored on a 4-point scale with higher scores indicating greater
quality. The essays were scored by a team of three research assistants, all graduate students specializing in education-related areas with prior experience as classroom teachers and blind to the study questions. In the group training for scoring team, a training set of essays was scored by all three scorers guided by the holistic writing rubric, which included anchor essays at each level. After this training, high inter-rater reliability was achieved on the basis of 20% of the sample, with Kendall's Coefficient of Concordance for Ordinal Response higher than .92 on all dimension scores (i.e., Position: .92; Development of Ideas: .99; Organization: .98; Clarity: .99).

**Vocabulary in Writing Measures**

Guided by prior research, as introduced in previous sections, five conceptually complementary dimensions were identified as promising for capturing the variability of *Vocabulary in Writing (VW)*. The dimensions include: vocabulary diversity, lexical density, word rarity, lexical specificity, and academic vocabulary. The current study selected one measure for each dimension. Computer programs were used to automatically calculate the values on each measure.

**Lexical Diversity: the Index $D$.**

Vocabulary diversity was measured using the index $D$ using the *Child Language Analysis (CLAN)* program (MacWhinney, 2000). The index $D$ is calculated based on adjusted type-token ratios fitting a probabilistic model.
(Malvern et al., 2004). The CLAN program calculated the $D$ in three steps. First, the program generated random subsamples of words within each text. Second, the type-token ratio for each subsample is calculated by dividing the number of unique words by the total number of words in the subsample. Third, the type-token ratios from subsamples were fitted in a probability curve to determine the best fit of $D$ for the text. The $D$ values tended to range from 10 to 100, and higher $D$ values indicate larger vocabulary diversity (McCarthy & Jarvis, 2010). Previous research has found average $D$ values around 50 at fourth grade and around 80 at seventh grade in expository writing (Berman & Verhoeven, 2002).

**Lexical Density: Proportion of Content Words.**

Lexical Density was measured as the proportion of content words per total words per text (Johansson, 2009; Perfetti, 1969), using the Child Language Analysis (CLAN) program (MacWhinney, 2000). Content words refers to nouns, non-auxiliary verbs, adjectives, and adverbs. Content words contrast with function words, such as auxiliary verbs, pronouns, articles, and prepositions. The possible range of Lexical Density is 0-1. Higher proportion of content words per total words represents higher Lexical Density. Previous research has found proportions around .30 in fourth and seventh grade students’ writing (Berman & Verhoeven, 2002; Strömqvist et al., 2002) and around .40 or higher in mature writers’ texts (Ure, 1971).
Lexical Rarity: Corpus-based Range Transformed.

Lexical Rarity was calculated as a transformation of corpus-based range scores in four steps. First, Contemporary American English – Spoken Subcorpus (COCA) (Davies, 2009) was chosen as the reference corpus for calculation because it corresponds to the geographical language varieties used by the current study’s participants. Second, for each content word (i.e., including nouns, non-auxiliary verbs, adjectives, and adverbs; not including articles, prepositions, conjunctions, pronouns, and auxiliary verbs) in a student’s essay, a word-specific range value was calculated as the proportion of transcripts in the reference corpus in which this word occurs. Third, the range score per essay, on the scale of 0 to 1, was calculated by adding all word-specific range values and dividing the sum by the total number of words added. The first three steps were conducted using the TAALES program (Kyle et al., 2018; Index name: COCA_spoken_Range_CW). Higher range scores, which by definition correspond to higher prevalence of the words in the language environment, represent mastery of more frequent and typically earlier acquired vocabulary. In contrast, lower range scores represent rarer words, and thus mastery of words typically acquired later. This directionality is opposite to all other measures included in the current study. For the purpose presentation clarity, as a final step, I transformed the range score per essay by multiplying by -1 and then adding 1, so that the final scores were aligned in directionality with other measures in the current study and stayed on a scale of 0-1.
Lexical Specificity: Position in a Semantic Hierarchy.

Lexical Specificity refers to the degree of precision in word meanings measured as their positions in a hierarchical semantic framework (Fellhaum, 1998) as included in the TAALES program (Kyle et al., 2018; Index name: hyper_noun_verb_s1_p1). First, each noun and verb in an essay received a specificity value; if the noun or verb had multiple meanings, the value was calculated using its most frequent meaning. Then, the lexical specificity score per essay was calculated by adding all word-specific values and dividing the sum by the total number of words added. Higher scores indicate higher skills in using specific and precise words. Given the algorithm was recently developed and not widely used, the range of possible scores for writing upper elementary and middle school grades was not found to have been reported by researchers; Nonetheless, previous research has found that in English language textbooks which used authentic texts targeting at the beginner-level ESL learners, the average score is 1.89 for verbs and 5.07 for nouns (Crossley et al., 2007).

Academic Vocabulary: Proportion of Academic Words.

Academic Vocabulary was calculated as the proportion of cross-disciplinary academic words per total words in a text. First, each word in a student’s text was identified as belonging to the Academic Word List or not (AWL; Coxhead, 2000). Second, the number of AWL words in the text was
divided by the total number of words in the text. The resulting number indicates the text’s Academic Vocabulary score as a normed count of academic words in the text. The Academic Vocabulary score was calculated by using the TAALES program (Kyle et al., 2018; Index name: all_awl_normed). The possible range of Academic Vocabulary is 0-1, with higher scores indicating higher Academic Vocabulary. Previous research found that the Academic Vocabulary scores were about .10 (i.e., 10% of the words in a text were academic vocabulary) for academic research articles (Vongpumivitch et al., 2009). About .07 for secondary school science textbooks (Coxhead et al., 2010), and on average .01 for fifth graders’ expository writing (Olinghouse & Wilson, 2013).

**Data Analysis**

For RQ1 that tests whether the established and potential dimensions can indeed jointly indicate one skill domain of *Vocabulary in Writing* (VW) for adolescent written argumentation, I used structural equation modeling to specify and confirm a measurement model reflecting VW. First, the five candidate measures (i.e., lexical diversity, lexical density, lexical rarity, lexical specificity, and academic vocabulary) were entered as observed indicators within a unidimensional measurement model. Second, Confirmatory Factor Analysis (CFA) was conducted to examine whether the measures jointly reflect a latent variable of VW. Given two of the VW candidate measures (i.e., lexical diversity
and academic vocabulary) are continuous variables with non-normal distributions, asymptotic distribution free method was applied for the estimation. Third, I accepted the measurement model on condition that it has: RMSEA ≤ .08, CFI ≥ .90, SRMR ≤ .08 (Hu & Bentler, 1999). For each indicator, if the standardized loading was ≥ .40, I accepted this measure for the latent construct of VW. If the standardized loading was < .40, I dropped this measure, conducted the CFA again, and re-checked the model fit.

For RQ2 that examines the whether the VW latent construct predicts the essays’ overall Writing Quality, I first used structural equation modeling to specify and confirm a measurement model where Writing Quality was jointly indicated by the four holistically scored dimensions (i.e., Position, Development of Idea, Organization, and Clarity), following the same CFA process and condition as for VW. As the four candidate measures for Writing Quality are continuous variables with non-normal distributions, asymptotic distribution free method was applied for estimation. Then, I tested whether the latent construct VW predicts the latent construct Writing Quality in a structural model by examining the significance and coefficient of the direct path from VW to Writing Quality. Last, I specified a different structural model by using the five individual indicators, rather than the single latent construct, of VW to predict the latent construct Writing Quality and examined the significance and coefficients of the five individual paths.

For RQ3 that explores the developmental trends for VW, I generated factor scores for VW based on the measurement model and examined whether students’
grade levels are associated with their VW factor scores, controlling for their sociodemographic backgrounds (i.e., gender, socioeconomic status, and English language learner status) in multiple regressions. I moved to a regression framework rather than conducting a different structural model for two reasons. First, with the current sample size, it is challenging for such a structural model with a large number of sociodemographic background variables as covariates and a comparatively small sample size at each grade level to achieve model convergence. Second, the factor scores have the advantage of providing numerical values of the latent construct for direct comparison. In the multiple regressions, I tested for the association between students’ grade levels and the VW factor score, controlling for students’ sociodemographic background. In the modeling process, I used the grade levels as categorical variables, with fifth grade as the reference group, to examine if there is statistically significant between-grade difference in VW factor scores, after controlling for students’ sociodemographic background (i.e., students’ gender, socio-economic status, and English language learner status). Students’ sociodemographic background variables were sequentially entered in the series of models. Significant control variables were retained in the final model, based on which I conducted pairwise comparison between any two grades. Then, using the same model that predicted the VW latent construct, based on the final model accepted for the VW factor scores, I fit a set of OLS regressions to examine the developmental trends for each of the five individual dimensions respectively. I conducted five different regressions to
examine the associations between students’ grade levels and each individual indicator respectively. I used the grade levels as categorical variables, with fifth grade as the reference group, to examine if there is statistically significant between-grade difference in an individual dimension. Based on that, I used pairwise comparisons to examine the between-grade difference on each individual indicator.

All statistical analyses were conducted using the STATA16 program. Given the lexical diversity score requires a minimum of 50 words in a text to be calculated, 38 essays with word counts of less than 50 ($M = 36, SD = 11, Min = 6, Max = 49$) were not included in the analyses, resulting in the final sample size of 474.

**Results**

**Descriptive Statistics: Vocabulary in Writing Candidate Measures and Writing Quality Dimension Scores**

Summary statistics of the *Vocabulary in Writing (VW)* individual measures and the *writing quality* dimension scores are reported in Table 1.2, and their correlations are reported in Table 1.3. All variables except lexical density, lexical rarity, and academic vocabulary displayed non-normal distributions. The five vocabulary measures displayed moderate or moderately strong correlations with each other: the weakest correlation was between lexical density and academic vocabulary ($r = .21$), whereas the strongest correlation was between lexical rarity
and lexical specificity ($r = .60$). For writing quality, the four quality dimensions showed moderate to strong correlations with each other: the weakest correlation is between Position and Organization ($r = .36$), whereas the strongest correlation is between Development of Ideas and Organization ($r = .60$). The correlations between individual $VW$ dimensions and individual $writing$ $quality$ dimensions were non-significant or weak (i.e., $r \leq .19$).

**Confirmatory Factor Analysis: Vocabulary in Writing ($VW$)**

As shown in Figure 1.1, the model for $VW$ fit the data well ($\chi^2 = 10.004, df = 5, p = .075$, RMSEA = .046, CFI = .970, SRMR = .026), confirming that this is an acceptable measurement model. All five standardized factor loadings were equal or larger than .4. Therefore, all five candidate measures (i.e., lexical diversity, lexical density, lexical rarity, lexical specificity, and academic vocabulary) were kept in the model as joint indicators for the latent construct $VW$.

**Vocabulary in Writing ($VW$) Latent Construct Predicting Writing Quality**

As shown in Figure 1.2, the measurement model for $Writing$ $Quality$ fit the data well ($\chi^2 = 1.705, df = 2, p = .426$, RMSEA = .000, CFI = 1.000, SRMR = .012), confirming that this is an acceptable model. All four standardized factor loadings were larger than .4. Therefore, all four candidate measures were kept as joint indicators for the latent construct $Writing$ $Quality$. 
As shown in Figure 1.3, a structural regression model was specified using the latent variable VW to predict the latent variable Writing Quality with asymptotic distribution free method estimation. The model fit the data well ($\chi^2 = 47.848$, $df = 26$, $p = .006$, RMSEA = .042, CFI = .947, SRMR = .043). VW positively predicted Writing Quality with a moderately strong strength ($r = .38$, $z = 7.56$, $p < .001$).

**Vocabulary in Writing Individual Dimensions Predicting Writing Quality**

Another structural model was specified using the five individual indicators for Vocabulary in Writing to predict the latent variable Writing Quality with asymptotic distribution free method estimation. The model fit the data well ($\chi^2 = 37.737$, $df = 17$, $p = .003$, RMSEA = .051, CFI = .918, SRMR = .030). As shown in Figure 1.4, the paths originating from Lexical Diversity, Lexical Density, and Lexical Specificity were not statistically significant. The path from Lexical Rarity was statistically significant and moderately positive ($r = .23$, $z = 3.68$, $p < .001$). The path from Academic Vocabulary was also statistically significant and positive, but with a weak strength ($r = .12$, $z = 2.14$, $p < .05$).

**Exploring the Developmental Trends of Vocabulary in Writing (VW)**

After the measurement model for VW was confirmed, factor scores were generated based on the model. The factor scores show a normal distribution ($M = .35$, $SD = 10.05$). The mean factor scores for each grade level were: -3.08 (8.80).
for fifth grade, .15 (10.14) for sixth grade, .52 (9.93) for seventh grade, and 4.66 (10.05) for eighth grade. Essay examples with low (10th percentile), medium (50th percentile), and high levels (90th percentile) of VW factor scores are presented in Appendix 1.2. The sample descriptive statistics showed a developmental trend, such that students in higher grade levels on average tended to have higher factor scores.

As shown in Table 1.4, the multiple regressions to predict VW factor scores showed that, after dropping the non-significant control variables, the final model (Model 5) included grade levels as the predictors and students’ socioeconomic status as a control variable. After controlling for students’ socioeconomic status, on average fifth- and sixth-grade essays were not statistically significantly different in VW factor scores, but seventh-grade essays were statistically significantly higher than those of fifth graders ($\beta = 3.26, SE = 1.24, p < .01$) and so were eighth grade essays ($\beta = 6.44, SE = 1.52, p < .001$). Post hoc pairwise comparison was conducted to further test for the difference between sixth, seventh, and eighth grade scores. Results showed that on average sixth and seventh grade essays were not statistically significantly different in VW scores, but eighth grade essays were statistically significantly higher than sixth grade ($F (1, 507) = 8.54, p < .01$), as well as higher than seventh grade ($F (1, 507) = 5.79, p < .05$).

Exploring Developmental Trends of Individual Dimensions
As shown in Table 1.5, the multiple regressions to predict each individual dimension showed that, after controlling for students’ socioeconomic status, significant higher performance of sixth, seventh, and eighth than fifth grade were found for Lexical Rarity \((F\ (4,\ 507) = 10.87, p < .001, R^2 = .08)\) as well as for Academic Vocabulary \((F\ (4,\ 507) = 9.88, p < .001, R^2 = .07)\); whereas Lexical Diversity, Lexical Density, and Lexical Specificity do not show statistically significant difference between fifth grade and other grade levels despite of some trends in their sample statistics. Post hoc pairwise comparison for Lexical Rarity showed no significant difference between sixth and seventh grade, but eighth grade essays are significantly higher than sixth grade \((F\ (1,\ 507) = 9.43, p < .01)\) and seventh grade \((F\ (1,\ 507) = 8.14, p < .01)\) respectively. Similarly, post hoc pairwise comparison for Academic Vocabulary shows no significant difference between sixth and seventh grade, but eighth grade essays are significantly higher than sixth grade \((F\ (1,\ 507) = 11.39, p < .001)\) and seventh grade \((F\ (1,\ 507) = 5.61, p < .05)\) respectively.

**Discussion**

The current study established a unitary yet multifaceted construct of *Vocabulary in Writing (VW)* including five indicators: Lexical Diversity, Lexical Density, Lexical Rarity, Lexical Specificity, and Academic Vocabulary. This novel measurement model expanded the repertoire of vocabulary measures for adolescent writing research. The latent construct provides a more informative and
more comprehensive measure found to be predictive of students’ essays’ overall writing quality and sensitive of developmental trends between grades 5 and 8.

**A Unitary Multifaceted Construct of Vocabulary in Writing (VW)**

To my knowledge, the current study is the first to integrate the five dimensions of VW into the same construct for analyzing developing academic writers in a diverse sample of US mid-adolescent students. It confirms that Lexical Diversity \( (r = .40) \) and Lexical Density \( (r = .55) \), the two dimensions that have been commonly used in prior English monolingual (EO) adolescent writing research, constitute important indicators of VW. It also confirms that Lexical Rarity, Lexical Specificity, and Academic Vocabulary, three dimensions examined in ESL/EFL writing research, also function as relevant indicators of the VW for EO adolescent writers.

The integration of dimensions from these two lines of research expands the repertoire of vocabulary measures in EO adolescent writing research. The three novel VW indicators (i.e., Lexical Rarity, Lexical Specificity, and Academic Vocabulary) have a few advantages. First, they provide a more precise conceptualization and more direct operationalization for EO adolescent writers’ lexical sophistication. Prior research has defined lexical sophistication as the extent to which a word is abstract, rare, and academic, but has emphasized the overlap of these aspects and typically used remote measures such word length and word origin to identify sophisticated words (Berman & Nir-Sagiv, 2010; Berman...
Ravid, 2009; Berman & Verhoeven, 2002; Olinghouse & Wilson, 2013). The current studyunpacks the lexical sophistication concept by addressingunique variation of each aspect represented by individual dimensions. Especially, the Lexical Specificity dimension directly addresses the expectation of “using precise words” in educational standards and national assessment rubrics (Common Core Standards Initiative, 2010; NAEP, 2011). The current study also provides direct operationalizations, such as the corpus-based range scores for Lexical Rarity and the percentage points for Academic Vocabulary, that are more direct measures of the relevant lexical domains. The second advantage of the novel VW indicators is that they provide more efficient and transparent operations with automated tools rather than relying on human scoring. Third, they expand the humanly scored word complexity scales that include only nouns to other parts of speech.

Among the indicators of VW, Lexical Rarity displayed the strongest factor loading \(r = .86\), followed by Lexical Specificity \(r = .72\), suggesting that they are especially sensitive indices of individual differences in students’ VW. Academic Vocabulary displayed a moderately strong factor loading \(r = .55\). Previous studies have found that Academic Vocabulary items were rarely produced by developing academic writers, such that less than 1% of the words in fifth graders’ essays were academic and therefore, in this prior research, this measure was eliminated from further analysis (Olinghouse & Wilson, 2013). Similarly, the current study found low production of Academic Vocabulary items in the fifth-to-eighth graders’ essays, that is, on average only 2% of the words in
an essay were academic at fifth grade, and only 3% of the words in an essay were academic overall. However, the differences, though seemingly small in scale, have been found to contribute to the variability of VW.

In short, the current study provides evidence that the various dimensions can jointly function as a valid indicator of a multifaceted construct of VW. Although the five indicators describe different characteristics of lexical performance as exhibited in students’ written products, their variance-covariance patterns analyzed through structural equation modeling suggest that they indicate one underlying skill.

**Vocabulary in Writing (VW) Predicting Writing Quality**

Students’ VW moderately and positively predicted argumentative essays’ Writing Quality ($r = .38$); in contrast, each individual indicator’s prediction was much weaker. Lexical Diversity, Lexical Density, and Lexical Specificity were not significantly associated with Writing Quality, while Lexical Rarity ($r = .23$) and Academic Vocabulary ($r = .12$) only weakly associated with Writing Quality.

Some of the findings on the individual indicators are consistent with the extant few studies on EO adolescent persuasive writing. For example, the current study found that the individual indicator Lexical Diversity is not a significant predictor of Writing Quality, which echoes the findings from Olinghouse and Wilson’s (2013) study that showed a non-significant association between Lexical Diversity and persuasive Writing Quality for fifth grade students. The current
study’s finding on the positive contributions of Lexical Rarity and Academic Vocabulary are consistent with prior research on ESL/EFL writing which has identified the two indicators as predictors of Writing Quality (Kyle & Crossley, 2016; Kim & Crossley, 2018; Vögelin et al., 2019; Yoon, 2018). Some of the findings on the individual indicators are slightly different from prior research. For example, Olinghouse and Wilson’s (2013) study found non-significant association between Academic Vocabulary and persuasive Writing Quality for fifth grade students, whereas the current study did find a weak positive association, perhaps because students in their study showed floor effect on Academic Vocabulary (less than 1% production), while students in the current study had higher production and were able to display larger variability.

The results of the current study support the hypothesis that as the five indicators in the VW measurement model conceptually complement one another, when they fit together, the latent construct can encompass more variability than individual indicators, and in turn serves as a more robust predictor for Writing Quality. For example, the current study found that Lexical Diversity did not significantly predict Writing Quality by itself; however, in the novel measurement model it is a significant indicator of VW, and VW predicts Writing Quality with moderate strength ($r = .38$). In other words, the results of the current study suggest that the latent construct VW has advantages over individual indicators in representing students’ productive vocabulary skills across different domains of lexical performance, and in turn in explaining more variability in Writing Quality.
Developmental Trends in Vocabulary in Writing (VW)

In the exploration of developmental trends, the current study found that fifth graders were not significantly different in VW factor scores than sixth graders, but were significantly lower than seventh graders, and seventh graders were in turn significantly lower than eighth graders, after controlling for students’ sociodemographic backgrounds. On the individual indicators, Lexical Rarity and Academic Vocabulary showed this same developmental trend, but Lexical Diversity, Lexical Density, and Lexical Specificity did not show any between-grade difference, after controlling for students’ sociodemographic backgrounds.

The finding on the VW factor scores is consistent with the general conclusions drawn from previous EO adolescent writing research that fourth graders had lower vocabulary performance than seventh graders in expository writing (Berman & Nir-Sagiv, 2010; Berman & Ravid, 2009; Berman & Verhoeven, 2002). Given that previous studies by Berman and colleagues covered fourth grade and seventh grade only, the current study adds to this body of research by including more grade levels and describing more detailed between-grade differences in upper elementary and middle school. In addition, the current study has an advantage of using a large sample with more than 500 students, in comparison to Berman and colleagues’ previous research which included about 20 students per grade level. Compared with the more homogeneous middle class
sample, the current study includes a more socioeconomic diverse sample that is representative of U.S. urban public schools.

On the other hand, the current study’s findings on the individual indicators have differences from the prior EO adolescent writing research. The current study found that Lexical Diversity and Lexical Density did not differ between any two grades, after controlling for students’ sociodemographic backgrounds, while the previous studies reported differences between fourth and seventh grade on the two dimensions. There are a few possible explanations for the discrepant findings. One possibility is that there might have been significant increase on the two dimensions between fourth and fifth grade, but it was out of the scope for the current study as the sample did not include fourth grade. Another possibility is that Berman and colleagues studied expository writing, in which the students could either express their opinions or provide information, while the current study focuses on argumentative writing, in which the students were only expected to take a position and convince a potential audience. The differences in students’ vocabulary performance need to be further examined across genres.

**Implications**

The current study responds to the urgency of understanding how to best support adolescents in argumentative writing by focusing on language as a potential area in need of instructional attention. The study focuses on vocabulary, a skill domain that has been broadly described as expected to be *precise and*
specific words or clear and appropriate word choices in educational standards, or in holistic scoring rubrics of essay quality (e.g., appropriate and specific word choice; inappropriate and unspecific word choice) embedded. Moving beyond a broad and vague description of vocabulary expectations in standards and holistic rubrics, the concept of Vocabulary in Writing (VW) proposed in this study has several implications.

First, echoing and expanding previous research on adolescent writing, the study highlights that VW consists of several individual dimensions that can make the abovementioned characteristics (i.e., precise, specific, clear, or appropriate) more specific and measurable.

Second, the study provides an efficient tool to expedite data processing for researchers, making it more plausible to analyze larger samples beyond the constraints of human scorer availability. The corpora on which the measurement model was built can be used as references or guide for educational practices. For example, curriculum developers may draw on the Academic Vocabulary list to include target words that would support students’ communications in the intellectual context in textbooks and design learning activities for this purpose. Although it is challenging for practitioners to directly adopt the measurement model, research and development specialists may potentially offer a service package that practitioners could outsource. The service package would include writing test administration, student output analyses, and score interpretation with individualized feedback.
Last but not least, the study advocates for dialogues between different traditions of writing research. The confirmed integration of indicators generated from EO adolescent writing research and ESL/EFL writing research supports the view that there is commonality among developing academic writers despite their first language backgrounds, and the two traditions of writing research can learn from each other.

**Limitations**

The study has several limitations. First, the measurement model for VW adopted in the current study may be one of many possible variations. For each of the five dimensions of VW examined, only one of many available measures was selected. For example, the proportion of academic words (Coxhead, 2000) was used as the Academic Vocabulary indicator in the current study, whereas indices based on other corpora (e.g., Corpus of Contemporary American English - Academic Text Subcorpus) could also potentially serve the same purpose.

The second limitation is that the results reflect students’ immediate performance, not edited careful rewriting; the results also only reflect one instance of writing, thus it reflects the proficiency as displayed in one writing performance, not a writer’s profile. Given all the written responses were based on one prompt, the type of words students produced were constrained by the nature of the topic. The 20-to-25-minute writing time only allowed a student to produce a first draft. Furthermore, the study only tested for students’ productive vocabulary without
testing their receptive vocabulary knowledge. In this design, if a word of interest
is not present in a student’s essay, it is unclear whether it is because the student
has not known the word, or has known the word but not retrieved it from memory
to integrate it into this particular draft.

In addition, the study has a limited scope of its generalizability. It examined
only one genre (i.e., argumentative) of writing. It used a cross-sectional, rather
than longitudinal sample, to analyze between-grade differences. Causal inferences
between VW and Writing Quality cannot be made, as the current study only tested
the relation as association. It is also unknown whether improvement on VW would
lead to higher scores on Writing Quality.

**Future Research**

Future research could be conducted to address the limitations in the current
study. More measures on each dimension of VW can be explored, and more
dimensions may potentially be identified. Future research can examine a variety of
writing prompts and genres as well as elicit responses from students at multiple
time points. Longitudinal samples could be analyzed in order to have more
accurate description of the developmental trends. Intervention studies on
productive vocabulary with randomized control design could be conducted to test
for the potential causal relations between VW and Writing Quality.

Furthermore, future research on adolescent writing could explore linguistic
domains besides vocabulary, such as syntax and discourse structures. Different
indicators and algorithms could be developed to measure each domain. This is important, in particular, because a high lexical performance might not necessarily coincide with a higher level of argumentation, for instance. Texts written in English by learners with different first language backgrounds or texts written in different languages could be analyzed and compared. Studies could be conducted on more age groups such as students in high school, college, or graduate school, to depict a comprehensive picture of academic writing development. Establishing a corpus on academic writing would be helpful for more detailed analyses and exploration.

**Conclusion**

The study constructed and confirmed a measurement model of *Vocabulary in Writing (VW)* for a cross-sectional sample of fifth-to-eighth grade students’ argumentative essays. The VW latent construct was jointly indicated by five dimensions: Lexical Diversity, Lexical Density (both established in adolescent writing research), Lexical Rarity, Lexical Specificity, and Academic Vocabulary (the last three adopted from ESL/EFL research on older learners). The VW latent construct positively and moderately predicted essays’ overall writing quality, whereas the individual dimensions of VW showed weakly positive or non-significant relations to writing quality. After controlling for students’ socioeconomic status, the VW factor scores for eighth graders were significantly higher than those for fifth, sixth, or seventh graders; among the three lower grade
levels, fifth graders were not significantly different from sixth graders but significantly lower than seventh graders in VW factor score. When examining developmental trends in individual indicators, two of the five indicators --i.e. Lexical Rarity and Academic Vocabulary-- showed the same trend as the VW factor score, while the other three individual indicators did not show any between-grade differences. The study suggests that Vocabulary in Writing (VW) is a complex domain that could be jointly indicated by various complementary dimensions, and therefore the latent construct can serve as a more robust predictor for writing quality and a more sensitive detector of developmental trends than the dimensions in singularity. The study provides evidence for the potential educational relevance of describing and evaluating the language skills for developing academic writers using a more fine-grained, quantifiable, direct, and efficient approach.
# Tables

## Table 1.1

*Participants’ Socio-Demographic Information (N = 512)*

<table>
<thead>
<tr>
<th>Socio-demographic Background</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>261</td>
<td>51%</td>
</tr>
<tr>
<td>Male</td>
<td>251</td>
<td>49%</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free/reduced lunch Eligible</td>
<td>345</td>
<td>67%</td>
</tr>
<tr>
<td>Free/reduced lunch non-eligible</td>
<td>167</td>
<td>33%</td>
</tr>
<tr>
<td><strong>Language Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Language Learner</td>
<td>14</td>
<td>3%</td>
</tr>
<tr>
<td>Non-English Language Learner</td>
<td>498</td>
<td>97%</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>208</td>
<td>41%</td>
</tr>
<tr>
<td>Black</td>
<td>209</td>
<td>41%</td>
</tr>
<tr>
<td>Asian</td>
<td>8</td>
<td>1.6%</td>
</tr>
<tr>
<td>Latinx</td>
<td>67</td>
<td>13%</td>
</tr>
<tr>
<td>Native/Pacific</td>
<td>2</td>
<td>0.4%</td>
</tr>
<tr>
<td>Mixed/Other</td>
<td>12</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>95</td>
<td>19%</td>
</tr>
<tr>
<td>6th</td>
<td>150</td>
<td>29%</td>
</tr>
<tr>
<td>7th</td>
<td>182</td>
<td>36%</td>
</tr>
<tr>
<td>8th</td>
<td>85</td>
<td>17%</td>
</tr>
</tbody>
</table>
Table 1.2

Summary Statistics of Dimension Scores: Vocabulary in Writing and Writing Quality ($N = 512$)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Vocabulary in Writing**

- **Lexical Diversity**: 72.12 (24.07), 78.46 (29.31), 79.25 (24.96), 81.77 (24.01), 78.15 (26.13)
- **Lexical Density**: .48 (.05), .48 (.05), .48 (.05), .49 (.05), .48 (.05)
- **Lexical Rarity**: -.46 (.06), -.44 (.06), -.44 (.06), -.41 (.06), -.44 (.06)
- **Lexical Specificity**: 4.07 (.56), 4.11 (.58), 4.19 (.54), 4.22 (.52), 4.15 (.55)
- **Academic Vocab**: .02 (.01), .03 (.02), .03 (.02), .04 (.02), .03 (.02)

**Writing Quality**

- **Position**: 2.96 (.79), 2.85 (.82), 2.92 (.86), 3.20 (.89), 2.96 (.85)
- **Develop of Ideas**: 2.45 (.75), 2.72 (.74), 2.78 (.86), 3.00 (.87), 2.74 (.82)
- **Organization**: 2.30 (.71), 2.64 (.81), 2.63 (.87), 2.97 (.95), 2.63 (.86)
- **Clarity**: 2.41 (.64), 2.63 (.64), 2.71 (.75), 2.96 (.84), 2.67 (.74)
<table>
<thead>
<tr>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Diversity</td>
<td>Domain</td>
<td>Specialty</td>
<td>Ranking</td>
<td>Vocabulary</td>
<td>Academic Position</td>
<td>Development</td>
<td>Organization</td>
</tr>
<tr>
<td>2.</td>
<td>Lexical</td>
<td>Lexical</td>
<td>Lexical</td>
<td>Lexical</td>
<td>Lexical</td>
<td>Lexical</td>
<td>Lexical</td>
<td>Lexical</td>
</tr>
</tbody>
</table>

Table 1.3 Pearson Correlations among Vocabulary in Writing Candidate Indices and Holistic Writing Quality (N=512)
Table 1.4

<table>
<thead>
<tr>
<th>VW</th>
<th>Model 1 VW</th>
<th>Model 2 VW</th>
<th>Model 3 VW</th>
<th>Model 4 VW</th>
<th>Model 5 VW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 6</td>
<td>2.933*</td>
<td>3.057*</td>
<td>2.617*</td>
<td>2.740*</td>
<td>2.480</td>
</tr>
<tr>
<td></td>
<td>(2.28)</td>
<td>(2.37)</td>
<td>(2.03)</td>
<td>(2.11)</td>
<td>(1.93)</td>
</tr>
<tr>
<td>Grade 7</td>
<td>3.600**</td>
<td>3.704**</td>
<td>3.379**</td>
<td>3.465**</td>
<td>3.258**</td>
</tr>
<tr>
<td></td>
<td>(2.90)</td>
<td>(2.97)</td>
<td>(2.71)</td>
<td>(2.78)</td>
<td>(2.63)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>7.741***</td>
<td>7.873***</td>
<td>6.592***</td>
<td>6.743***</td>
<td>6.439***</td>
</tr>
<tr>
<td></td>
<td>(5.28)</td>
<td>(5.36)</td>
<td>(4.32)</td>
<td>(4.39)</td>
<td>(4.23)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.422</td>
<td>-0.284</td>
<td>-0.293</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.48)</td>
<td>(-0.33)</td>
<td>(-0.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>^1FRL</td>
<td>-2.763**</td>
<td>-2.632**</td>
<td>-2.796**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.85)</td>
<td>(-2.69)</td>
<td>(-2.90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>^2 ELL</td>
<td>-2.564</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>-3.079**</td>
<td>-2.997**</td>
<td>-0.745</td>
<td>-0.851</td>
<td>-0.725</td>
</tr>
<tr>
<td></td>
<td>(-3.06)</td>
<td>(-2.71)</td>
<td>(-0.55)</td>
<td>(-0.63)</td>
<td>(-0.56)</td>
</tr>
<tr>
<td>R²</td>
<td>0.053</td>
<td>0.055</td>
<td>0.070</td>
<td>0.072</td>
<td>0.069</td>
</tr>
</tbody>
</table>

Note. Grade 5 set as the reference group

^1FRL: Free-reduced lunch status; ^2 ELL: English Language Learner Status

t statistics in parentheses  * p < 0.05, ** p < 0.01, *** p < 0.001
Table 1.5
Vocabulary in Writing Individual Dimensions Predicted by Grade Levels
(N = 474)

<table>
<thead>
<tr>
<th></th>
<th>Lexical Diversity</th>
<th>Lexical Density</th>
<th>Lexical Rarity</th>
<th>Lexical Specificity</th>
<th>Academic Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 6</td>
<td>4.674</td>
<td>-0.005</td>
<td>0.0192*</td>
<td>0.030</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>(1.32)</td>
<td>(-0.68)</td>
<td>(2.36)</td>
<td>(0.41)</td>
<td>(2.51)</td>
</tr>
<tr>
<td>Grade 7</td>
<td>6.070</td>
<td>-0.008</td>
<td>0.022**</td>
<td>0.106</td>
<td>0.011***</td>
</tr>
<tr>
<td></td>
<td>(1.78)</td>
<td>(-1.28)</td>
<td>(2.76)</td>
<td>(1.51)</td>
<td>(3.78)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>4.935</td>
<td>0.005</td>
<td>0.046***</td>
<td>0.121</td>
<td>0.018***</td>
</tr>
<tr>
<td></td>
<td>(1.19)</td>
<td>(0.63)</td>
<td>(4.73)</td>
<td>(1.40)</td>
<td>(5.13)</td>
</tr>
<tr>
<td>FRL†</td>
<td>-10.46***</td>
<td>-0.007</td>
<td>-0.018**</td>
<td>-0.064</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(-3.99)</td>
<td>(-1.36)</td>
<td>(-2.97)</td>
<td>(-1.17)</td>
<td>(-1.73)</td>
</tr>
<tr>
<td>_cons</td>
<td>80.76***</td>
<td>0.490***</td>
<td>-0.447***</td>
<td>4.127***</td>
<td>0.025***</td>
</tr>
<tr>
<td></td>
<td>(23.02)</td>
<td>(73.64)</td>
<td>(-54.75)</td>
<td>(56.48)</td>
<td>(8.70)</td>
</tr>
</tbody>
</table>

R²       | 0.046             | 0.015           | 0.079          | 0.012               | 0.072               |

Note. Grade 5 set as the reference group
†FRL: Free-reduced lunch status; ‡ELL: English Language Learner Status
t statistics in parentheses  * p < 0.05, ** p < 0.01, *** p < 0.001
Figures

Figure 1.1

Vocabulary in Writing (VW) Measurement Model  \( (N = 474) \)

Note. Standardized factor loadings displayed
Figure 1.2

*Writing Quality Measurement Model with Standardized Factor Loadings*

*(N = 474)*

*Note.* Standardized factor loadings displayed
Vocabulary in Writing (VW) Latent Construct Predicting Writing Quality

Figure 1.3

(\textit{N} = 474)
Appendices

Appendix 1.1

Argumentative Writing Prompt

Should we allow iPads in our classrooms?

Imagine that your school decided to give iPads to all students, so everyone has the opportunity to use iPads for learning. Most teachers like the iPads because all students can use online dictionaries and search the internet for interesting information. But in the last few weeks some students have used their iPads to post videos with mean comments about other kids online. Now, many parents are worried about this situation. Because of this problem, your Principal has decided that iPads will no longer be allowed in school!

Write an article for the school newspaper that argues for or against allowing iPads at your school.

* Make sure to give specific reasons to support your position and to convince the people who read the article to agree with you.
* Explain how the Principal’s decision can impact you and others.
* Discuss other things that the school community could do to solve the iPad problem.
Appendix 1.2

Sample Essays with Low, Medium, and High Vocabulary in Writing (VW) Factor Scores

Vocabulary in Writing (VW) Factor Score = -11.63 [Low: 10th percentile]

I think that they should take iPads. I think that because if they keep the iPads. It would get worser and worse. Also if they do no take it there would be many fights. You would no want fight like that. I think they should keep them away from the bad people. I agree because no kid should be allowed to bully. It is not only bullying it is cyber bullying. The bad things they are doing with iPads are embarrassing to the principal also to the school. Bullying is the wrong thing to do especially if you are getting bullied. This should stop. The school community could solve this iPad problem from discipline. I say discipline because the cyber bullying has to stop. My idea is to take the iPads away from the schools that are using them to bully other kids. Another idea is to try to get their parent to sit in the school with their kids like kindergarteners because they do no know how to get.

This is what I think.

[ID: 2C51404020024]

Basic Sociodemographic Information

- Grade: 5
- Gender: Female
Vocabulary in Writing (VW) Dimension Scores

- Lexical Diversity: 58.33
- Lexical Density: 0.43
- Lexical Rarity: 0.50
- Lexical Specificity: 3.58
- Academic Vocabulary: 0.01

Writing Quality Dimension Scores

- Position: 3.5
- Development of Ideas: 2
- Organization: 2
- Clarity: 3

Vocabulary in Writing (VW) Factor Score = .38 [Medium: 50th percentile]

I think that children and all students all over the world do not really need iPads in order to learn. All kids need to learn by going outside and learning about nature. Sure iPads are good for taking photos. But that is what cameras are for. And iPads are great at calculating information. My point is when I grow up I would love to be a second grade school teacher. And I do not want my kids looking up definitions all day on electronics. They are going to be outside maybe
counting how many trees there are around the school then count how many flowers there are then find the difference between the two numbers. You can pretty much learn or do any subject outside when it is nice out of control. During the winter time you can utilize your smart board and chalk board to teach lessons. I think all kids learn better if they are all on the same page going at the same pace with iPads you could finish before someone else and just try to get it done so fast that you do not learn anything. However when the teacher sets a good pace for the kids' brains to seek information at everyone is getting educated since kids' brains absorb most information when they are young. Kids can communicate and make friends easier if they are working on a worksheet together. When with iPads you can not practice making friends because all you are doing is maybe playing a school related game or looking up information when you could be learning how to add and laughing and playing with your friends outdoors or indoors at the same thing.

[ID: C20104020011]

**Basic Sociodemographic Information**

- Grade: 6
- Gender: Female
- Free/Reduced Lunch Status: No

**Vocabulary in Writing (VW) Dimension Scores**
In light of the recent decision to disallow school iPads I would like to personally note that they were a terrible idea in the first place. I think they were a desperate attempt to bring technology into the classroom and I have no idea how it was expected that anyone do something productive with them. As firmly that I believe technology in the classroom could work indiscriminately giving everyone an iPad is not the way to do it. I suggest the issuing of laptops to students with a passing grade for a number of reasons. One laptops have a physical keyboard making it feasibly possible to type a long paper. Two the windows operating system has a broad set of restriction tools to keep students from doing anything
non-educational. Three laptops are cheaper. Four the windows operating system has better software. In conclusion taking away the iPads away was more of an ultimate solution although giving them out in the first place was a mistake. [ID: C20106020017]

**Basic Sociodemographic Information**

- Grade: 8
- Gender: Male
- Free/Reduced Lunch Status: No

**Vocabulary in Writing (VW) Dimension Scores**

- Lexical Diversity: 90.5
- Lexical Density: 0.49
- Lexical Rarity: 0.65
- Lexical Specificity: 5.21
- Academic Vocabulary: 0.05

**Writing Quality Dimension Scores**

- Position: 3.5
- Development of Ideas: 4
- Organization: 4
- Clarity: 4
References


Consortium. Retrieved from www.natcorp.ox.ac.uk


Vögelin, C., Jansen, T., Keller, S. D., Machts, N., & Möller, J. (2019). The


Study 2.

Diversity of Advanced Syntactic Structures (DASS) in Writing Predicts Argumentative Writing Quality and Receptive Academic Language Skills of Fifth-to-Eighth Grade Students

Abstract

Prior research on adolescent writing tends to use omnibus length-based measures, such as Mean Length of Clauses (MLC), to describe and evaluate students’ syntactic performance in writing. However, such measures provide insufficient descriptive information about students’ production of the syntactic structures that support writing at school. This study aims to: (1) develop and introduce a novel index, Diversity of Advanced Syntactic Structures (DASS), to measure the variability in fifth-to-eighth graders’ syntactic performance in argumentative essays; and (2) provide evidence of the validity of the DASS by examining this index in relation to participants’ grade levels, their argumentative writing quality, and their receptive academic language skills. To develop DASS, I selected 7 types of syntactic structures that have been identified as characteristic of school-based texts in adolescence: adverbial clause, clausal complement, clausal prepositional complement, relative clause as modifier, clausal subject, noun as modifier, and passive voice. Students’ essays were coded for the presence or absence of each advanced syntactic structure, and the total number of types of structures present in a text determined the DASS score. A cross-sectional sample
of fifth to eighth graders ($N = 512$) wrote argumentative essays responding to a school policy controversy. DASS scores in fifth grade were significantly lower than those in seventh and eighth grade. DASS significantly and positively predicted students’ writing quality as well as receptive academic language with a moderately strong strength, controlling for students’ grade, gender, and socio-economic status, and even MLC. This study suggests that the DASS offers a promising novel index to capture syntactic performance in emerging academic writers, and effectively captures those aspects of syntax that are most associated with writing quality.

**Introduction**

As students enter upper elementary and middle school grades, school contents and tasks become increasingly challenging and require students to produce written language that differs in systematic ways from their more colloquial communications with peers (Cummins, 1979; Schleppegrell, 2001). Students need to express complex thoughts in writing, and their mastery of a repertoire of linguistic resources to convey sophisticated meanings to a distant audience supports such communication at school. National assessments in the U.S. have revealed documented that only 30% of tested fourth graders and eighth graders performed at or above the proficiency level for argumentative writing (NAEP, 2011). Against this backdrop, it is imperative for educational researchers
to analyze the language skills of adolescents as a potential area in need of instructional support for effective written communication.

The current educational standards are not sufficiently informative in describing the language skills. Syntax is one of the main language skill domains that constitute writing, but is described only vaguely in the standards. Syntax refers to “the systematic ways in which discrete units (e.g., words) can be combined to create meaningful utterances (e.g., sentences)” (Fromkin et al., 2013, as cited by Kyle, 2016). The Common Core State Standards (CCSS, 2010) described a general expectation on upper elementary and middle school grades as “Each year in their writing, students should demonstrate increasing sophistication in all aspects of language use, from vocabulary and syntax to the development and organization of ideas…” Besides an emphasis on recognizing and correcting syntactic errors, the descriptions on “increasing syntactic sophistication” by grade are: fifth graders are expected to “link opinion and reasons using words, phrases, and clauses” in writing; sixth to eighth graders are expected to “use words, phrases, and clauses to clarify the relationships [among argumentative moves]” in writing. Although the standards have identified basic units for syntactic analysis (e.g., words, phrases, and clauses) and their functions (e.g., to link ideas and display their interrelations), additional details on which kinds of syntactic structures form part of a continuum of increase sophistication throughout the middle school grades still remain unspecified. Greater detail on the repertoire of syntactic structures that show developmental trends and positively associated with
writing quality can shed new light on the design of innovative instructional approaches.

Syntactic skills in holistic writing assessment rubrics are also described in broad terms. The National Writing Assessment Framework for fourth and eighth graders’ persuasive writing describes sentence structure in a high-quality essay as “well controlled and varied” and in a low-quality essay as “sometimes correct but little variety” (NAEP, 2011). Given that scoring rubrics only include a handful of sample essays to illustrate these quality differences, analyses derived from samples of students representative of the US school population are needed to examine the vast variety of skills in students’ syntactic production and to examine the relation between an essay’s syntactic diversity and its overall writing quality. Most current writing assessments are based on holistic rubrics and, thus, provide limited guidance to teaching and learning in the area of syntactic resources for writing.

In short, in order to facilitate improvement on adolescents’ academic writing, especially argumentative writing, beyond the broad educational standards and assessment rubrics, a more fine-grained analysis is needed to offer insights on syntax, a language domain that might benefit from instructional attention. Specifically, in this study, I developed a new index named Diversity of Advanced Syntactic Structures (DASS) to operationalize the diversity of advanced syntactic structures produced by adolescents in their argumentative writing. In the following sections, I first review prior research that has measured syntactic complexity in
adolescent writers’ text production or described syntactic characteristics of academic writing. Next, I propose my new syntactic index which is guided by an integration of the types of syntactic structures identified from prior research as relevant for academic writing and likely to develop throughout the adolescent years. Then, I examine the validity of this new index by testing its between-grade differences as well as its prediction of argumentative essays’ overall writing quality and of students’ receptive academic language skills. Finally, I discuss the research and practice implications of my new syntactic index and suggested directions for future research.

**The Role of Syntax in Adolescent Writing: Prior Research**

Research on syntactic development in productive language has largely been focused on oral language during early childhood (e.g. Brown & Fraser, 1963; Brown, 1973; Dromi & Berman, 1986; Huttenlocher et al., 2002; Tomasello, 2000; Tomasello & Brooks, 1999). How young children develop from producing simple syntactic forms, such as one-word utterances, to more complex syntactic forms, such as embedded clauses, is well documented. The length of children’s utterances (Brown, 1973; Klee et al., 1989) and diversity of structures produced (Berninger et al., 2011; Sagae et al., 2005; Scarborough, 1990) indicate developmental and individual differences in syntactic skills as a part of children’s overall oral language development.
In contrast, research on syntax beyond early childhood is comparatively scarce, and even more limited when one looks for research on productive syntax in writing. A search for studies of K-12 students’ English writing among five databases yields a total of only 36 published empirical studies that explicitly measure students’ syntactic performance in writing in the last thirty years (Jagaiah et al., 2020). However, syntax is an essential skill that students need to master to navigate through school literacy contexts, especially in upper elementary and middle school grades when students are required to make the transition into forms of written language for academic purposes that are less familiar to the language they use outside of school or in the narratives they have read in elementary school. The following sections reviewed research on how syntactic complexity in adolescents’ writing was operationalized and how syntactic features of school-based texts were described.

**Conventional Measures for Syntactic Complexity in Adolescent Writing**

Previous studies that focus on analyzing syntactic performance in the writing of upper elementary and middle school students predominantly use “omnibus measures” that describe the global syntactic complexity of the text “in a single quantitative variable” (Biber et al., 2020). These omnibus measures focus on calculating the average length of various syntactic units in a text. The most widely adopted indices are Mean Length of T-units (MLT), Mean Length of Clauses (MLC), and Mean Number of Clauses per T-unit (CT) (see summary in
Jagaiah et al., 2020). T-unit, which stands for *Terminable Unit*, is defined as "a main clause plus all subordinate clauses and non-clausal structures attached to it or embedded in it” (Hunt, 1970, p. 4). In other words, a T-unit may be a unit that consists of one independent clause without any attached clauses (e.g., *We will not go out*); one main clause with a subordinate clause (e.g., *We will not go out because it is raining*), or a complex sentence with more than one embedded clause (e.g., *The installation of the new surveillance cameras has caused individuals who engage in small group smoking outside the office building when the weather is good considerable distress*). In operation, the length in MLT and MLC is typically measured in words. To calculate MLT, MLC, or CT, a text is segmented into T-units and/or clauses. MLT is calculated as the total number of words divided by the total number of T-units in text; MLC is calculated as the total number of words divided by the total number of clauses in text; CT is calculated as the total number of clauses divided by the total number of T-units in text.

**Length-Based Measures Indicating Genre-Specific Syntactic Development**

Although the seminal research by Hunt (1970) found among fourth to 12th graders a consistent pattern that students at higher grades produced greater MLT, MLC, and CT, more recent studies on these length-based omnibus measures have revealed genre-based differences in adolescents’ writing development. First, evidence was found that adolescents’ expository writing seems to be more
syntactically complex than narrative writing. Researchers have found that MLC showed a higher value in expository writing than in narrative writing among high school students; the same trend was found for the mean proportion of relative clauses among all clauses, a measure similar to CT (Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009). Second, researchers have found syntactic complexity in expository writing showed a steeper developmental slope than narrative writing during upper elementary and middle school. For example, comparing fourth graders’ and seventh graders’ writing of the two genres, Berman and Verhoeven (2002) found that MLC in narrative writing was around 5.6 at both grade levels, whereas MLC in expository writing was around 5 at fourth grade and around 7 at seventh grade.

However, the length-based omnibus measures were not always consistent in reflecting developmental trends, even in the same writing genre. For example, Beers and Nagy’s (2011) longitudinal study on persuasive writing found MLC and CT negatively correlated with each other; not surprisingly, MLC was found to be lower at third and fifth grade than at seventh grade, whereas CT did not show between-grade difference. It seems MLC was more sensitive to developmental trends in adolescence.

Further studies on the association between syntactic complexity of essays and the overall essay quality also suggested that MLC seemed to be appropriate for the expository genre specifically. For example, Beers and Nagy (2009) found seventh and eighth graders’ persuasive writing quality was positively predicted by
MLC, negatively predicted by MLT, and not at all by CT; for the narrative genre, on the other hand, writing quality was positively predicted by MLT but not predicted by MLC or CT. As explained by the authors, narrative writing is a genre that is more similar to speech in expressing sequential events and concatenating ideas, resulting in longer utterances (i.e., larger MLT) that consist of collocated but not embedded phrases or clauses; in contrast, expository writing entails a higher level of information packing.

**Textual Linguistics Research on Syntactic Features of School-based Texts**

Beyond calculating the length, researchers have provided more detailed descriptions for various syntactic features of academic texts, a challenging genre that developing writers aim to master. Primarily, compared with oral language utterances with short conversational turns, the written language features longer utterances with more dense information (Snow & Uccelli, 2009; Uccelli 2019). As reflected in syntactic features, the higher information density is typically achieved by organizing and linking language structures within a clause or between clauses. In addition, syntactic features of the written language may not correspond to longer utterances but may serve the purpose of heightening communicative effectiveness of complex communications, such as foregrounding the information that the writer intends to highlight. Besides studies on adolescent writers, recent findings from learners of English as a second/foreign language (ESL/EFL)
corroborate the identification of certain syntactic features as indicators of syntactic performance in writing.

**Syntactic Features of Within-Clause Information Packing**

Textual linguistics research has identified several syntactic features of school-based texts that display the specific approaches used to achieve phrase-level information packing within a clause, which may correspond to higher MLC. Ravid and Berman (2010) identified noun phrase structure as a key area of syntactic development in upper elementary and middle school grades. Based on analyses of English language grammar, Biber et al. (2020) have provided a sociolinguistic descriptive framework that differentiates syntactic features in academic writing from those used in conversation; specifically, academic writing features four main approaches to elaborating noun phrases: attributive adjectives (e.g., *conversational practices*), nouns as noun modifiers (e.g., *aviation security committee*), preposition phrases as noun modifiers (e.g., *the scores for male and female students*), and appositive noun phrases (e.g., *Two Stuart monarchs, Charles I and Charles II*). These approaches to modifying and elaborating noun phrases are all ways of packing more information within a clause.

Studies on English-as-a-second-or-foreign language (ESL/EFL) learners have provided evidence that these syntactic structures predict writing proficiency. For example, Crossley and McNamara (2014) found that college-level EFL learners used larger number of modifiers per noun phrase after a semester-long
academic writing course. Kyle (2016) found that a composite score on noun phrase elaboration was positively correlated with higher argumentative writing quality among adolescent and adult EFL learners. As these specific structures for noun phrase elaboration have rarely been examined for adolescent English monolingual students, another group of developing academic writers, it is worth exploring whether any of the structures are produced in adolescent writing.

**Syntactic Features of Between-Clause Information Packing**

Linguistic analyses of adolescent writing have identified specific syntactic structures for between-clause information packing, i.e. embedding one or more clauses under another clause. As Nippold (2006) has summarized, embedded clauses may be relative (e.g., *This flower which only grows in the tropics is very rare*), adverbial (e.g., *The flower blooms when the temperature is above 95 degrees*), or nominal (e.g., *Whoever discovered the flower was a great scientist*) (Nippold, 2004). Embedded clauses were found to be distinctive and prevalent in complex written language in secondary school academic texts (Christie & Derewianka, 2008; Berman & Ravid, 2009; Schleppegrell 2001). For example, the sentence *This flower which only grows in the tropics is very rare* includes an embedded *which*-clause; without the embedded relative clause, the same meaning would be expressed as two main clauses, that is, two separate sentences as *This flower is very rare. It only grows in the tropics*. The embedded clause version is more likely to occur in a science text for adolescents, whereas the two-sentence
version is more likely to occur in conversation or texts written for younger students.

Research on students from third to ninth grade has found that older students produced more adverbial clauses in sentence completion tasks (McClure & Steffensen, 1985 as cited in Nippold, 2006). ESL/EFL writing research has provided evidence of embedded clauses indicating writing proficiency. For example, De Clercq and Housen (2017) found in French-speaking secondary school EFL learners that higher proportions of adverbial and relative clauses indicate higher English proficiency levels. However, research on embedded clause production in English monolingual adolescent writing has revealed some confusing findings. Berman and Nir-Sagiv (2007) reported a non-linear developmental pattern in expository writing; although the relative clauses were rarer at fourth grade than at 11th grade, the percentages at seventh grade were even lower than at fourth grade. Given the conceptual importance of utilizing various embedded clauses in writing but the irregular pattern revealed by the length-based measure, it is possible that the focus of adolescents’ development on between-clause information packing is not on generating embedded clauses beyond single independent clauses, but rather on expanding the diversity of embedded clauses. The possibility of establishing a new index reflecting the types rather than the frequency should be explored.

**Other Syntactic Features for Effective Communication on Complex Topics**
Textual linguistics research has also identified other syntactic features that are typically acquired later in development and are characteristic of school-based texts and which are not based on the length of the unit. First, researchers have identified the use of passive voice, a low frequency and late-developing linguistic structure (Berman & Ravid, 2009; Nippold, 2006). Passive voice has the advantage of highlighting the experiencer of an action, rather than the performer of an action, by positioning it as the subject of a sentence. For example, the passive voice sentence *Kennedy was killed* was able to highlight *Kennedy* by putting it in the sentence subject position, in comparison to the active voice sentence *Someone killed Kennedy* which highlights the assassin. In this case, the passive voice and active voice sentences have the same length. Research on English expository writing found that 15-to-16-year old students used a larger number of passives than 12-to-13-year old students, who in turn use fewer passive structures than 9-to-10-year old students (Jisa, et al., 2002).

A second distinct feature of the language of school texts is the use of nominal clauses as sentence subjects. As explained by Schleppegrell (2001), the majority of sentence subjects in conversation are pronouns such as *I, You, She* and *He*. In contrast, sentence subjects in academic texts tend to be predominantly nouns (e.g., *Water*), noun phrases (e.g., *Sedimentary rocks*), or nominal clauses (e.g., *The formation of sedimentary rocks; Analyzing the formation of sedimentary rocks*). By using a clause as the sentence subject, the writer is able to direct the reader’s attention to the content of the clausal subject. For example, the sentence
Having technology in our classrooms is important implies the focus of discussion is on having or not having technology.

Gaps in Research on Measuring and Describing Syntax in Adolescents’ Expository Writing

Prior research relevant to adolescents’ academic writing can be synthesized in a conceptual framework as shown in Figure 2.1. On the one hand, textual linguistics research has identified syntactic features of school-based texts that serve the written communication purposes of information packing between or within clauses and foregrounding the writer’s intention. These studies typically use individual syntactic structures as predictors of proficiency level or writing quality, with the aim of identifying the strongest predictors. However, academic writing can be simultaneously characterized as displaying multiple types of syntactic structures which could potentially be adopted by the writers who have acquired them. It is evidence of access to a variety of syntactic structures, rather than any single one of them, that marks the skillful writer.

On the other hand, the body of adolescent writing research on the widely adopted length-based omnibus measures identified reliable genre-specific developmental trends and individual differences in syntactic complexity of writing, with accumulating evidence that MLC is the most promising measure for characterizing adolescents’ expository writing. Nonetheless, this line of research has also revealed a few gaps in understanding expository writing development in
adolescence. The simple index of MLC is minimally informative because many different syntactic structures might display the same number of words per clause. Beyond just a number, a menu of the types of syntactic structures that support argumentative writing in adolescents can be promising to design targeted syntactic scaffolds. Furthermore, the inconsistent findings from MLC and CT have not been fully explained. Conceptually, both MLC and CT can quantify information packing; the difference is the former measures within-clause and the latter between-clause length (Beers & Nagy, 2009; 2011). One possible explanation for the developmental trends found for MLC but not for CT is that students in mid-adolescence are still developing the ability to generate sophisticated phrases within clauses, but have already achieved the ability to produce clauses within a T-unit. Given the length-based measures are not able to provide information about the variety of embeddings at clause or phrase levels, the plausibility of these possible explanations remains unclear. The length-based measures may have obscured underlying variability in adolescents’ syntactic complexity in writing.

There is also a lack of research examining the relation between syntactic complexity of student essays and the essays’ overall writing quality as well as examining between-grade differences in one mid-adolescent group with diverse sociodemographic backgrounds. Many extant studies (e.g., Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009; Berman & Verhoeven, 2002) focused on describing and comparing average performance at given grade levels. Although the general developmental trends have provided valuable information on the students per
grade, the individual differences within each grade level also need to be revealed and examined. Charting the more nuanced variability to be described within and between grades would be helpful for examining the relation between an essay’s syntactic complexity and the essay’s overall writing quality. In addition, most of these studies were based on small and relatively homogeneous groups of students, with sample sizes around twenty at each grade level. It is unclear whether the patterns found would also apply to students across different sociodemographic backgrounds.

In short, the gaps in prior research suggest the potential value of generating a new syntactic index that can a) represent the variety of target syntactic structures that adolescents aim to master in their academic language; and b) reflect the degree to which adolescent writers produce these structures by quantifying their occurrence in written texts. Such an index, if valid, should also be sensitive to developmental trends and reflect the variability in the overall writing quality as well as in the students’ receptive academic language skills beyond the conventionally adopted length-based measure for expository writing. Therefore, the research questions for the current study are:

RQ 1. Can a novel index based on the diversity of adolescents’ syntactic production (Diversity of Advanced Syntactic Structures, or DASS) identify individual variability in argumentative writing produced by upper elementary and middle school students’?
RQ 2. Does the novel index *Diversity of Advanced Syntactic Structures (DASS)* capture developmental differences in students’ syntactic performance in argumentative writing from upper elementary to middle school grades overall?

RQ 3. Are students’ syntactic performance in argumentative writing scored by the novel index *Diversity of Advanced Syntactic Structures (DASS)* associated with: a) students’ argumentative essays’ holistic quality overall, or b) students’ receptive academic language skills, even when controlling for Mean Length of Clauses (MLC)?

For RQ 1, I hypothesized that adolescent students’ syntactic performance in argumentative writing can be conceptualized as the variety of advanced syntactic structures produced, and it can be operationalized as a novel index *Diversity of Advanced Syntactic Structures (DASS)*. For RQ 2, I hypothesized that the DASS scores can reflect developmental trends among students, with higher grade students in general receiving higher scores. For RQ 3, I hypothesized that the DASS scores of students’ argumentative essays would be positively and significantly associated with these essays’ holistic writing quality and receptive academic language skills respectively, even when controlling for MLC.
Methods

Participants

The full sample of the study included 512 fifth-to-eighth graders from Title 1 urban public schools in the Northeastern and Mid-Atlantic regions of the United States. Participating students were part of the control group in a large-scale literacy intervention. Since the current study aims to investigate general developmental patterns and individual differences, rather than a treatment effect, the treatment group was not included in the current study. Participants’ socio-demographic backgrounds are shown in Table 2.1. About half of the participants were female; about two-thirds of the participants were eligible for free/reduced-price lunch. The vast majority (97%) were native English speakers. The two largest race/ethnicity sub-groups in the sample were White (41%) and Black (41%), followed by Latinx (13%). The sample consisted of 20% fifth graders, 30% sixth graders, 30% seventh graders, and 20% eighth graders.

Procedures

I focused on participants’ responses to one writing prompt administered at the end of spring 2014. The writing prompt was: Should we allow iPads in our classrooms? The writing task was developed by the IES-funded Catalyzing Comprehension through Discussion and Debate (CCDD) team (Jones et al., 2019; LaRusso et al., 2016; Lawrence et al., 2015; Snow et al., 2009) to assess upper elementary and middle school students’ writing. Participants were given 20 to 25
minutes to write an argumentative essay and were provided with the following scenario: their school principal had decided to stop the school’s policy of providing iPads to students, thus participants were asked to take a position and to write an argumentative essay to be published by their school newspaper. Participants read a brief description of why iPads had been popular and why they were subsequently prohibited. In their essay, students were asked to give reasons to support their position, to try to convince people, to explain the impact on others, and to discuss potential alternative resolutions to the problem. Participants wrote the essays in the paper-and-pencil format (see full prompt in Appendix 2.1).

**Data Preparation**

Prior to analysis, all the hand-written essays were transcribed using the *Code for the Human Analysis of Transcripts (CHAT)* conventions (MacWhinney, 2000). All spelling errors were corrected in the transcribed essay data in order to assure that human scorers of writing quality were not negatively biased by non-relevant misspellings or other orthographic features. Original files with misspellings were also preserved. Then, the spelling error free texts were saved as .txt files in order to be processed in the automated language analysis software.

**Measures**

**Writing Quality**

Students’ responses were scored using a holistic rubric developed by a team
of language and writing researchers and informed by the NAEP (2011) Writing Framework. The rubric includes four dimensions: (1) Position: the number of sides that the essay considers; (2) Organization: the extent to which the essay is coherently structured. (3) Development of Ideas: the degree of depth, complexity, elaboration, and connectedness of ideas provided; (4) Clarity: the extent to which the essay conveys information in a precise and unambiguous manner. Each dimension was scored on a 4-point scale, from which the overall writing quality score was generated on a 6-point scale. Essays with higher scores on multiple dimensions were rated with higher overall writing quality score. The essays were scored by a team of three research assistants, all graduate students specializing in education-related areas with prior experience as classroom teachers and blind to the study questions. In the group training for scoring team, a training set of essays was scored by all three scorers guided by the holistic writing rubric, which included anchor essays at each level. After this training, a high inter-rater reliability was achieved on the basis of 20% of the sample, with Kendall's Coefficient of Concordance for Ordinal Response higher than .92 on all dimension scores (i.e., Position: .92; Development of Ideas: .99; Organization: .98; Clarity: .99) and .99 on the overall writing quality.

Receptive Academic Language | Core Academic Language Skills (CALS)

Instrument

Participants’ receptive academic language skills were measured using the
**Core Academic Language Skills (CALS) Instrument**, a researcher-developed, paper-and-pencil assessment for students in grades 4 to 8 (Barr et al., 2019; Uccelli et al., 2015). The **CALS** Instrument measures seven domains of academic language skills: unpacking dense information, connecting ideas logically, tracking participants, interpreting writers’ viewpoints, understanding metalinguistic vocabulary, understanding text organization, and recognizing academic register. It includes two vertically equated forms: *Form 1* for fourth, fifth, and sixth graders ($\alpha = .90$, total items = 49) and *Form 2* for seventh and eighth graders ($\alpha = .86$, total items = 46). Scores were generated using Rasch item response theory analysis.

**Length-based Measure for Syntactic Complexity**

For each essay, the mean number of words per clause is calculated as of Mean Length of Clauses (MLC). For this measure, each essay was processed in the *Syntactic Complexity Analyzer* module (Lu, 2010) within the *Tool for the Automatic Assessment of Syntactic Sophistication and Complexity (TAASSC)* program (Kyle, 2016). Prior work has shown that MLC averages around 7.2 ($SD = 1.2$) in persuasive essays of seventh and eighth grade English speaking students (Beers & Nagy, 2009), and ranges from 8.8 to 9.6 in argumentative writing for college students learning English as a second/foreign language (Lu, 2010).
Development of A Novel Index: Diversity of Advanced Syntactic Structures (DASS)

Framework of Identifying Syntactic Structures.

The list of advanced syntactic structures used in my analysis was selected from Kyle’s (2016) clausal and phrasal complexity indices, which are based on previous studies using a dependency parsing framework (De Marneffe et al., 2006; Chen & Manning, 2014). Dependency parsing is a labelling system that describes the relationships among words, phrases, or other linguistic elements in a sentence. The labelled relationships in a sentence are mutually exclusive, enabling simultaneous identification of a variety of syntactic structures at between-clause or within-clause levels. Unlike constituency parsing which represents linguistic elements nesting within each other in a hierarchy, dependency parsing typically uses the finite verb of the independent clause as the structural center, and linearly labels other elements in the sentence according to their direct or indirect relationship to the center (Caroll et al., 1999; King et al., 2003, as cited in De Marneffe et al., 2006). For example, in the sentence The moon rose as night fell, the word rose is labelled as the center; the word the is labelled as determiner of the word moon, which is in turn labelled as the nominal subject of rose; the clause as night fell is labelled as the adverbial clause of the word rose. Using the finite verb of the independent clause as the center, Kyle (2016) identified 29 structures that are directly linked to the center and 10 structures that are indirectly linked to the center, resulting in 39 syntactic structures according to dependency parsing. From
the 39 structures in Kyle (2016)’s framework, I selected seven target structures for the current study’s analysis.

**Identifying Advanced Syntactic Structures for Adolescent Writers.**

The seven target syntactic structures for adolescent academic writers, as shown in Appendix 2.2, were selected based on prior research situated in the conceptual framework. First, I selected two target structures that serve the purpose of between-clause information packing using embedded clauses (Christie & Derewianka, 2008), including: 1) clausal complement (e.g., *I think that the principal should allow iPads*) and 2) clausal prepositional complement (e.g., *The punishment should depend on how serious their mistake is*). As the structure names suggested, the difference between the two structures was that the latter began with a preposition.

Second, embedded clauses may begin with subordinating conjunctions (Nippold, 2006), a group of adverbs serving as intrasentential cohesion devices such as *after, because, if, when*. The structure signaled by subordinating conjunctions was labelled as adverbial clause in the dependent parsing framework. Embedded clauses may also begin with pronouns, such as *who, which, whose* that lead a relative clause. Therefore, the third and fourth target structures I selected were: 3) adverbial clause (e.g., *We should allow iPads because they help us learn*), and 4) relative clause as modifier (e.g., *We need to carry heavy textbooks everywhere which is a pain*).
Third, the packing of information may also occur in the independent clause. The subjects of a sentence in conversations were typically single nouns or pronouns, whereas subjects of a sentence in school textbooks tended to be longer as a description of a scenario (Schleppegrell, 2001). Therefore, the fifth target structure I selected was to identify the lexicalized sentence subjects: 5) clausal subject (e.g., *Having iPads in the classroom can help us learn.*)

In addition to identifying different clause linking patterns, I selected the nouns as noun modifier for within-clause information packing in academic texts. Biber et al. (2020) has identified that nouns as noun modifiers in a phrase are common in academic writing. Although Biber et al. (2020) have also identified three other structures for noun phrase elaboration, as reviewed in the previous section, noun as noun modifier stood out as it was also identified in analysis on school-based texts for adolescents (Schleppegrell, 2001). Therefore, I included 6) noun as modifier (e.g., *We can ask the whole school community*) as another target structure for the current study.

I also chose to include passive voice as it has generally exhibited a low frequency in adolescent writing (Jisa et al., 2002; Nippold, 2007). In Kyle (2016)’s dependent parsing framework, different types or parts of a passive voice structure were labelled separately, as *agent* in the passive structure, *passive auxiliary verb*, *passive clausal subject*, or *passive nominal subject*. Since the aim of the current study is not to identify the nuances within the structure, I did not differentiate
these structures; rather, I relabeled all such structures unitarily as the final target structure: 7) passive voice (e.g., *IPads should not be allowed at my school*).

In short, through the processes above, I identified a total of seven advanced syntactic structures that could potentially represent the characteristics in adolescents’ written language. The structures include: 1) clausal complement, 2) clausal prepositional complement, 3) adverbial clause, 4) relative clause as modifier, 5) clausal subject, 6) noun as modifiers, and 7) passive voice.

**Constructing the Scores for the Diversity of Advanced Syntactic Structures (DASS).**

The essays in the sample were imported as .txt files into the *Tool for the Automatic Assessment of Syntactic Sophistication and Complexity (TAASSC)* program (Kyle, 2016) for automated analysis. Sentences with grammatical errors were not included in analysis. The *TAASSC* program’s default setting is to calculate the mean frequency of a structure per clause or per phrase for an essay. Because the aim of the current study is to capture the diversity of syntactic structures rather than the quantity of each structure, I transformed the mean frequency to a binary variable of incidence: if any one of the seven structures was produced in an essay, the structure was coded as 1 (i.e. present); if a structure was not produced in an essay, it was coded as 0 (i.e. absent). After that, I calculated the sum of 1s within each essay as the score of Diversity of Advanced Syntactic
Structures (DASS). Therefore, the possible DASS score for an essay ranges from 0 to 7.

Results

DASS as A Novel Index of Syntactic Performance in Writing

Descriptive statistics show that all seven types of advanced syntactic structures (henceforth structures) were present in the sample, but with varying incidence rates. In other words, some structures are produced by more students than other structures. As shown in Table 2, more than 90% of the students produced clausal complements in their essays. In contrast, only 26%, produced clausal subjects, and a mere 1% of the students produced clausal prepositional complement. The other four structures were produced by 60-80% of the students in the sample.

There was considerable variation in DASS scores. No individual student produced all seven advanced syntactic structures. As shown in Figure 2.1, 12% of the students in the sample scored 6 on DASS (i.e. produced six types of the structures); 30% of the students scored 5; 31% scored 4; 17% scored 3. The last 10% of the student in the sample scored below 3 on DASS, including three students (0.5% of the sample) who did not produce any of the seven structures. Students in the sample have a mean score of 4.12 on DASS with a standard deviation of 1.27. Sample essays with low (score of 2, in the 10th percentile),
medium (score of 4, in the 50th percentile), and high (score of 6, in the 90th percentile) DASS scores are presented in Appendix 2.3.

The descriptive statistics for DASS, MLC, writing quality, and scores on students’ receptive academic language skills for each grade are summarized in Table 2.3. Pearson correlations among DASS, MLC, writing quality, and receptive academic language skills are reported in Table 2.4. DASS showed a moderately strong, positive, and statistically significant correlation with writing quality ($r = .52$) and with receptive academic language skills ($r = .35$).

**Between-Grade Differences in DASS**

I fit a set of multiple regression models to examine the developmental trends in DASS. In the modeling process, I used the grade levels as categorical variables, with fifth grade as the reference group, to examine if there was a statistically significant between-grade difference in DASS, after controlling for students’ sociodemographic background (i.e., students’ gender, socio-economic status indicated by the free/reduced lunch status, and English learner designation). As shown in Table 2.5, students’ sociodemographic background variables were sequentially entered in the series of models. After dropping the non-significant control variables, the final model (Model 3) included grade levels as the predictor and students’ gender and socioeconomic status as control variables. Results showed that after controlling for students’ gender and socioeconomic status, on average fifth and sixth grade essays were not statistically significantly different in
DASS, but seventh grade DASS scores were statistically significantly higher than those in fifth grade ($\beta = .32, SE = .15, p < .05$) and so were eighth grade essays ($\beta = .46, SE = .19, p < .05$). In other words, on average, a seventh grade essay contains .32 more types of advanced syntactic structure and an eighth grade essay contains .46 more advanced syntactic structure than a fifth grade essay, controlling for the writer’s gender and socioeconomic status. Post-hoc pairwise comparison between sixth, seventh, and eighth grade essays found that on average there was no statistically significant difference among the three grades.

**Writing Quality of Argumentative Essays Predicted by DASS**

I fit a set of multiple regression models to examine whether students’ DASS scores predicts their scores on their essays’ writing quality. In the modeling process, I used DASS as the independent variable, controlling for MLC, the categorical variables of students’ grade levels (i.e., using fifth grade as the reference group) and sociodemographic background (i.e., students’ gender, socioeconomic status indicated by the free/reduced lunch status, and English learner designation). As shown in Table 2.6, the control variables were sequentially entered in the series of models. After dropping the non-significant control variables, the final model (Model 7) showed that after controlling for students’ grade levels, gender, and socioeconomic status, students’ DASS significantly and positively predicts their writing quality ($\beta = .40, SE = .04, p < .001$); MLC was not statistically significant in any of the models. On average, students who produced
one additional type of advanced syntactic structure are predicted to score .40 point higher in their writing quality score, controlling for their gender and socioeconomic status. The prediction to the writing quality scores is substantial, as the .40 point difference corresponds to about a third of the writing quality scores’ standardized deviation.

**Receptive Academic Language Skills (CALS) Predicted by DASS**

I fit another set of multiple regression models to examine whether students’ DASS scores predicts their scores on receptive academic language skills. In the modeling process, I used DASS as the independent variable, controlling for MLC, the categorical variables of students’ grade levels (i.e., using fifth grade as the reference group) and sociodemographic background (i.e., students’ gender, socioeconomic status indicated by the free/reduced lunch status, and English learner designation). As shown in Table 2.7, the control variables were sequentially entered in the series of models. After dropping the non-significant control variables, the final model (Model 7) showed that after controlling for students’ grade levels and socioeconomic status, on average students’ DASS significantly and positively predicted their essays’ writing quality ($\beta = .24, SE = .04, p < .001$). MLC was not statistically significant in any of the models. On average, students who produced one additional type of advanced syntactic structure tended to score .25 point higher in their receptive academic language skills, controlling for their grade and socioeconomic status. The prediction to the receptive academic
language skills is moderate, as the .25 point difference corresponds to about 20% the receptive academic language skills scores’ standardized deviation.

**Discussion**

Motivated by identifying indicators of syntactic complexity exhibited in adolescents’ academic writing and by assessing adolescents’ syntactic performances beyond the conventional clause-length calculation approach, in this study I developed a new index, Diversity of Advanced Syntactic Structures (DASS). The DASS score robustly predicted the essays’ overall writing quality and students’ receptive academic language, even after controlling for MLC, the widely used length-based syntactic measure shown to be sensitive to differences in expository writing. Results also provide robust evidence of this index sensitivity to between-grade variability. DASS scores of fifth graders showed statistically significantly lower scores than seventh graders and eighth graders, after controlling for students’ gender and socioeconomic status.

**Identifying Specific Syntactic Structures to Generate an Overall Score**

The current study contributes to the field of adolescent writing research by proposing the newly developed DASS index which presents dual benefits. On one hand, it represents a variety of target syntactic structures characteristic of academic texts, as identified in textual linguistics. By identifying the specific structures that writers produce as they expand or link words or phrases within a
clause or between clauses, the DASS score becomes more interpretable and transparent. On the other hand, DASS has the advantage of conventional length-based omnibus measures in that it provides a quantifiable value for the full text.

Furthermore, the association of DASS with writing quality adds to the scarce research examining the role of productive syntax in adolescents’ writing. In addition, although DASS is a productive syntactic measure, it showed a significant and positive association with receptive academic language skills, which represents a construct more distant from syntactic production than writing. This latter association supports the view of a common underlying academic language proficiency on which students rely to comprehend as well as to produce language.

**DASS Accounting for Variability Beyond MLC**

The findings in the current study on students’ between-grade differences corroborate and elaborate findings from previous research. Previous research has set the foundation of describing the developmental trends by using MLC to compare fourth grade or fifth grade with seventh grade students’ argumentative writing (Beers & Nagy, 2011; Berman & Nir-Sagiv, 2007; Berman & Ravid, 2009; Berman & Verhoeven, 2002). The current study, using a new type-based index, detected the same developmental trends showing fifth grade essays significantly lower than seventh grade. Furthermore, it contributes to this body of research by analyzing all grade levels spanning upper elementary and middle school within one study. It added more details to the developmental trend.
description by analyzing essays from sixth grade, an intermediate grade level within this age range, and extending the description to eighth grade. The consistent trends yielded from a new syntactic measure other than MLC corroborates the view that productive syntactic skills continue to develop in adolescence, especially as students are learning the argumentative genre.

The current study found MLC was not significantly associated with writing quality. The findings was different from previous research which found MLC as a significant predictor (Beers & Nagy, 2007). One possible explanation for the different results might be the different school contexts and participants’ socioeconomic background in the two studies, which may in turn be associated with the variability among the participants. The participants in Beers and Nagy (2007)’s study were students from suburban middle schools, whereas participants in the current study were from urban schools with more diverse socioeconomic background. More research is needed to understand the generalizability of these results across different populations.

Implications

The study has several implications for research. For adolescent writing research, it suggests that students’ productive syntactic skills may play a more important role in writing than shown in previous research. The development of DASS can help shed light on the language domain of syntax that has been broadly and vaguely described in educational standards and assessment rubrics as "well-
controlled sentence structures” (NAEP, 2017). The study suggests the plausibility of identifying local syntactic structures and integrating them as a text-wide score. Detailed textual linguistic analyses of academic language are fruitful.

For practice, the study suggests that DASS could be applied as a diagnostic or formative assessment tool to inform instruction. The DASS score, which is the total number of advanced syntactic structure types, is easily interpretable for teachers. Instead of simply suggesting “the more words in a clause, the better”, the study found specific syntactic structures that students are expected to learn as a part of their language for school to both show developmental trends and show an association with writing quality. Thus, teachers and curriculum developers can design specific materials and practices to scaffold the mastery of these structures as resources for students’ to express their own meanings through writing. The types of advanced syntactic structures can be integrated in curricula, lesson plans, or a reference of providing feedback to improve students’ writing.

**Limitations**

The study has several limitations. First, the study gave all selected syntactic structures in DASS equal weight. However, some structures are more prevalent than others in the sample of the current study. It is possible that some structures should receive heavier weight. Second, DASS provides an initial exploratory set of syntactic structures relevant for adolescent argumentative writing, rather than prescribing a definite or complete set of structures. Additional syntactic forms
could potentially be identified as advanced structures for this age group. Third, the study only tested for students’ productive syntax without testing their receptive syntactic knowledge, which is the basis of production. Furthermore, the writing task elicited only one response per student, and all were on the same prompt; thus, the types of syntactic structures produced reflect only a writing performance, not the syntactic profile of a writer. In addition, the study used a cross-sectional, rather than longitudinal, sample to analyze between-grade differences. Causal inferences between DASS and writing quality cannot be made, as the current study only tested the relation as association. It is unknown whether improvement on DASS would lead to higher scores on writing quality. Finally, it is unknown to what extent these findings can be generalized outside this particular sample of students.

**Future Research**

To address the abovementioned limitations, future research can further analyze whether some advanced syntactic structures are more likely to be produced by adolescent students than other structures, and can expand the search to identify other syntactic structures that may be sensitive in capturing variability for this age group. A larger variety of writing prompts and topics could be used to elicit responses from students, preferably at multiple time points. Future studies should also analyze longitudinal samples in order to have more accurate description of the developmental trends. Intervention studies on advanced
syntactic structures with randomized control design could be conducted to test for
the potential causal relations between DASS and writing quality.

In future research, receptive as well as productive syntactic skills could
both be included in the investigations to explore learners’ academic language
proficiency. For example, research could be conducted to examine which syntactic
structures have higher frequency in students’ literacy environment, such as in their
reading materials, textbooks, or classroom discussions, or which structures have
received more instructional time than others. The results from such analyses could
be compared with the syntactic structures in students’ oral or written production,
from which more specific inferences might be drawn on the connections between
structures that students are exposed to and the structures they produce.

Last but not least, it would be helpful to establish a corpus of adolescent
writing for more detailed analyses and exploration on the linguistic indicators of
the texts. The corpus could include a large variety of scenario-based prompts,
ranging from more spontaneous writing (e.g., emailing a professor) to more
structured writing (e.g., writing for high-stake standardized assessments). Not only
the text products but the writing processes, such as the drafting, revising, editing,
or oral discussions regarding the text, could be recorded. Products from learners
with different English learning history or different first language background
could also be analyzed.
Conclusions

In the current study, I developed a novel index, Diversity of Advanced Syntactic Structures (DASS), to indicate adolescents’ syntactic performance in argumentative writing. I found that DASS is a robust predictor of writing quality as well as of receptive academic language skills, even after controlling Mean Length of Clauses (MLC), a widely adopted syntactic complexity measure. DASS scores at fifth grade are lower than seventh and eighth grade. The study builds on and expands prior research that characterizes the written production of developing academic writers by providing an operationalizable index to measure students’ syntactic performance. This index complements the information provided by the predominantly applied omnibus length-based measures.
### Table 2.1

**Participants’ Socio-Demographic Information (N = 512)**

<table>
<thead>
<tr>
<th>Socio-demographic Background</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>261</td>
<td>51%</td>
</tr>
<tr>
<td>Male</td>
<td>251</td>
<td>49%</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free/reduced lunch Eligible</td>
<td>345</td>
<td>67%</td>
</tr>
<tr>
<td>Free/reduced lunch non-eligible</td>
<td>167</td>
<td>33%</td>
</tr>
<tr>
<td><strong>Language Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Language Learner</td>
<td>14</td>
<td>3%</td>
</tr>
<tr>
<td>Non-English Language Learner</td>
<td>498</td>
<td>97%</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>208</td>
<td>41%</td>
</tr>
<tr>
<td>Black</td>
<td>209</td>
<td>41%</td>
</tr>
<tr>
<td>Asian</td>
<td>8</td>
<td>1.6%</td>
</tr>
<tr>
<td>Latinx</td>
<td>67</td>
<td>13%</td>
</tr>
<tr>
<td>Native/Pacific</td>
<td>2</td>
<td>0.4%</td>
</tr>
<tr>
<td>Mixed/Other</td>
<td>12</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>95</td>
<td>19%</td>
</tr>
<tr>
<td>6th</td>
<td>150</td>
<td>29%</td>
</tr>
<tr>
<td>7th</td>
<td>182</td>
<td>36%</td>
</tr>
<tr>
<td>8th</td>
<td>85</td>
<td>17%</td>
</tr>
<tr>
<td>Types of Advanced Syntactic Structures</td>
<td>Number of Students who Produced this Type</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>clausal complement</td>
<td>470 (92%)</td>
<td></td>
</tr>
<tr>
<td><em>(I think that the principal should allow iPads.)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adverbial clause</td>
<td>412 (80%)</td>
<td></td>
</tr>
<tr>
<td><em>(We should allow iPads because they help us learn.)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>noun as modifier</td>
<td>387 (76%)</td>
<td></td>
</tr>
<tr>
<td><em>(We can ask the whole school community.)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relative clause as modifier</td>
<td>375 (73%)</td>
<td></td>
</tr>
<tr>
<td><em>(We need to carry heavy textbooks everywhere which is a pain.)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>passive voice</td>
<td>327 (64%)</td>
<td></td>
</tr>
<tr>
<td><em>(IPads should not be allowed at my school.)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clausal subject</td>
<td>132 (26%)</td>
<td></td>
</tr>
<tr>
<td><em>(Having iPads in the classroom can help us learn.)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clausal prepositional complement</td>
<td>5 (1%)</td>
<td></td>
</tr>
<tr>
<td><em>The punishment should depend on how serious their mistake is.</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.3

Summary Statistics of Scores on Diversity of Advanced Syntactic Structures (DASS), Mean Length of Clauses (MLC), Writing Quality, and Receptive Academic Language Skills (CALS) (N=512)

<table>
<thead>
<tr>
<th></th>
<th>Grade</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>DASS</td>
<td>3.77</td>
<td>4.11</td>
<td>4.15</td>
<td>4.45</td>
<td>4.12</td>
</tr>
<tr>
<td></td>
<td>(1.35)</td>
<td>(1.17)</td>
<td>(1.22)</td>
<td>(1.38)</td>
<td>(1.27)</td>
</tr>
<tr>
<td>MLC</td>
<td>7.29</td>
<td>7.96</td>
<td>8.27</td>
<td>8.36</td>
<td>8.01</td>
</tr>
<tr>
<td></td>
<td>(1.51)</td>
<td>(1.55)</td>
<td>(2.37)</td>
<td>(1.48)</td>
<td>(1.90)</td>
</tr>
<tr>
<td>Writing Quality</td>
<td>2.75</td>
<td>3.08</td>
<td>3.24</td>
<td>3.79</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>(.95)</td>
<td>(.96)</td>
<td>(1.17)</td>
<td>(1.42)</td>
<td>(1.17)</td>
</tr>
<tr>
<td>CALS</td>
<td>.56</td>
<td>1.32</td>
<td>1.30</td>
<td>2.51</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>(.93)</td>
<td>(1.30)</td>
<td>(1.21)</td>
<td>(1.26)</td>
<td>(1.29)</td>
</tr>
</tbody>
</table>
Table 2.4

Pearson Correlations among Scores on Diversity of Advanced Syntactic Structures (DASS), Mean Length of Clauses (MLC), Writing Quality, and Receptive Academic Language Skills (CALS) (N=512)

<table>
<thead>
<tr>
<th></th>
<th>DASS</th>
<th>MLC</th>
<th>Writing Quality</th>
<th>CALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASS</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLC</td>
<td>-.00</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Quality</td>
<td>.52***</td>
<td>.07</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CALS</td>
<td>.35***</td>
<td>.12*</td>
<td>.45***</td>
<td>1</td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01, *** p < 0.001
Table 2.5

Diversity of Advanced Syntactic Structures (DASS) Scores Predicted by Grade Levels (N = 512)

<table>
<thead>
<tr>
<th></th>
<th>Model 1 DASS</th>
<th>Model 2 DASS</th>
<th>Model 3 DASS</th>
<th>Model 4 DASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 6</td>
<td>0.345*</td>
<td>0.375*</td>
<td>0.295</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>(2.09)</td>
<td>(2.35)</td>
<td>(1.87)</td>
<td>(1.81)</td>
</tr>
<tr>
<td>Grade 7</td>
<td>0.380*</td>
<td>0.374*</td>
<td>0.315*</td>
<td>0.310*</td>
</tr>
<tr>
<td></td>
<td>(2.38)</td>
<td>(2.43)</td>
<td>(2.07)</td>
<td>(2.03)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>0.679***</td>
<td>0.695***</td>
<td>0.462*</td>
<td>0.453*</td>
</tr>
<tr>
<td></td>
<td>(3.60)</td>
<td>(3.83)</td>
<td>(2.47)</td>
<td>(2.41)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.711***</td>
<td>0.736***</td>
<td>0.736***</td>
<td>0.736***</td>
</tr>
<tr>
<td></td>
<td>(6.62)</td>
<td>(6.95)</td>
<td>(6.95)</td>
<td>(6.95)</td>
</tr>
<tr>
<td>FRL</td>
<td></td>
<td>-0.503***</td>
<td>-0.511***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-4.25)</td>
<td>(-4.26)</td>
<td></td>
</tr>
<tr>
<td>ELL</td>
<td></td>
<td></td>
<td>0.146</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.44)</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>3.768***</td>
<td>3.392***</td>
<td>3.803***</td>
<td>3.809***</td>
</tr>
<tr>
<td></td>
<td>(29.12)</td>
<td>(24.86)</td>
<td>(22.99)</td>
<td>(22.93)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.025</td>
<td>0.104</td>
<td>0.135</td>
<td>0.135</td>
</tr>
</tbody>
</table>

*Note.* Grade 5 set as the reference group

1FRL: Free-reduced lunch status; 2ELL: English Language Learner Status

$t$ statistics in parentheses $^*$ $p < 0.05$, $^** p < 0.01$, $^*** p < 0.001$
<table>
<thead>
<tr>
<th>Grade</th>
<th>Coefficients</th>
<th>Standard Errors</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.218***</td>
<td>0.018</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>1.377**</td>
<td>0.023</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>3</td>
<td>1.544</td>
<td>0.028</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>1.712**</td>
<td>0.032</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>5</td>
<td>1.880</td>
<td>0.036</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>6</td>
<td>2.047**</td>
<td>0.040</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>7</td>
<td>2.214</td>
<td>0.044</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>8</td>
<td>2.381**</td>
<td>0.048</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Note: Grades 5 & 6 are the reference group.

Table 2.6: Input of Advanced Spanish Structure (DASS) Scores Predicting Writing Quality (N = 512)
<table>
<thead>
<tr>
<th>Year</th>
<th>Grade</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0'36</td>
<td>0'36</td>
<td>0'36</td>
<td>0'36</td>
<td>0'36</td>
</tr>
<tr>
<td>0'35</td>
<td>0'35</td>
<td>0'35</td>
<td>0'35</td>
<td>0'35</td>
</tr>
<tr>
<td>0'34</td>
<td>0'34</td>
<td>0'34</td>
<td>0'34</td>
<td>0'34</td>
</tr>
<tr>
<td>0'33</td>
<td>0'33</td>
<td>0'33</td>
<td>0'33</td>
<td>0'33</td>
</tr>
<tr>
<td>0'32</td>
<td>0'32</td>
<td>0'32</td>
<td>0'32</td>
<td>0'32</td>
</tr>
<tr>
<td>0'31</td>
<td>0'31</td>
<td>0'31</td>
<td>0'31</td>
<td>0'31</td>
</tr>
<tr>
<td>0'30</td>
<td>0'30</td>
<td>0'30</td>
<td>0'30</td>
<td>0'30</td>
</tr>
<tr>
<td>0'29</td>
<td>0'29</td>
<td>0'29</td>
<td>0'29</td>
<td>0'29</td>
</tr>
<tr>
<td>0'28</td>
<td>0'28</td>
<td>0'28</td>
<td>0'28</td>
<td>0'28</td>
</tr>
<tr>
<td>0'27</td>
<td>0'27</td>
<td>0'27</td>
<td>0'27</td>
<td>0'27</td>
</tr>
<tr>
<td>0'26</td>
<td>0'26</td>
<td>0'26</td>
<td>0'26</td>
<td>0'26</td>
</tr>
<tr>
<td>0'25</td>
<td>0'25</td>
<td>0'25</td>
<td>0'25</td>
<td>0'25</td>
</tr>
<tr>
<td>0'24</td>
<td>0'24</td>
<td>0'24</td>
<td>0'24</td>
<td>0'24</td>
</tr>
<tr>
<td>0'23</td>
<td>0'23</td>
<td>0'23</td>
<td>0'23</td>
<td>0'23</td>
</tr>
<tr>
<td>0'22</td>
<td>0'22</td>
<td>0'22</td>
<td>0'22</td>
<td>0'22</td>
</tr>
<tr>
<td>0'21</td>
<td>0'21</td>
<td>0'21</td>
<td>0'21</td>
<td>0'21</td>
</tr>
<tr>
<td>0'20</td>
<td>0'20</td>
<td>0'20</td>
<td>0'20</td>
<td>0'20</td>
</tr>
<tr>
<td>0'19</td>
<td>0'19</td>
<td>0'19</td>
<td>0'19</td>
<td>0'19</td>
</tr>
</tbody>
</table>

**Notes:**
- Gender: Male, Female
- Grade: 1st to 8th
- Total: Combined Male and Female scores

Table 2.7

Gender of Advanced Speaking Subskills (DASS) Scores Predicting Academic Language Skills (CALS)
Figures

Figure 2.1

*Conceptual Framework for Syntax in Adolescent Writing*
Figure 2.2

Distribution of Scores on Diversity of Advanced Syntactic Structures (DASS)
Appendices

Appendix 2.1

Argumentative Writing Prompt

Should we allow iPads in our classrooms?

Imagine that your school decided to give iPads to all students, so everyone has the opportunity to use iPads for learning. Most teachers like the iPads because all students can use online dictionaries and search the internet for interesting information. But in the last few weeks some students have used their iPads to post videos with mean comments about other kids online. Now, many parents are worried about this situation. Because of this problem, your Principal has decided that iPads will no longer be allowed in school!

Write an article for the school newspaper that argues for or against allowing iPads at your school. * Make sure to give specific reasons to support your position and to convince the people who read the article to agree with you. * Explain how the Principal's decision can impact you and others. * Discuss other things that the school community could do to solve the iPad problem.
Appendix 2.2

Types of Advanced Syntactic Structures (adapted from Kyle, 2016)

<table>
<thead>
<tr>
<th>Structure</th>
<th>Index Name and Examples in TAALES Program (Kyle, 2016)</th>
<th>Examples from the current study under the topic of <em>Should we allow iPads in our classroom?</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>adverbial clause</td>
<td>advcl</td>
<td><em>They should be allowed</em> [because they help students write essays]. 2C20106990007</td>
</tr>
<tr>
<td></td>
<td><em>The accident happened</em> [as night fell].</td>
<td></td>
</tr>
<tr>
<td>clausal complement</td>
<td>ccomp</td>
<td><em>I think</em> [that the principal should allow students to use iPads]. 2C30105020025</td>
</tr>
<tr>
<td></td>
<td><em>I am certain</em> [that he did it].</td>
<td></td>
</tr>
<tr>
<td>clausal prepositional complement</td>
<td>pcomp</td>
<td><em>Any students caught misusing iPads would get various punishments depending on</em> [how serious the rule breaking was]. C20104040018</td>
</tr>
<tr>
<td></td>
<td><em>They heard about</em> [you missing classes].</td>
<td></td>
</tr>
<tr>
<td>relative clause as modifier</td>
<td>rcm</td>
<td><em>I saw the man</em> [you love].</td>
</tr>
<tr>
<td></td>
<td>mod</td>
<td></td>
</tr>
<tr>
<td>clausal subject</td>
<td>csubj</td>
<td><em>[Having technology in the classroom] can help take advantage of our technological advances for the better of our learning and teaching.</em> 2C51406030009</td>
</tr>
<tr>
<td></td>
<td><em>What he said</em> is not true.</td>
<td></td>
</tr>
<tr>
<td>noun as modifier</td>
<td>nn</td>
<td><em>They are better than</em> [school] computers. 2C50905010020</td>
</tr>
<tr>
<td></td>
<td><em>Oil</em> prices are rising.</td>
<td></td>
</tr>
<tr>
<td>passive voice</td>
<td>n/a</td>
<td><em>[IPads should not be allowed] at my school.</em> 2C50705020017</td>
</tr>
<tr>
<td></td>
<td><em>Kennedy has been killed</em></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2.3

Sample Essays with Low, Medium, and High Diversity of Advanced Syntactic Structures (DASS) Scores

**DASS Score = 2 [Low: 10th percentile]**

*I think the principal should allow iPad. Read on to find out why. iPad are good for looking stuff up like word. Instead of taking two hours looking for words in a dictionary take one minute to find a word in an iPad. Plus when kids are good [adverbial clause] they can play on their iPads. They could download games on it. Instead of everybody going on the computer to look stuff up they could use their iPad. The principal could take the iPad away if he she does not deserve it. That is why kids would have iPad [clausal complement].*

[ID: 2C30104030012]

*Note.* One example of each target syntactic structure type present in the essay is marked for clarity purpose. The tagging is not exhaustive.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Gender</th>
<th>Free/Reduced Lunch</th>
<th>Writing Quality Score</th>
<th>CALS Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Male</td>
<td>No</td>
<td>3</td>
<td>.88</td>
</tr>
</tbody>
</table>

**DASS Score = 4 [Medium: 50th percentile]**

*Hey student iPad users! the principal has took away the iPads. We the students think classrooms should be allowed [passive voice] to have iPads for*
three reasons. Reason one! students can use it as a resource and get information off the Internet. Reason two! it gets students interesting in learning about the most boring topics. Reason three! it allows them to learn how to work technology for high school, college and later on in life. What the principal did was wrong. His decision impacted everyone. One way it impacted us is we might have a difficulty learning. Another reason is some textbooks might not have up-to-date information. A final reason is students could get frustrated because they were no causing the problem. We can solve this problem by doing many things. One is to limit access to websites. A second one is the teacher can get an app that can monitor student use. A third one is if a student misuses it they will not be able to use the iPad. With these rules we will be able to get back the iPads and have a tolerance policy against misuse.

[ID: 2C51405020020]

Note. One example of each target syntactic structure type present in the essay is marked for clarity purpose. The tagging is not exhaustive.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Gender</th>
<th>Free/Reduced Lunch</th>
<th>Writing Quality Score</th>
<th>CALS Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Female</td>
<td>No</td>
<td>4</td>
<td>3.32</td>
</tr>
</tbody>
</table>
IPads should be allowed [passive voice] at school. They are a great tool for learning and can help students achieve many different things. First students can use the iPads as agendas and can set reminders on them. So they can remember when assignments are due [adverbial clause]. Also students can create presentations on the iPads and those presentations can be projected on the board when they are presenting another thing to take into consideration is that kids can access their work on the iPads [clausal complement]. Some cons that go along with the iPad [relative clause as modifier] might be students can access music games [noun as modifier] and the Internet which can be a big distraction. A way to fix that problem would be to limit the time in class that iPads can be used. Also monitoring and blocking sites that seem to take up the most time with kids might help. To prevent hurtful things being said over the Internet [clausal subject] iPads could be taken away as a punishment if someone is caught. I think that despite the few cons of having the iPads they should be allowed in schools. They are a great thing students are faculty to have and redeem benefits from.

[ID: C20106020001]

Note. One example of each target syntactic structure type present in the essay is marked for clarity purpose. The tagging is not exhaustive.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Gender</th>
<th>Free/Reduced Lunch</th>
<th>Writing Quality Score</th>
<th>CALS Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Female</td>
<td>No</td>
<td>6</td>
<td>4.33</td>
</tr>
</tbody>
</table>
References


Study 3.

Developing Argumentation Complexity Scale (ACS) to Characterize and Evaluate Fifth-to-Eight Grade Argumentative Discourse

Abstract

The current study examined individual variability and developmental trends in argumentation complexity as displayed in mid-adolescents’ written argumentative essays. The study had three aims: 1) to describe and compare the incidence of various argumentative elements in mid-adolescents’ essays; 2) to explore a novel scale to score each essay for argumentative complexity; and 3) to test the validity of the novel scale by assessing its association with students’ grade levels; essays’ writing quality; and students’ receptive academic language skills. The analytical sample included essays produced by a cross-sectional sample of fifth to eighth graders (N = 363) from urban school districts in the New England and Mid-Atlantic regions of the United States. First, all essays were coded using the researcher-developed coding scheme informed by data-driven insights, as well as by the integration two lines of research: structural approach (i.e., differentiating claim vs. support in argumentation; Toulmin, 1958/2003) and perspective approach (i.e., differentiating writers’ level of engagement with an alternative position in argumentation; Kuhn & Cromwell, 2011). The coding scheme enabled the identification of the following argumentative elements: Own Claim, Mitigated Claim, Counter Claim; Own Support, Solution Support, Critique Support, and
Counter Support. Results revealed that, as expected, mid-adolescent writers were more likely to generate Own Claims than Own Support; however, unexpectedly, students were more likely to generate Counter Supports than Counter Claims. After the incidence of each element (i.e., the proportions of essays that included a given element) was calculated, elements were ranked based on their comparative incidence. A 5-point *Argumentation Complexity Scale (ACS)* was generated based on the general patterns of the element combinations and the individual differences in written production, with higher scores given to essays that included argumentative elements representing writers’ higher levels of engagement with positions different from their own. Student in eighth grade received significantly higher ACS scores than those in fifth, sixth, or seventh grade. Using multiple regression approaches, essays’ scores on ACS were found to have significant positive associations with their traditionally scored writing quality receptive and students’ academic language skills and, controlling for students’ sociodemographic background.
Introduction

Argumentation is “the act or process of forming reasons and of drawing conclusions and applying them to a case in discussion” (Merriam-Webster, n.d.). The ability to clearly express the reasoning that justifies taking a particular position on a topic has been acknowledged as an important goal of literacy education (Crowhurst, 1990; Ferretti & Lewis, 2013; NAEP, 2011; Newell, Beach, Smith, & VanDerHeide, 2011). The Common Core State Standards (CCSS, 2010), as well as several other college-and-career readiness standards, confirm the centrality of argumentative writing skill. The CCSS define argumentative writing requirements for upper elementary and middle school students as “a reasoned, logical way of demonstrating that the writer’s position, belief, or conclusion is valid.” Specifically, fifth graders are expected to “write opinion pieces on topics or texts, supporting a point of view with reasons and information;” eighth graders should “distinguish the claim(s) from alternate or opposing claims...and maintain a formal style”. However, despite an overall consensus that students should be prepared to be proficient argumentative writers, U.S. students have long been struggling with this skill. More than two-thirds of fourth and eighth graders in the U.S. have performed consistently below grade level on evaluations of argumentative writing over the last three decades (Applebee, 1986; Graham et al., 2014; NAEP, 2011; Persky et al., 2003). The most recent national writing assessment found that 76% of eighth graders did not reach the proficient level in argumentative writing (NAEP, 2011). This is a persistent educational challenge for
which research should offer insights, on identifying the elements that constitutes the argumentation discourse and describing the characteristics of the texts that differentiate levels of argumentative writing quality.

**Quantifying Argumentation Writing Quality: Prior Research**

Whereas research on argumentative writing during the college years and beyond is extensive, very little is known about this genre for students in mid-adolescence (approximately 10 to 13 years of age), the age group of interest in the current study. Writing proficiency levels are often determined, both for educational and research purposes, using broad rubrics (e.g. Andrade et al., 2010; Beard et al., 2016; Beers & Nagy, 2009; Figueroa et al., 2018; McNamara et al., 2010; Olinghouse & Wilson, 2013; Vera et al., 2016). Most often, these rubrics consist of holistic scoring for general features of writing, such as development of ideas and organization of ideas (NAEP, 2011), typically without close attention to the genre-specific elements that comprise the ideas. As a result it is still unclear, from the currently available data, to what extent students can produce the core elements of argumentative writing mandated in the standards (CCSS, 2010), such as stating claims from their own or the opposing position, or providing support -- evidence or explanations-- for these positions. Motivated by the gap in this pedagogically relevant area of research, I chose to focus on investigating the elements, rather than holistic features, of argumentative writing produced by young adolescents.
Beyond the study of broad argumentation quality features, two major approaches have focused on the types of argumentative moves that writers make to advance their stand in writing. One approach is Toulmin’s argumentation model and its adaptations which focused on identifying the structural elements in argumentation (hereafter *structural element approach*) (Toulmin, 1958/2003; Belland, 2010; Glassner et al., 2005; Knudson, 1992; McCann, 1989; McNeill, 2011; Moore & MacArthur, 2012; O’Hallaron, 2014; VanDerHeide & Newell, 2013). The other approach is Kuhn and her colleagues’ idea unit scheme which focuses on categorizing and ranking argumentative moves based on the writer’s perspective (hereafter *perspective element approach*) (Kuhn & Crowell, 2011; Kuhn et al., 2016). In the next two sections, I synthesize the findings from the two approaches and their implications to analyzing argumentation writing development.

**Structural Element Approach: Toulmin’s Model of Argument and Its Adaptations**

Toulmin’s seminal study (1958/2003) identified six types of argumentative moves produced by mature adults: *claim, ground, warrant, backing, qualifier,* and *rebuttal*. The central argumentative move is *Claim*, defined as “an assertion put forward publicly for general acceptance” (Toulmin et al., 1979, p. 29). The types of argumentative moves that serve to justify the claim include: *ground* (i.e. the evidence on which the assertion is based), *warrant* (i.e. explanations that link the
evidence to the assertion), and backing (i.e., additional explanations that advance warrants from a different angle). As the three argumentative moves all directly serve the purpose of supporting the claim, I collectively label these argumentative moves as support. Apart from claim and support, a rebuttal is an acknowledgement of an alternative view of the situation; a qualifier is a word such as “mostly” or “usually” that indicate the scope of the argumentative moves.

Studies on adolescent writing found that students’ use of claims developed earlier than their use of support. Earlier studies found that although almost all students produced claims, sixth graders hardly produced any support, whereas ninth graders produced some support but with poor quality. In these studies, the argumentative writing quality was mostly accounted for by claim quality. More recent research on fifth graders’ argumentation in a science context found that among the students who were able to produce written argumentation, three quarters of them produced both claim and support, while a quarter of them produced just claims without support (McNeil, 2011). There is a lack of recent research using the structural element approach to analyze writing in middle school. One study on seventh graders’ oral scientific argumentation found that both claim and support were present in students’ speech, but students’ language was less clear and less relevant when providing support than when stating claims (Belland, 2010). In the abovementioned studies, the type of support produced by students, if any, was mostly warrant or ground; in addition, studies found that backing was an
element that was almost non-existent in upper elementary and middle school (Belland, 2010; Knudson, 1992; McCann, 1989; McNeil, 2011).

There was little research on adolescent students’ use of rebuttal, that is, their acknowledgement of an alternative view of the situation. Earlier studies attempted to examine this argumentative move but found near-zero frequency in this category (McCann1989; Knudson, 1992). One study of fifth-grade English learners found three out of a total fifteen writers “anticipates and responds to an opposing position” (O’Hallaron, 2014:312), an argumentative move that corresponds to the definition of rebuttal. However, the generalizability of this study was unclear due to its small sample size.

In short, studies following the structural element approach identified a set of argumentative moves that constitute mature argumentation, from which two main distinct categories emerged in analyses on the upper elementary and middle school grade levels: 1) claim, which is the thesis at the center of the argumentation, and 2) support, which is provided in service of validating the thesis. These studies also found that claim emerges earlier, (i.e., it is present in essays produced in earlier grades), than support. Nonetheless, this line of research has not offered sufficient insights into how writers acknowledge and respond to a position different from their own.

**Perspective Element Approach: Kuhn et al.’s Idea Unit Coding Scheme**
Kuhn and colleagues used a different approach to identify the argumentative moves in young adolescents’ writing. This approach focused on identifying text segments according to the types of perspectives included in the text (hereafter *perspective element approach*). The text segment was called *idea units*, defined as “a claim together with any reason and/or evidence supporting it… [that] most often consisted of a single sentence but could be up to two or three sentences in length.” (Kuhn et al., 2016, p. 100). The *idea units* were then categorized based on the writer’s perspective as *own-side only perspective*, *dual-perspective*, *integrated-perspective*, etc.¹ Specifically, an *own-side only perspective idea unit* is one in which the writers support their favored position by describing its positives; in other words, it does not include any engagement with the writer’s opposing position. A *dual-perspective idea unit* is one in which the writers support their own position by critiquing an alternative view; therefore, it represents a higher level of the writer’s engagement with the opposing position. An *integrated-perspective idea unit* is one in which the writers state the positives of an alternative view or the negatives of their own view; in other words, it represents the highest level of writers’ engagement with the opposing position.

The perspective element approach provides a refined lens to detect how students engage with the opposing position, an action that requires both language

---

¹ An *own-side idea unit* is also named as a *support-my-own idea unit*. A *dual-perspective idea unit* is also named as an *weaken-other idea unit*. An *integrated-perspective idea unit* is also named as a *weaken-my-own or support-other idea unit*. Kuhn and colleagues’ coding scheme also included idea units which are *no argument*, *repeated argument*, and *however argument*, which are not reviewed in detail here since they are not directly related to the current study.
and thinking skills. Upper elementary school writers able to include positions counter to their own in their essays also tend to perform better on syntactic complexity and verbal analogical reasoning tasks (Nippold & Ward-Lonergan, 2010). The perspective element approach studies revealed that sixth graders generally were not able to explicitly acknowledge the opposing position in writing, as shown in the absence of *integrated-perspective idea units* in their essays (Kuhn & Crowell, 2011; Kuhn et al., 2016). However, even without explicit acknowledgment, some of these young writers were able to critique the opposing position in an effort to support their own position, as shown in the presence of *dual-perspective idea units* in their essays. For example, when responding to the prompt “Do you agree with experience-based pay or equal pay for teachers?”, writers supported the experience-based pay by pointing out the negative consequences of equal pay: “If new teachers got the same pay, experienced teachers would get fed up and quit” (Kuhn & Crowell, 2011; Kuhn et al., 2016).

Furthermore, the perspective element approach documented levels of complexity in students’ argumentative writing analyses of sixth to eighth graders showed that among the three types of idea units, the *own-side only perspective* idea units were the most frequent, followed by *dual-perspective*, and the *integrated-perspective* ones were the least frequent (Kuhn & Crowell, 2011; Kuhn et al., 2016). Kuhn and her colleagues interpreted the production of less commonly seen types of idea units, those that entail engaging with positions beyond one’s own, as indicating a developmentally higher level of argumentation.
Nonetheless, I argue in this paper that the perspective element approach does not fully capture the variability in students’ argumentation, particularly within dual-perspective idea units. These studies defined dual-perspective idea units as “the negatives of the opposing position” (Kuhn & Crowell, 2011, p. 548), but this definition might ignore other emerging argumentative moves through which writers strengthen their own position with some level of engagement with the opposing position. For instance, young writers’ inclusion of contingencies or action plans might offer an emerging, even if implicit, response to an alternative perspective. In responding to the prompt of “do you agree with experience-based pay or equal pay for teachers”, for example, writers may express their endorsement of the experience-based pay by stating “Teachers with more experience should get paid more if they help new teachers with their work”. In this case, the statement does not fit into the dual-perspective definition, as it does not point out the negatives of the opposing position. Rather, the writer mitigates his/her own position on experience-based pay by attaching a contingency in the form of the if clause. In another example, the statement “We can ask the government to set up extra fund for experienced teachers” offers an action plan proposed to offset a potential problem with the opposing position (e.g., We don’t have extra money to pay the experienced teachers). In the two examples above, the contingency or solution demonstrates an implicit engagement with a potential opposing position; however, it is unclear how such segments would be coded in the perspective element approach studies.
In short, the perspective element approach segments an argumentative text into *idea units* and categorizes them by levels of writers’ engagement with the opposing position, as *own-side only perspective*, *dual perspective*, and *integrated perspective*. Findings suggest that the three perspectives represent a hierarchy in writing development. Nonetheless, this approach is limited in describing the variability in writers’ engagement with the opposing position while strengthening their own, a crucial area in argumentation.

**Integrating the Structural Element and Perspective Element Approaches**

The two major approaches to analyzing the arguments produced by young adolescents, the structural element approach and the perspective element approach, have advantages and limitations. For upper elementary and middle school grades, the structural element approach is most relevant in (a) differentiating *claims* from *support*; and in (b) documenting that *claims* develop earlier than *support*. However, this approach offers no categories that capture the writers’ levels of engagement with the opposing position. On the other hand, the perspective element approach does not distinguish *claim* from *support* but does capture gradual advances in developing writers’ incorporation of perspectives beyond their own. This approach differentiates not only the writer’s own position and the opposing position, but also a more intermediate level engagement with the opposing position (*dual perspective*, i.e. the writer weakens the opposing position by providing its negatives). This distinction is developmentally relevant because
the weakening of the opposing position has been shown to develop earlier than the opposing position itself (Kuhn & Crowell, 2011; Kuhn et al., 2016). Nonetheless, the perspective element approach is limited in its lack of claim-support distinction. By definition, an idea unit is “a claim together with any reason and/or evidence supporting it” (Kuhn & Crowell, 2011; Kuhn et al., 2016). The claim-support distinction is important, though, in students’ argumentative writing development as shown in studies carried out with the structural element approach. Without this distinction, it is unclear whether the increased presence of dual-perspective idea units found as a result of the intervention consists of more claims, more support, or both (Kuhn & Crowell, 2011; Kuhn et al., 2016).

The complementary strengths of the structural and perspective element approaches suggests the value of integrating them in a single analytic scheme. Furthermore, no study to my knowledge has constructed a scoring scale from an integrated approach of the argumentative elements. Thus, the current study’s first aim was to identify and compare the incidence of various argumentative elements in mid-adolescents’ essays. The second aim was to explore generating a novel scale to score each essay based on the combination of higher- and lower-incidence argumentative elements. The third aim was to examine the evidence on the validity of the novel scale by assessing the scores’ association with a) students’ grade levels; b) essays’ writing quality; and c) students’ receptive academic language skills. Therefore, the research questions for the current study are:
RQ 1: Based on fifth to eighth graders’ argumentative essays, what elements can be identified in adolescents’ argumentative writing?

RQ 2: Can an Argumentation Complexity Scale be generated based on an integrated analysis of structural and perspective element patterns?

RQ 3: Is there evidence to support the validation of the Argumentation Complexity Scale?

RQ 3a: Did students’ performance scored by the Argumentation Complexity Scale exhibit differences between grade levels?

RQ 3b: Did students’ performance scored by the Argumentation Complexity Scale predict the overall writing quality?

RQ 3c: Is students’ performance scored by the Argumentation Complexity Scale associated with students’ receptive academic language skills?

For RQ 1, I developed an Argumentative Element Coding Scheme that integrates the structural elements approach and the perspective elements approach. I hypothesized based on the structural elements approach, that students would be more likely to generate claim than support, at all levels of engagement of the opposing position; I also hypothesized based on the perspective elements approach, that students would be less likely to generate elements with higher levels of engagement of the opposing position, on both claim and support. For RQ
2, I generated the *Argumentation Complexity Scale*. Based on theory-based assumptions and data-driven insights, I anticipated that I would find a complexity gradient manifested in different types and combinations of claims and supports per essay. For RQ 3, I anticipated that students at higher grade levels would tend to score higher on the *Argumentation Complexity Scale*, that the *Argumentation Complexity Scale* would be positively associated with essays’ traditionally scored writing quality, and that students’ scores on the *Argumentation Complexity Scale* would be positively associated with their receptive academic language skills.

**Methods**

**Participants**

The full sample of the study included 512 fifth-to-eighth graders from Title 1 urban public schools in the Northeastern and Mid-Atlantic regions of the United States. Participating students were part of the control group in a large-scale literacy intervention. Since the current study aims to investigate general developmental patterns and individual differences, rather than a treatment effect, the treatment group was not included in the current study. Participants’ socio-demographic backgrounds are shown in Table 3.1. About half of the participants were female; about two-thirds of the participants were eligible for free/reduced-price lunch. The vast majority (97%) were native English speakers. The two largest race/ethnicity sub-groups in the sample were White (41%) and Black...
(41%), followed by Latinx (13%). The sample consisted of 20% fifth graders, 30% sixth graders, 30% seventh graders, and 20% eighth graders.

**Procedures**

I focused on participants’ responses to one writing prompt administered at the end of spring 2014. The writing prompt was: *Should we allow iPads in our classrooms?* The writing task was developed by the IES-funded *Catalyzing Comprehension through Discussion and Debate (CCDD)* team (Jones et al., 2019; LaRusso et al., 2016; Lawrence et al., 2015; Snow et al., 2009) to assess upper elementary and middle school students’ writing. Participants were given 20 to 25 minutes to write an argumentative essay and were provided with the following scenario: their school principal had decided to stop the school’s policy of providing iPads to students, thus participants were asked to take a position and to write an argumentative essay to be published by their school newspaper. Participants read a brief description of why iPads had been popular and why they were subsequently prohibited. In their essay, students were asked to give reasons to support their position, to try to convince people, to explain the impact on others, and to discuss potential alternative resolutions to the problem. Participants wrote the essays in the paper-and-pencil format (see full prompt in Appendix 3.1).

**Data Preparation**
Prior to analysis, all the hand-written essays were transcribed using the *Code for the Human Analysis of Transcripts (CHAT)* conventions (MacWhinney, 2000). All spelling errors were corrected in the transcribed essay data in order to assure that human scorers of writing quality were not negatively biased by non-relevant misspellings or other orthographic features. Original files with misspellings were also preserved.

**Measures**

**Writing Quality Measure: Dimension Scores**

Two dimensions of writing quality, scored as part of a holistic writing rubric, were included in this analysis: *Organization* and *Development of Ideas*. Students’ responses were scored using a holistic rubric. The rubric, informed by the NAEP (2011) Writing Framework, includes four dimensions: (1) Position: the number of sides that the essay considers; (2) Organization: the extent to which the essay is coherently structured. (3) Development of Ideas: the degree of depth, complexity, elaboration, and coherence of reasons provided; (4) Clarity: the extent to which the essay conveys information in a precise and unambiguous manner. Each dimension was scored on a 4-point scale, from which the overall writing quality score was generated on a 6-point scale. The dimension of Position was scored with reference to the coding scheme developed for the current study. The dimension of Clarity is not related to the research questions of the current study. Therefore, the two dimensions were not included in the validity check for the
novel instrument developed in the current study. Only the Organization and Development of Ideas dimensions were included in the analyses. The essays were scored by a team of three research assistants who are graduate students specializing in education-related areas with prior experience as classroom teachers. The scoring team were trained with argumentative essays during group sessions. In the group training, each essay was scored by all three scorers guided by the holistic writing rubric, which included anchor essays at each level. A high inter-rater reliability was achieved on the basis of 20% of the sample, with Kendall's Coefficient of Concordance for Ordinal Response higher than .92 on all dimension scores (i.e., Position: .92; Development of Ideas: .99; Organization: .98; Clarity: .99).

Receptive Academic Language | Core Academic Language Skills (CALS)

Instrument

Participants’ receptive academic language skills were measured using the Core Academic Language Skills (CALS) Instrument, a researcher-developed, paper-and-pencil assessment for students in grades 4 to 8 (Barr et al., 2019; Uccelli et al., 2015). The CALS Instrument measures seven domains of academic language skills: unpacking dense information, connecting ideas logically, tracking participants, interpreting writers’ viewpoints, understanding metalinguistic vocabulary, understanding text organization, and recognizing academic register. It includes two vertically equated forms: Form 1 for fourth, fifth, and sixth graders
(\(\alpha = .90\), total items = 49) and Form 2 for seventh and eighth graders (\(\alpha = .86\),
total items = 46). Scores were generated using Rasch item response theory analysis.

**Analytical Approach**

A mixed-method approach was adopted for the current study. First, I developed a qualitative coding scheme that includes the argumentative elements derived from integrating the structural and perspective element approach as well as those that emerged in the coding process. Then, I conducted proportion tests to examine the hypothesized complexity difference among the elements, based on the elements’ presence or absence in essays. After that, I proposed an *Argumentation Complexity Scale (ACS)* to evaluate the full text, taking into consideration the patterns of element combinations and students’ individual differences in text generation. Finally, I conducted multiple regressions to test for the validation of ACS. A set of regressions were conducted to test if there is any between-grade difference among students on ACS, controlling for students’ sociodemographic background. An additional series of regressions were conducted to test if the ACS scores are significantly and positively associated with the two discourse dimensions (i.e., Organization, Development of Ideas) of the essays’ holistic writing quality and with the students’ scores on their receptive academic language skills.
Argumentative Element Coding Scheme

I developed an Argumentative Element Coding Scheme (see Appendix 3.2) integrating the structural and perspective elements approaches. Each essay was coded line by line. Each sentence or part of a sentence in an essay received one of the eight mutually exclusive codes. The definitions and examples for the codes are as follows:

- **Own Claim**: An assertion that declares the writer’s own position without consideration of the opposing position, or a direct objection to the opposing position. (e.g., *iPads should be allowed in our school.*)

- **Mitigated Claim**: An assertion that declares the writer’s own position with consideration of the opposing position, such as contingency or concession. (e.g., *iPads should be allowed in our school if students can follow the rules.*)

- **Counter Claim**: An assertion that declares the opposing position. (e.g., *Some people think iPads should not be allowed in our school.*)

- **Own Support**: The advantages of the writer’s own position. (e.g., *We can make powerpoints on iPads.*)

- **Solution Support**: Action plans proposed to solve a problem that may potentially be raised from the opposing position. (e.g., *We can block the bad apps on iPads.*)
- **Mitigated Support**: Critiques of the writer’s opposing position. (e.g., *Students will be upset if iPads are taken away.*)

- **Counter Support**: Advantages of the writer’s opposing position; disadvantages of the writer’s own position. (e.g., *Some students play video games on iPads.*)

- **Other**: Non-argumentative or unclear utterances

Most argumentative elements were directly derived from the integration of the structural and perspective elements approaches. Own Claim and Own Support were identified by further categorizing Kuhn et al. (2011, 2016)’s “Own-side only” argument into claim and support, which was defined according to Toulmin (1958/2003)’s school of research. Similarly, Counter Claim and Counter Support were identified by further categorizing Kuhn et al. (2011, 2016)’s “Integrative perspective” argument. Mitigated Support corresponds to Kuhn et al. (2011, 2016)’s definition of “Dual perspective” argument. Mitigated Claim in the current coding scheme represents an intermediate level of engagement with the opposing position stated in the form of claim. It was not explicit how such content would have been coded in Kuhn et al. (2011, 2016)’s framework. In addition, during the pilot coding process, Solution Support emerged as a stand-alone element which was present even when Mitigated Support or Counter Support was not. Given the student-proposed solution is addressing to an audience who hold an opposing
position, but the solution itself is not a direct confrontation or acknowledgment of the audience, this element is coded as an independent element as an emerging engagement with the opposing position.

**Qualitative Coding**

Essays in the whole sample ($N = 512$) were coded in three steps. The first step was identifying essays with a clear stance to determine whether they favored allowing iPads or not allowing iPads. Essays with unclear stances ($n = 37$) were excluded from the current analysis. The second step was differentiating affirmative-stance essays ($n = 363$) from negative ones ($n = 112$). This step is necessary in the procedure because the directionality of the stance determines the coding of the writer’s own position and the opposing position. For example, the statement “Some people said iPads can help us learn better” can be an Own Support in an affirmative stance essay, but would be a Counter Support in a negative stance essay. After each essay received a line-by-line coding, the presence or absence of each code within an essay was marked as 1 (i.e., present) or 0 (i.e., absent). A team of three research assistants all coded 20% of the whole sample in MAXQDA, a qualitative coding software. They reached high inter-rater reliability (PABAK > .90) on each of the seven argumentative elements. After that, each research assistant worked on a different subset of the sample.

**Final Analytical Sample**
As pilot coding suggested that affirmative and negative stance essays exhibit different argumentative component distributions, I chose to focus on the affirmative essays as the final analytical sample of the current study ($N = 363$, 71% of the full sample). As shown in Table 3.1, the final analytical sample has comparable socio-demographic background with the full sample.

**Results**

**Patterns of Argumentative Elements**

The incidence of each argumentative element (i.e., the proportion of students in the final analytical sample who produced this element) is reported in Table 3.2. Descriptive statistics showed that for *claims*, Own Claim was the most common type, with an incidence of 97%, whereas Counter Claim was the rarest type, with an incidence of 12%; for *supports*, Own Support was the most common type, with an incidence of 92%, whereas Counter Support was the rarest type, with an incidence of only 29%.

To test the hypotheses for RQ 1, I conducted proportion tests to compare the incidence between the argumentative elements. The first set of proportion tests was conducted to compare claim and support at different levels of engagement with the opposing position. As shown in Figure 3.1, the incidence of Own Claim (97%) was significantly higher than that of Own Support (92%) ($z = 3.25; p$  

---

2 The results on the negatives stance essays in comparison with the affirmative essays will be reported in a separate paper (Deng, in preparation).
In contrast, the incidence of Counter Claim (12%) was significantly lower than that of Counter Support (30%) \((z = -6.25; p < .001)\); similarly, the incidence of Mitigated Claim (17%) was significantly lower than that of either Solution Support (74%) \((z = -15.34; p < .001)\) or Critique Support (45%) \((z = -7.97; p < .001)\).

The next set of proportion tests were conducted to compare different levels of engagement with the opposing position for claim and for support. For claim, the incidence of Own Claim (97%) was significantly higher than that of Mitigated Claim (18%) \((z = 21.56; p < .001)\), which in turn was significantly higher than that of Counter Claim (12%) \((z = 2.50; p < .05)\). Similarly, for support, the incidence of Own Support (92%) was found to be significantly higher than that of Solution Support (73%) \((z = 6.55; p < .001)\), which was significantly higher than that of Critique Support (44%) \((z = 8.13; p < .001)\), which in turn was significantly higher than that of Counter Support (29%) \((z = 3.93; p < .001)\).

**Constructing the Argumentation Complexity Scale (ACS)**

The presence and absence of the seven argumentative elements could possibly form 128 (i.e., \(2^7\)) unique combinations. The final analytical sample included 48 unique combinations. For RQ 2, I explored generating an *Argumentation Complexity Scale (ACS)* to rate the argumentative element combinations.
Complexity Gradients of Claim and Support Element Combinations

I generated a complexity gradient for claim element combinations and one for support element combinations respectively, and then integrated the two gradients as the ACS. For either gradient, I followed three criteria to rank the element combinations:

1) Rarity. As informed by the RQ 1 results, rarer elements would generally be rated as more complex. For example, Critique Support was produced by 44% of the students, while Solution Support was produced by 73% of the students, a statistically significantly higher percentage. The result supported rating Critique as more complex than Solution.

2) Competence Scope. Students who have produced a more complex element were expected to have possessed the competence of producing a less complex element. For example, among the students who produced Counter Support (n = 111), 78% of them also produced Solution Support in their essays, a percentage statistically significantly higher than chance (.5). The result supported rating Counter as more complex than the Solution. Another example is that Mitigated Claim would be rated as more complex than Own Claim because the former by definition is the latter plus contingency or concession.

3) Diversity. Essays including a larger variety of elements would be rated as more complex than those including a smaller variety. For example, although Critique Support and Counter Support were not found to differ in complexity according to the previous two criteria, essays which included both elements would
be rated as more complex than essays which included only one of the two elements.

The three criteria were simultaneously applied when rating the essays. As shown in Table 3.3, the claim element combinations were categorized as two complexity gradients; as shown in Table 3.4, the support element combinations were categorized as four complexity gradients.

**Integrating Claim and Support Complexity Gradients to Generate ACS**

In order to generate a single dimension for the Argumentation Complexity Scale (ACS), I integrated the two-level claim complexity gradients and the four-level support complexity gradients. The ACS used the support level as the baseline score. For essays which were at the lower claim level, their ACS score would correspond to their support level, ranging from 1 to 4 points. For essays which were at the higher claim level, their ACS score would be 1 point higher than their support level.

As shown in Table 3.5, the ACS scores for all essays in the sample ranged from 1 to 5 points. Essays with a point of 1 on ACS were those with Own Claim and Own Support only, without any engagement with the opposing position. Essays with a point of 2 on ACS were those one higher level engagement with the opposing position at either claim or support. Essays with point of 5 on ACS have the highest level on both claim and support. The example essays at each point of ACS were presented in Appendix 3.
Examining Evidence on the Validation of the Argumentation Complexity Scale (ACS)

The descriptive statistics of students’ scores on ACS, the two holistic writing quality dimensions considered in this study (i.e., Development of Ideas and Organization), and the receptive academic language (i.e., Core Academic Language Skills, or CALS) are reported in Table 3.6. The distribution of the ACS scores was shown in Figure 3.2. In the sample, 16% of the essays received a score of 1, 20% received a score of 2, 37% received a score of 3, 19% received a score of 4, and 8% received a score of 5. Shapiro-Wilk test for normality showed that the ACS formed a normal distribution ($z = -1.64, p = .95$). Students’ mean ACS score was 2.84 points ($SD = 1.15$), indicating that on average, students were at intermediate level of engagement with the opposing position. The average students may have possessed the competence of providing Solution Support and approaching the status of generating Critique Support or Counter Support, which is governed by an elementary level of claim (i.e., Own Claim); or the average students may have generated a Mitigated Claim or Counter Claim, which was bolstered by an elementary level of Support (i.e., Own Support). As displayed in the correlation matrix for Table 3.7, ACS scores showed moderately positive correlation with Development of Ideas ($r = .34, p < .001$), Organization ($r = .22, p < .001$), and CALS ($r = .31, p < .001$).
Developmental Trends Reflected by Scores on the Argumentation Complexity Scale (ACS)

I fit a set of multiple regressions to examine the developmental trends in ACS scores. In the modeling process, I used the grade levels as a set of binary variables, with fifth grade as the reference group, to examine if there is statistically significant between-grade difference in ACS scores, after controlling for students’ sociodemographic background (i.e., students’ gender, socioeconomic status as indicated by the free/reduced lunch status, and English language learner status). As shown in Table 3.8, students’ sociodemographic background variables were sequentially entered in the series of models. After dropping the non-significant control variables, the final model (Model 3) included grade levels as the predictor, with students’ gender and socioeconomic status as control variables. Regression results showed that after controlling for students’ gender and socioeconomic status, on average eighth grade essays were scored significantly higher on the ACS than fifth grade ($\beta = .77, SE = .21, p < .001$). The between-grade difference was substantial, as the .77 point difference corresponded to more than 60% of the standard deviation in ACS score. Post-hoc pairwise comparison results showed that eighth grade essays were also significantly higher than sixth grade ($F(1, 347) = 8.94, p < .01$) and seventh grade ($F(1, 347) = 15.35, p < .001$), respectively. There was no statistically significant difference in ACS scores between fifth, sixth and seventh grade.
Scores on Argumentation Complexity Scale (ACS) Predicting Writing Quality and Receptive Academic Language

I fit three sets of multiple regressions to examine whether students’ scores on ACS could predict their scores on Writing Quality or Receptive Academic Language (Core Academic Language Skills CALS). In the modeling process, I used ACS as the independent variable to predict Development of Ideas, Organization, or CALS, respectively, controlling for students’ grade levels and sociodemographic background (i.e., students’ gender, socio-economic status, and English language learner status). Students’ sociodemographic background variables were sequentially entered for each set of models. For the prediction to Development of Ideas, as shown in Table 3.9, after dropping the non-significant control variables, the final model (Model 4) showed that ACS scores positively and significantly predict the Development of Ideas dimension of writing quality, controlling for students’ grade level, gender, and socio-economic status ($\beta = .17, SE = .04, p < .001$). The prediction of ACS scores was substantial, as 1 point difference in ACS score corresponded to .17 point difference, that is, about a fifth of the standard deviation difference, in the Development of Ideas score. In the same vein, as shown in Table 3.10, the final model (Model 3) showed that ACS scores also positively and significantly predict the Organization dimension of writing quality, controlling for students’ grade level and gender ($\beta = .10, SE = .04, p < .01$). Similarly, as shown in Table 3.11, the final model (Model 5) showed that ACS scores also positively and significantly predict the CALS scores,
controlling for students’ grade level, socioeconomic status, and English language learner status ($\beta = .17, SE = .06, p < .01$).

**Discussion**

The current study has three aims: 1) to identify and describe the patterns of elements that constitutes adolescents’ argumentative discourse, 2) to generate an *Argumentation Complexity Scale (ACS)* based on the patterns, and 3) to examine the evidence on the validation of the new scale. The results showed that first, argumentative elements based on an integration of structural and perspective element approaches can be identified in students’ writing. Specifically, by integrating the two approaches and grounded-theory coding, I identified three new elements that described students’ emerging or intermediate engagement with the opposing position: Solution Support, Critique Support, and Mitigated Claim. Their patterns in which students generated argumentative elements shows that support is easier than claim to be when students are engaging with the opposing position. Second, a 5-point *Arguementation Complexity Scale (ACS)* was generated based on the complexity gradients of structural (claim vs. support) and perspective (level of engagement of the opposing position) elements based on the criteria of reflecting general patterns and individual differences in students’ production of argumentative elements. Third, evidence was found in support of validating the ACS: eighth grade showed significantly higher ACS than fifth, sixth, or seventh grade; ACS positively predicted traditionally holistic writing quality scores on
Development of Ideas and Organization; ACS also positively predicted students’ receptive academic language skills.

**Novel Patterns on Structure**

The current study showed that the patterns of claim and support incidence differed by the writer’s level of engagement with the opposing position. Students were more likely to produce Own Claim than Own Support, but more likely to produce Counter Support than Counter Claim, and also more likely to provide Solution Support or Critique Support than Mitigated Claim. This finding partly revises conclusions drawn from previous studies using Toulmin et al.’s coding. Knudson (1992) and McCann (1989) suggested that claims developed earlier than support based on their findings that sixth and ninth graders produced claims but rarely produced support. McNeill (2011) also found that among fifth graders who wrote arguments on science topics, one-fourth of them produced just claims without any support. Partly consistent with the previous studies, the current study found that that young adolescents in this sample were more likely to produce claim than support when advancing their own position. However, the difference in the current study is significant but small in scale, as the incidence was higher than 90% for both Own Claim and Own Support. This may reflect the fact that schools and educators have been actively responding to rising standards (CCSS, 2010) on argumentative writing, by incorporating instructions on argumentation in English Language Arts. Presumably participants in our study have been also exposed to
these changes in U.S. curricular standards and consequently, showed higher awareness and greater skill in advancing their position. An engaging and familiar topic (i.e., the use of tablets in school) might have also provided conditions that led to higher performance.

More interesting and intriguingly, the current study identifies a novel pattern: when engaging with an opposing position, students are more likely to generate support than claims. This pattern is the reverse of what happens when young adolescent writers advance their own position. One possible explanation for the low incidence of Mitigated Claim is that a contingency or concession needs to be embedded in the form of a dependent clause, which may pose syntactic challenges for many students. One possible explanation for the low incidence of Counter Claim is that students may feel unnecessary to produce this element if they have already provided Counter Support, as the differentiation between the two elements were not required in instruction; another possible explanation for the low incidence of Counter Claim is that acknowledging the opposing position is not recognized as a helpful or even necessary move in written argumentation for most students in this age group. Indeed, the CCSS (2010) only require students to differentiate own claims from counter claims in writing starting at eighth grade, without any requirement on providing support at different levels of engagement with the opposing position. Even though the current study did not have information on pedagogical practices that students received or the students’ mental activities during their writing process to explore the possible explanation, it adds
more evidence to support that *claim* and *support* are two independent argumentative elements.

**Elaborated Patterns on Perspective**

The current study found that higher engagement with the opposing position indicates higher challenge for young adolescents’ writing. Student almost always stated their own position (i.e. generating Own Claim or Own Support), but were less likely to have emerging or intermediate engagement with the opposing position (i.e. generating Mitigated Claim, Solution Support, or Critique Support), and even more rarely have high level of engagement with the opposing position (i.e. generating Counter Claim or Counter Support). The finding is consistent with Kuhn et al.’s finding of a frequency difference among three types of idea units along the perspective spectrum: *own-side only perspective, dual perspective*, and *integrated perspective*. Furthermore, the current study expands Kuhn et al.’s findings by separately confirming the hierarchy in the area of *claim* and *support*, and by using incidence rather than frequency of argumentative components as the measurement unit. Although frequency can describe the variability in the volume of argumentative component production, incidence is a better reflection of emerging competence.

**Solution as an Initial Attempt to Engage with the Opposing Position**
One contribution of the study is its identification of Solution Support as an emerging attempt to engage with the opposing position. According to the argumentation complexity level indicated by types of support, about a quarter of students in the sample ($n = 94$) provided Solution Support beyond providing explanations or evidence for their favored position. Solution support was provided still in the absence of critiquing or acknowledging the opposing position. To my knowledge, no previous studies on young adolescents’ argumentative writing has reported such finding. One possible explanation for the problem-solving orientation is that previous studies did not code Solution as a separate category. Another possible explanation is that young adolescents regard solutions as the most efficient tool to refute the opponents and then close the argument when they first start developing their argumentation skills. An alternative explanation is that the participants were affected by the specific writing prompt. Indeed, the writing prompt includes explicit request for solutions, which may have led participants to produce this component. However, it should be noted that the writing prompt also provided scaffold for Critique Support by requiring participants to explain the potential impact of the principal’s decision, but the proportion of essays that exhibited Critique Support was significantly lower than that of Solution. Therefore, the strong tendency to produce solutions cannot be solely attributed to the request from the writing prompt.
Element-Focused Approach in Measuring Argumentative Writing Complexity

In the current study I identified argumentative complexity elements from integrating the structural and perspective elements approach and data-driven insights, based on which I generated an *Argumentation Complexity Scale (ACS)*. The current study found that 37% of the essays in the sample (n=135) received a score of 3 on ACS. In other words, these students have shown intermediate engagement with the opposing position, a concept closely aligned with Kuhn et al.’s *dual perspective argument*. The level identified the current study is similar to the percentage of control group students who generated *dual perspective idea units* in the Kuhn and Crowell (2011) study, 19% to 38%. However, previous studies did not generate an evaluation of the argumentative writing quality from the argumentative elements. Instead, I adopted a bottom-up approach in measuring argumentative writing quality. In other words, my scoring process starts from identifying microscope features of the discourse (i.e., the argumentative complexity elements), to analyze the patterns of the combinations of the microscope features within each text, and finally generates a macroscope score for by the ranking of the combinations. This is in contrast with the traditionally used holistic approach to the analysis of Argumentative writing (e.g., NAEP 2011), which starts at and ends with treating the full text as the unit of analysis and generates scores on dimensions such as *development of ideas* or *organization of ideas*. Although the holistic approach can yield reliable scores, it is less
informative for supporting students’ argumentation as it focuses on general dimensions of writing and does not identify the conceptual content of the ideas being developed or organized. In contrast, I ultimately constructs and applies a 5-point scale to a full text. The bottom-up process of generating the scores entails a detailed understanding of what types of argumentative moves a writer made, and therefore entails a more precise scoring, not of general writing quality, but instead of the variability found in argumentative writing complexity during mid-adolescence. Even though the element-focus scoring approach in the current study is more labor intensive operationally than the traditional holistic scoring approach and therefore challenging to implement in large scale summative assessments, it can serve as an insightful tool in discourse analysis research on developing academic writers.

**Developmental Trends between Fifth-to-Eighth Grade**

The Argumentation Complexity Scale (ACS) delineate the five levels at which writers increasingly engage with the opposing position. The developmental trend was not found to be progressively linear across grades. Instead, ACS scores were similar across fifth to seventh grade, while significantly higher at eighth grade. On average fifth, sixth, and seventh graders scored below 3 points. In other words, on average students in these grades are already capable of providing solutions, demonstrating an emerging awareness of the opposing position. However, on average fifth, sixth, or seventh graders did not demonstrate the
ability to critique the opposing position or to embed a contingency or concession in their thesis. On the other hand, eighth graders show a significantly higher level than earlier grades. The eighth-grade essays received a mean ACS score of 3.55 points; in other words, on average eighth graders demonstrate their competence of explicitly engaging with the opposing position either in support or in claim, outperforming fifth, sixth, seventh graders who typically generated only a Solution Support as the highest element to engage with the opposite position. This finding is different from Kuhn et al. (2011, 2016)’s, which reported that on average their control group students had not showed improvement in dual perspective production from sixth to eighth grade. One possible explanation for the different finding is that, the participants in the respective studies likely received different instruction in their school settings and thus performed differently in argumentative writing. Another possible explanation is that the respective studies have different writing prompts in terms of the degree of scaffolding provided on background information and content, which elicited different responses from students. As the current study is purely descriptive without investigating explanatory factors related to the described variability in argumentation complexity, it is unclear to what extent the differences found between grades are associated with pedagogical content, testing materials, or developmental progressions.

Implications to Research and Practice

The current study contributes to the body of adolescent writing research by
integrating two existing approaches on identifying the elements in written argumentation: the structural and the perspective element approach. In the process, new argumentative elements were identified from the integration and emerged from data-driven insight. It suggests that detailed discourse analysis with ground-theory approach can shed light on understanding the ideas and content that students produce in their writing. The Argumentation Complexity Scale (ACS) has the potential to serve as a sensitive tool to measure treatment and control group difference in interventions that aim to improve adolescents’ argumentative writing skills. Given ACS delineates students’ emerging and intermediate levels in argumentation, especially in engaging with the opposing position, it has the potential to detect nuances which might have not been found from traditional holistic scoring.

The study has several implications to educational practice such as curriculum development and instruction. It identifies argumentation complexity as an area in need of instructional support and offers evidence of the strengths and needs of a diverse sample of public school students, which in turn can potentially inform the design of future interventions. Instructors can actively raise students’ awareness in detecting the argumentative elements in reading comprehension, in producing them in classroom activities such as discussion and debate, and in including them in their writing output. Furthermore, instructors can use ACS as a lens to analyze students’ writing samples as a diagnostic or formative assessment, for the purpose of identify a student’s zone of proximal development in
argumentation as an instructional target, and thus achieve higher efficiency in writing instruction.

**Limitations**

The current study has several limitations. First, students in the study produced argumentative essays based on a specific prompt and were tested only once. The content produced by students was constrained by the nature of the topic and task. Therefore, the findings presented here reflect the analysis of one piece of writing, and thus, are interpreted as the skills exhibited in one writing performance, not as the full profile of the participating writers. It is possible that a different prompt, for example, a topic on history or social sciences that is outside the everyday school context, would elicit different patterns in argumentation. The current prompt also provided considerable scaffolding; a prompt with minimal or less elaborated scaffolding might have generated less sophisticated responses.

Second, the study only analyzed students’ affirmative essays (i.e., essays whose writers’ own positions is “yes we should allow iPads” and the opposing position is “no we should not allow iPads”). Although the affirmative essays represented the majority (71%) of the sample, it is possible that the negative essays (23% of the full sample) would reveal different patterns. In addition, a small percentage of essays (6% of the full sample) did not show a clear preference in the stance they chose: 2% students in the full sample (n=9) had a thesis of no-preference such as “Both are fine” or “I don’t care”; 3% students (n=16) declared
self-contradictory stances within an essay; 1% students (n=3) did not produce argumentative texts. These essays, though they exhibit illuminating diversity in students’ real-world responses to a writing prompt, were not included in the analyses due to the limited scope of this paper.

Third, the study used a cross-sectional, rather than longitudinal sample, to analyze between-grade differences. The study only tested for students’ argumentative production without testing their knowledge of the argumentative genre. Causal inferences between ACS and the traditional holistic writing quality or students’ receptive academic language skill scores cannot be made, as the current study only tested the relations as association. It is unclear to what degree the results drawn from the current study could be generalized to other student samples.

Finally, the current study only analyzes students’ generation of argumentative elements, one aspect of discourse, in writing quality. It did not analyze other discourse features such as students’ production of transition sentences or organizational markers. The study did not include analysis on the quality or richness of each argumentative element, such as whether the Solution Support a student provided was valid or plausible, or how elaborated a student provided Own Support. It also did not include other linguistic domains such as vocabulary diversity and syntactic complexity that contribute to writing quality. The current argumentation element coding scheme, due to its detailed line-by-line
human coding process, requires a large amount of time in data processing, which in turn limits the volume of texts that could be analyzed within one study.

**Future Research**

The current study suggests a few directions for future research on adolescent writing. Future studies can examine a variety of writing prompts and argumentation topics, as well as elicit responses from students at multiple time points, to be further validated for generalizability. Given the scarcity of research testing the effect of different levels of scaffolding in writing prompts, future research can investigate the relationship between levels of scaffolding and argumentation complexity of young adolescents’ essays. Analyses could be conducted on affirmative as well as negative essays, with additional examinations on the content quality of the argumentative elements and considerations of other non-discourse language domains such as vocabulary or syntax. In addition to the cross-sectional sample that was used in the current study, longitudinal or cohort-sequential samples could be used to further investigate the developmental patterns. Intervention studies on argumentative element instruction with randomized control design could be conducted to test for the potential causal relations among argumentation complexity, writing quality, and receptive academic language skills, with receptive knowledge as well as production of the argumentative elements both included in the intervention and analyses. Last but not least,
machine learning or natural language tools may be trained with the coding scheme and applied to a larger corpus of student essays.

**Conclusion**

In the current study, I identified elements in adolescents’ written argumentation (i.e., Own Claim, Mitigated Claim, Counter Claim, Own Support, Solution Support, Critique Support, and Counter Support) from a cross-sectional sample of fifth-to-eighth grade students by developing a qualitative coding scheme that integrates two major approaches in previous research (i.e., the structural and perspective element approaches) and that incorporates phenomena emerged from the coding process. Analyses on the argumentative element patterns revealed that it is easier for students to generate *claims* than *support* when advancing their own position, whereas it is easier for them to generate *support* than *claim* when they were engaging with the opposing position. Proceeding to directly acknowledge or strengthening the opposing position by stating a Counter Claim or providing a Counter Support, students tended to a contingency or concession (i.e., Mitigated Claim), action plans (i.e., Solution Support), or critiques (i.e., Critique Support), that is, the elements at different levels of engagement with opposing position, as a means to strengthen their own position. It suggests that students’ engagement with the opposing position may not emerge as a stand-alone element in an argumentative essay, but as elements within students’ thinking when they support their own position. The *Argumentation Complexity Scale (ACS)* generated from
the combinations of argumentative elements identified significantly higher performance at eighth grade than fifth, sixth, and seventh grade, positively predicted traditional holistic writing quality scores on Development of Ideas and Organization as well as students’ receptive academic language skills, providing evidence to support the validation of the new scale.
## Tables

### Table 3.1

**Participants’ Socio-demographic Background**

<table>
<thead>
<tr>
<th>Socio-demographic Background</th>
<th>Full Sample $(N = 512)$</th>
<th>Final Analytical Sample: Essays with Affirmative Stance $(N = 363)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>261</td>
<td>51%</td>
</tr>
<tr>
<td>Male</td>
<td>251</td>
<td>49%</td>
</tr>
<tr>
<td>SES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free/reduced lunch Eligible</td>
<td>345</td>
<td>67%</td>
</tr>
<tr>
<td>Free/reduced lunch non-eligible</td>
<td>167</td>
<td>33%</td>
</tr>
<tr>
<td>Language Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Language Learner</td>
<td>14</td>
<td>3%</td>
</tr>
<tr>
<td>Non-English Language Learner</td>
<td>498</td>
<td>97%</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>208</td>
<td>41%</td>
</tr>
<tr>
<td>Black</td>
<td>209</td>
<td>41%</td>
</tr>
<tr>
<td>Asian</td>
<td>8</td>
<td>1.6%</td>
</tr>
<tr>
<td>Latinx</td>
<td>67</td>
<td>13%</td>
</tr>
<tr>
<td>Native/Pacific</td>
<td>2</td>
<td>0.4%</td>
</tr>
<tr>
<td>Mixed/Other</td>
<td>12</td>
<td>2%</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>95</td>
<td>19%</td>
</tr>
<tr>
<td>6th</td>
<td>150</td>
<td>29%</td>
</tr>
<tr>
<td>7th</td>
<td>182</td>
<td>36%</td>
</tr>
<tr>
<td>8th</td>
<td>85</td>
<td>17%</td>
</tr>
</tbody>
</table>
Table 3.2

*Incidences of Argumentative Elements (N=363)*

<table>
<thead>
<tr>
<th></th>
<th>Number of Essays Containing this Element</th>
<th>Incidence (i.e., Percentage of Essays Containing this Element)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own Claim</td>
<td>353</td>
<td>97%</td>
</tr>
<tr>
<td>Mitigated Claim</td>
<td>66</td>
<td>18%</td>
</tr>
<tr>
<td>Counter Claim</td>
<td>42</td>
<td>12%</td>
</tr>
<tr>
<td>Own Support</td>
<td>333</td>
<td>92%</td>
</tr>
<tr>
<td>Solution Support</td>
<td>266</td>
<td>73%</td>
</tr>
<tr>
<td>Critique Support</td>
<td>158</td>
<td>44%</td>
</tr>
<tr>
<td>Counter Support</td>
<td>107</td>
<td>29%</td>
</tr>
</tbody>
</table>
Table 3.3

*Complexity Gradient on Claim Element Combinations (N = 363)*

<table>
<thead>
<tr>
<th>Claim Level</th>
<th>Types of Claims Included in the Essay</th>
<th>Number of Essays</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (low)</td>
<td>Own Claim</td>
<td>261 (72%)</td>
</tr>
<tr>
<td>2 (high)</td>
<td>(Own Claim) + Mitigated Claim <em>and/or</em> Counter Claim</td>
<td>102 (28%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>363 (100%)</td>
</tr>
</tbody>
</table>

*Note.* The element in parentheses is optional, but its presence in an essay is statistically significantly higher than chance (.5)

Table 3.4

*Complexity Gradient on Support Element Combinations (N = 363)*

<table>
<thead>
<tr>
<th>Support Level</th>
<th>Types of Support Included in the Essay</th>
<th>Number of Essays</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (lowest)</td>
<td>Own Support</td>
<td>58 (16%)</td>
</tr>
<tr>
<td>2</td>
<td>(Own Support) + Solution</td>
<td>94 (26%)</td>
</tr>
<tr>
<td>3</td>
<td>(Own Support) + (Solution) + Critique <em>or</em> Counter Support</td>
<td>157 (43%)</td>
</tr>
<tr>
<td>4 (highest)</td>
<td>(Own Support) + (Solution) + Critique <em>and</em> Counter Support</td>
<td>54 (15%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>363 (100%)</td>
</tr>
</tbody>
</table>

*Note.* The element in parentheses is optional, but its presence in an essay is statistically significantly higher than chance (.5)
Table 3.5

*Argumentation Complexity Scale (ACS): 1-to-5 Points (N=363)*

<table>
<thead>
<tr>
<th></th>
<th>Support Level 1</th>
<th>Support Level 2</th>
<th>Support Level 3</th>
<th>Support Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim Level 1</td>
<td>1 pt</td>
<td>2 pts</td>
<td>3 pts</td>
<td>4 pts</td>
</tr>
<tr>
<td>Claim Level 2</td>
<td>2 pts</td>
<td>3 pts</td>
<td>4 pts</td>
<td>5 pts</td>
</tr>
</tbody>
</table>
Table 3.6


<table>
<thead>
<tr>
<th>Grade</th>
<th>ACS (1-5 pts)</th>
<th>Writing Quality Dimensions (1-4 pts)</th>
<th>Receptive Academic Language (CALS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>ACS</td>
<td>2.56 (1.07)</td>
<td>2.84 (1.10)</td>
<td>2.68 (1.12)</td>
</tr>
<tr>
<td>(1-5 pts)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Development of Ideas</td>
<td>2.47 (.71)</td>
<td>2.74 (.76)</td>
<td>2.81 (.80)</td>
</tr>
<tr>
<td>- Organization</td>
<td>2.33 (.60)</td>
<td>2.70 (.78)</td>
<td>2.72 (.80)</td>
</tr>
<tr>
<td>Receptive Academic Language (CALS)</td>
<td>.56 (.93)</td>
<td>1.32 (1.30)</td>
<td>1.30 (1.21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.33</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>English Language</td>
<td>0.03</td>
<td>8.0</td>
</tr>
</tbody>
</table>

*Table 3.7*

**Table 3.7:** Pearson Correlations between the Argumentation Complexity Scale (ACS) Scores, Writing Quality Dimensions, and Academic Language (CAL). Pearson Correlations between the Argumentation Complexity Scale (ACS) Scores, Writing Quality Dimensions, and Academic Language (CAL).
Table 3.8

*Argumentation Complexity Scale (ACS) Scores Predicted by Grade Levels*  
*(N = 363)*

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACS</td>
<td>ACS</td>
<td>ACS</td>
<td>ACS</td>
</tr>
<tr>
<td>Grade 6</td>
<td>0.279</td>
<td>0.281</td>
<td>0.222</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>(1.70)</td>
<td>(1.65)</td>
<td>(1.31)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>Grade 7</td>
<td>0.115</td>
<td>0.092</td>
<td>0.048</td>
<td>-0.036</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.54)</td>
<td>(0.29)</td>
<td>(-0.21)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>0.980***</td>
<td>0.949***</td>
<td>0.771***</td>
<td>0.637**</td>
</tr>
<tr>
<td></td>
<td>(4.96)</td>
<td>(4.70)</td>
<td>(3.71)</td>
<td>(3.04)</td>
</tr>
<tr>
<td>Female</td>
<td>0.372**</td>
<td>0.390***</td>
<td>0.399***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.15)</td>
<td>(3.34)</td>
<td>(3.40)</td>
<td></td>
</tr>
<tr>
<td>¹FRL</td>
<td></td>
<td>-0.391**</td>
<td>-0.426**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-3.04)</td>
<td>(-3.28)</td>
<td></td>
</tr>
<tr>
<td>²ELL</td>
<td></td>
<td></td>
<td></td>
<td>0.379</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.96)</td>
</tr>
<tr>
<td>_cons</td>
<td>2.566***</td>
<td>2.394***</td>
<td>2.704***</td>
<td>2.825***</td>
</tr>
<tr>
<td></td>
<td>(20.05)</td>
<td>(15.97)</td>
<td>(15.03)</td>
<td>(15.52)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.075</td>
<td>0.098</td>
<td>0.122</td>
<td>0.126</td>
</tr>
</tbody>
</table>

*Note.* Grade 5 set as the reference group  
¹FRL: Free-reduced lunch status; ²ELL: English Language Learner Status  
t statistics in parentheses  
*p < 0.05, ** p < 0.01, *** p < 0.001*
Table 3.9

*Argumentation Complexity Scale (ACS) Scores Predicting Essays’ Development of Ideas (N = 363)*

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Development of Ideas</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACS</td>
<td>0.234***</td>
<td>0.208***</td>
<td>0.185***</td>
<td>0.171***</td>
<td>0.153***</td>
</tr>
<tr>
<td></td>
<td>(6.79)</td>
<td>(5.91)</td>
<td>(5.28)</td>
<td>(4.87)</td>
<td>(4.20)</td>
</tr>
<tr>
<td>Grade 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.219</td>
<td>0.248*</td>
<td>0.222*</td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.92)</td>
<td>(2.22)</td>
<td>(1.99)</td>
<td>(1.50)</td>
<td></td>
</tr>
<tr>
<td>Grade 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.318**</td>
<td>0.321*</td>
<td>0.300**</td>
<td>0.254*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.79)</td>
<td>(2.89)</td>
<td>(2.71)</td>
<td>(2.24)</td>
<td></td>
</tr>
<tr>
<td>Grade 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.486***</td>
<td>0.504**</td>
<td>0.426**</td>
<td>0.387**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.50)</td>
<td>(2.71)</td>
<td>(3.06)</td>
<td>(2.74)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.318***</td>
<td>0.333***</td>
<td>0.341***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.08)</td>
<td>(4.28)</td>
<td>(4.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRL</td>
<td></td>
<td></td>
<td></td>
<td>-0.199*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-2.33)</td>
<td></td>
</tr>
<tr>
<td>ELL</td>
<td></td>
<td></td>
<td></td>
<td>-0.034</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-0.14)</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>2.110***</td>
<td>1.930***</td>
<td>1.823***</td>
<td>2.012***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(19.91)</td>
<td>(15.00)</td>
<td>(14.18)</td>
<td>(13.29)</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.116</td>
<td>0.150</td>
<td>0.189</td>
<td>0.201</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Grade 5 set as the reference group*

1FRL: Free-reduced lunch status; 2ELL: English Language Learner Status

*t statistics in parentheses  *p < 0.05,  **p < 0.01,  ***p < 0.001*
Table 3.10

*Argumentation Complexity Scale (ACS) Scores Predicting Essays’ Organization*

(N = 363)

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Organization</td>
<td>Organization</td>
<td>Organization</td>
<td>Organization</td>
<td>Organization</td>
</tr>
<tr>
<td>ACS</td>
<td>0.149*** (4.18)</td>
<td>0.114** (3.14)</td>
<td>0.096** (2.64)</td>
<td>0.085* (2.32)</td>
<td>0.071 (1.88)</td>
</tr>
<tr>
<td>Grade 6</td>
<td>0.344** (2.93)</td>
<td>0.366** (3.15)</td>
<td>0.345** (2.97)</td>
<td>0.300* (2.49)</td>
<td></td>
</tr>
<tr>
<td>Grade 7</td>
<td>0.379** (3.25)</td>
<td>0.382** (3.31)</td>
<td>0.366** (3.16)</td>
<td>0.320** (2.69)</td>
<td></td>
</tr>
<tr>
<td>Grade 8</td>
<td>0.647*** (4.54)</td>
<td>0.661*** (4.68)</td>
<td>0.599*** (4.13)</td>
<td>0.563*** (3.82)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.243** (3.00)</td>
<td>0.254** (3.14)</td>
<td>0.278** (3.36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>¹FRL</td>
<td></td>
<td>-0.158 (-1.77)</td>
<td>-0.139 (-1.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>²ELL</td>
<td></td>
<td>-0.05 (-0.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>2.268*** (20.58)</td>
<td>2.031*** (15.34)</td>
<td>1.949*** (14.58)</td>
<td>2.099*** (13.29)</td>
<td>2.167*** (13.08)</td>
</tr>
<tr>
<td>R²</td>
<td>0.047</td>
<td>0.103</td>
<td>0.128</td>
<td>0.134</td>
<td>0.120</td>
</tr>
</tbody>
</table>

*Note.* Grade 5 set as the reference group

¹FRL: Free-reduced lunch status; ²ELL: English Language Learner Status

*p* statistics in parentheses  

* p < 0.05, ** p < 0.01, *** p < 0.001
Table 3.11

*Argumentation Complexity Scale (ACS) Scores Predicting Receptive Academic Language (CALS) (N = 363)*

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CALS</td>
<td>CALS</td>
<td>CALS</td>
<td>CALS</td>
<td>CALS</td>
<td>CALS</td>
</tr>
<tr>
<td>ACS</td>
<td>0.350***</td>
<td>0.246***</td>
<td>0.242***</td>
<td>0.170**</td>
<td>0.152***</td>
<td>0.166**</td>
</tr>
<tr>
<td></td>
<td>(5.78)</td>
<td>(4.27)</td>
<td>(4.13)</td>
<td>(2.99)</td>
<td>(2.62)</td>
<td>(2.90)</td>
</tr>
<tr>
<td>Grade 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.706***</td>
<td>0.710***</td>
<td>0.626***</td>
<td>0.683***</td>
<td>0.673***</td>
<td>0.641***</td>
</tr>
<tr>
<td></td>
<td>(3.87)</td>
<td>(3.89)</td>
<td>(3.60)</td>
<td>(3.86)</td>
<td>(3.81)</td>
<td>(3.65)</td>
</tr>
<tr>
<td>Grade 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.703***</td>
<td>0.704***</td>
<td>0.637***</td>
<td>0.641***</td>
<td>0.641***</td>
<td>0.641***</td>
</tr>
<tr>
<td></td>
<td>(3.85)</td>
<td>(3.85)</td>
<td>(3.66)</td>
<td>(3.65)</td>
<td>(3.65)</td>
<td>(3.65)</td>
</tr>
<tr>
<td>Grade 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.729***</td>
<td>1.728***</td>
<td>1.408***</td>
<td>1.461***</td>
<td>1.468***</td>
<td>1.468***</td>
</tr>
<tr>
<td></td>
<td>(7.64)</td>
<td>(7.63)</td>
<td>(6.35)</td>
<td>(6.60)</td>
<td>(6.63)</td>
<td>(6.63)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.059</td>
<td>0.125</td>
<td>0.157</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(1.01)</td>
<td>(1.26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1FRL</td>
<td></td>
<td></td>
<td></td>
<td>-0.816***</td>
<td>-0.765***</td>
<td>-0.751***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-5.92)</td>
<td>(-5.52)</td>
<td>(-5.43)</td>
</tr>
<tr>
<td>2ELL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.175**</td>
<td>-1.158*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-2.64)</td>
<td>(-2.60)</td>
</tr>
<tr>
<td>cons</td>
<td>0.343</td>
<td>-0.080</td>
<td>-0.100</td>
<td>0.710**</td>
<td>0.735**</td>
<td>0.768***</td>
</tr>
<tr>
<td></td>
<td>(1.84)</td>
<td>(-0.39)</td>
<td>(-0.47)</td>
<td>(2.91)</td>
<td>(2.92)</td>
<td>(3.06)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.095</td>
<td>0.237</td>
<td>0.238</td>
<td>0.315</td>
<td>0.324</td>
<td>0.320</td>
</tr>
</tbody>
</table>

*Note.* Grade 5 set as the reference group

1FRL: Free-reduced lunch status; 2ELL: English Language Learner Status

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Figures

Figure 3.1

Incidences of Argumentative Elements

Figure 3.1A  Own Claim & Own Support

Figure 3.1B  Mitigated Claim, Solution Support, & Critique Support
Figure 3.1C  Counter Claim & Counter Support

Figure 3.2

Distribution of Essay Scores on the Argumentation Complexity Scale (ACS)
Appendices

Appendix 3.1

Argumentative Writing Prompt

Should we allow iPads in our classrooms?

Imagine that your school decided to give iPads to all students, so everyone has the opportunity to use iPads for learning. Most teachers like the iPads because all students can use online dictionaries and search the internet for interesting information. But in the last few weeks some students have used their iPads to post videos with mean comments about other kids online. Now, many parents are worried about this situation. Because of this problem, your Principal has decided that iPads will no longer be allowed in school!

Write an article for the school newspaper that argues for or against allowing iPads at your school.
* Make sure to give specific reasons to support your position and to convince the people who read the article to agree with you.
* Explain how the Principal's decision can impact you and others.
* Discuss other things that the school community could do to solve the iPad problem.
<table>
<thead>
<tr>
<th>Claim</th>
<th>Definition</th>
<th>Operational Definition in the Current Example</th>
<th>Argumentative Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>We should allow iPads if everyone follows the rules.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The position of banning iPads is wrong.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The position of allowing iPads is not a direct opposition to banning iPads.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The position of allowing iPads is a concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We should allow iPads even though some people don’t like the idea.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is wrong to ban iPads.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some people think we should not allow iPads.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The position of banning iPads is the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An assertion that declares the position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We believe that banning iPads is the position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A counter claim.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A concession, such as a concession of the position with concession of the opposite position.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Writer's position</td>
<td>Counter Support</td>
<td>Writer's position: disadvantages of the iPads</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td>-----------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Some students check bad websites on iPads.</td>
<td>The disadvantages of allowing</td>
<td>The disadvantages of using iPads.</td>
<td>Children of the writer's position.</td>
</tr>
<tr>
<td>Students will be very sad.</td>
<td>Counter Support</td>
<td>Students will be very sad.</td>
<td>Negatively consequences of</td>
</tr>
<tr>
<td>We can monitor what websites students check on iPads.</td>
<td>The advantages of allowing</td>
<td>The disadvantages of using iPads.</td>
<td>Advantages of the writer's position.</td>
</tr>
<tr>
<td>Would bring</td>
<td>the opposite</td>
<td>Problem that may potentially be</td>
<td>Action plans proposed to solve</td>
</tr>
</tbody>
</table>

Part B: Coding Scheme for Types of Support
Appendix 3.3

Sample Essays by Scores on Argumentation Complexity Scale (ACS)

1 Point: [ID: 2C50904020009; Female; 5th Grade]

Students should had iPads in school [Own Claim] so they can learn and look at your teacher so you can know what to look up on the Internet [Own Support]. They can help you with your projects if you need help with it [Own Support]. And you can show your teacher [Own Support]. I think that iPads is great in school for a reason [Own Claim]. It can be good for students to learn better and help you [Own Support].

2 points: [ID: 2C51305010007; Male; 6th Grade]

I think taking the Ipads away is a bad idea [Own Claim]. I think is a bad idea [Own Claim] because you can get through stuff faster [Own Support]. A reason I think is bad to take the iPads is that we do not have to go to a computer lab [Own Support]. Another reason is that we can go on websites and learn more [Own Support]. My last reason that you learn about more stuff like practice [Own Support]. To solve the problem of the iPads is that people should block the bad websites out [Solution Support]. They should give kids a lot of trouble if they do something bad [Solution Support].
3 points [ID: 2C20106990029; Male; 7th Grade]

*I think Ipads should be allowed to be in school [Own Claim]. I think that because first people might think that the kids are safe from bullying but they are not [Critique Support]. The bullies can still bully people face to face [Critique Support]. Second of all the Ipads have helped us improve our grades [Own Support]. For example if you forget you homework use Edmodo and ask your teacher [Own Support]. Finally it is not like people are going to look at porn or other bad things like Facebook et cetera [Critique Support]. Just block those websites so people will not use them [Solution Support].

4 points [ID: C20106040010; Male; 8th Grade]

*Ipads should not be banned from school [Own Claim]. Many people think that the Ipads are a waste of time a distraction or even a tag [Counter Support]. The Ipads are tools and should only be used as tools [Own Support]. A problem students are facing is getting distracted by online games videos or music all the students get so involved in all of these things [Counter Support]. But I do not think taking them away is the answer [Own Claim]. Teachers can block websites and control when the Ipads can be out [Solution Support]. Ipads can be very helpful in school [Own Support]. If a student were to have a school project like a Powerpoint they could easily work on said Powerpoint at home or at school [Own Support]. They can also be used for research homework or emailing your teacher and other school
related activities [Own Support]. If the principal were to take away the Ipads the productivity of students would decrease greatly [Critique Support]. It would be harder for students to do research projects and homework [Critique Support]. Therefore I believe students should have Ipads in school but should be limited to what activities they decide to do [Mitigated Claim].

5 points [ID: C20106020011; Female; 8th Grade]

I believe all students should have the opportunity to use Ipads while in school [Own Claim]. Without these electronic devices some students may have trouble finding access to other electronic devices in order to complete homework and school work [Critique Support]. I also believe that Ipads will be beneficial in a classroom because it will let students research topics and that research may be needed for an in school project [Own Support]. These are some of the pros to having access to Ipads during school [Own Support]. Although there are many pros to the Ipads there are also a few cons [Counter Claim]. Such as the Ipad being distracting for students [Counter Support]. The students with Ipads may be playing games or looking things up on the Internet that has nothing to do with the classroom topic [Counter Support]. Also the students may be posting harsh comment geared towards other students on social media websites during class or at home [Counter Support]. All of these problem can be fixed easily [Solution Support]. To cut down on the number of students playing games during class just
have the student keep the Ipads off and in their bags or under their seat until the teacher instructs them to take them out and use the Ipad for a certain purpose [Solution Support]. This reduce the number of cruel comments being posted during class. Another solution to stop mean comments from going viral at home is to have the students return the Ipads to a cart at the end of the day and receive them again in the morning [Solution Support]. Overall there are pros and cons to classroom Ipads. But the cons can be fixed with simple rules [Mitigated Claim]. This is why I believe the Ipads are an asset to the classroom [Own Claim].

Same 5-point Essay Coded under the Structural Element Approach and Perspective Element Approach

Structural Element Approach:
I believe all students should have the opportunity to use Ipads while in school [Claim]. Without these electronic devices some students may have trouble finding access to other electronic devices in order to complete homework and school work [Ground]. I also believe that Ipads will be beneficial in a classroom because it will let students research topics and that research may be needed for an in school project [Ground]. These are some of the pros to having access to Ipads during school [Ground]. Although there are many pros to the Ipads there are also a few cons [Claim]. Such as the Ipad being distracting for students [Ground]. The
students with Ipads may be playing games or looking things up on the Internet that has nothing to do with the classroom topic [Ground]. Also the students may be posting harsh comment geared towards other students on social media websites during class or at home [Ground]. All of these problem can be fixed easily [Ground]. To cut down on the number of students playing games during class just have the student keep the Ipads off and in their bags or under their seat until the teacher instructs them to take them out and use the Ipad for a certain purpose [Ground]. This reduce the number of cruel comments being posted during class. Another solution to stop mean comments from going viral at home is to have the students return the Ipads to a cart at the end of the day and receive them again in the morning [Ground]. Overall there are pros and cons to classroom Ipads. But the cons can be fixed with simple rules [Claim]. This is why I believe the Ipads are an asset to the classroom [Claim].

Perspective Element Approach:

I believe all students should have the opportunity to use Ipads while in school [Own-Side Only Perspective]. Without these electronic devices some students may have trouble finding access to other electronic devices in order to complete homework and school work [Dual Perspective]. I also believe that Ipads will be beneficial in a classroom because it will let students research topics and that research may be needed for an in school project [Own-Side Only Perspective].
These are some of the pros to having access to Ipads during school [Own-Side Only Perspective]. Although there are many pros to the Ipads there are also a few cons [Integrated Perspective]. Such as the Ipad being distracting for students [Integrated Perspective]. The students with Ipads may be playing games or looking things up on the Internet that has nothing to do with the classroom topic [Integrated Perspective]. Also the students may be posting harsh comment geared towards other students on social media websites during class or at home [Integrated Perspective]. All of these problem can be fixed easily [Dual Perspective]. To cut down on the number of students playing games during class just have the student keep the Ipads off and in their bags or under their seat until the teacher instructs them to take them out and use the Ipad for a certain purpose [Dual Perspective]. This reduce the number of cruel comments being posted during class. Another solution to stop mean comments from going viral at home is to have the students return the Ipads to a cart at the end of the day and receive them again in the morning [Dual Perspective]. Overall there are pros and cons to classroom Ipads. But the cons can be fixed with simple rules [Dual Perspective]. This is why I believe the Ipads are an asset to the classroom [Own-Side Only Perspective].
 References


Mahwah, NJ: Lawrence Erlbaum Associates.


Thesis Conclusion

In this thesis, I conducted three studies focused on the linguistic domains of argumentative writing: vocabulary, syntax, and discourse. In each study, I developed a new approach to conceptualize and measure a domain using quantitative, qualitative, or mixed methods, and provided evidence for validating my new approach.

In Study 1, I specified and examined a measurement model for vocabulary performance in fifth-to-eighth grade argumentative writing. The measurement model confirmed that lexical diversity, lexical density, lexical rarity, lexical specificity, and academic vocabulary jointly indicated a common underlying construct Vocabulary in Writing (VW). VW was found to be positively, and moderately associated with the holistic writing quality. The association was stronger than that between each individual indicator and the writing quality. The VW factor scores was found to display developmental trends from fifth to eighth grade, such that students in later grades tended to display higher VW scores.

In Study 2, I developed a novel measure of syntactic performance, Diversity of Advanced Syntactic Structures (DASS) score. DASS is calculated as the total types of a set of syntactic structures which identified as representative of academic language skill expectations for adolescents. The set includes: adverbial clause, clausal complement, clausal prepositional complement, relative clause as modifier, clausal subject, noun as modifier, and passive voice. DASS was significantly and positively associated with essays’ writing quality and students’ receptive academic
language skills, even after Mean Length of Clauses, a conventional syntactic complexity measure, was controlled for. DASS was also found to display developmental trends, in particular students in fifth grade displayed significantly lower DASS scores than students in seventh and eighth grade.

In Study 3, I developed a novel coding scheme that identified elements in the argumentative discourse: Own Claim, Own Support; Mitigated Claim, Solution Support, Critique Support; Counter Claim, Counter Support. I found that it was easier for young adolescents to generate claims than to generate supports when advancing their Own Argument, whereas it was easier for them to generate supports than to generate claims when engaging implicitly or explicitly with the opposing position, that is, when advancing Mitigated or Counter argument. Based on the complexity gradients identified by the coding scheme, I generated the Argumentation Complexity Scale (ACS). Similar to the VW and the DASS indices, the ACS displayed developmental trends in that eighth graders scored significantly higher in argumentative discourse performance at eighth grade than fifth, sixth, and seventh graders. Students’ scores on ACS were found to be significantly, positively, and moderately associated with essays’ writing quality and with students’ receptive academic language skills.

My thesis contributes to the body of language and literacy education research, specifically on adolescent writing, by providing a set of novel measures for the measuring the linguistic and argumentative features (i.e., vocabulary, syntax, and discourse) of adolescents’ written production. In each study, I took a
bottom-up approach in developing the new measurement tool. In other words, I first identified fine-grained characteristics in a linguistic domain, and then used quantitative, qualitative, or mixed methods to integrate these characteristics in order to construct a global index for this domain. This measurement approach is in contrast with the more widely adopted approach in current adolescent literacy research, which usually uses omnibus measures or broad dimensions as part of holistic rubrics to describe and evaluate students’ written products. The novel measures’ sensitivity to between-grade differences and significant associations with the traditionally scored writing quality offers robust evidence in support of their validity.

Overall, the three studies reveal the multifaceted nature of vocabulary, syntactic, and discourse performances that are only captured broadly and vaguely through holistic scoring. Besides offering a promising complementary set of measures to existing widely used approaches in research, these novel indices have a few advantages for education practice. The findings of these studies may shed light on the more specific delineation of learning objectives for writing pedagogy in standards, assessment criteria, and instructional practices. The new set of measures provides more detailed and quantifiable descriptions of students’ written texts. The automated linguistic analyses, especially for the domains of vocabulary and syntax, suggest their possible application in large-scale assessments. Admittedly, due to its modeling intricacy and coding complexity, the three measurement approaches pose challenges for practitioners to directly implement
them and interpret the scores. However, they open an opportunity for a promising field at the nexus of research and practice, where the work could be outsourced by teachers, schools, or districts to a group of liaison staff who provide a service package of data analysis and result interpretation. With this information, teachers can potentially conduct efficient diagnostics of students’ writing proficiency, and in turn design more targeted, individualized instruction. Future research may examine the relationships between the linguistic domains and how the domains jointly construct the overall language proficiency exhibited in students’ written production.