7-13-2017

Analysis of Syntactic Complexity and Its Relationship to Writing Quality in Argumentative Essays

Thilagha Jagaiah
University of Connecticut - Storrs, thilagha.jagaiah@uconn.edu

Analysis of Syntactic Complexity and Its Relationship to Writing Quality in Argumentative Essays

Thilagha Jagaiah, Ph.D.

University of Connecticut, 2017

Syntactic complexity has been recognized as an important construct in writing by numerous previous studies. However, there was no consensus on the precise and salient syntactic complexity measures (SCMs) to examine syntactic complexity. This is because most previous studies examined SCMs manually using a small sample size with few SCMs. In the current study, the author seeks to address these gaps using Confirmatory Factor Analysis (CFA) to test a hypothesized model of 28 SCMs and four latent variables (Sentence Pattern, Sentence Length, Sentence Connector, Sentence Sophistication). The data was analyzed using 1,029 eighth-grade, argumentative essays that were scored using an automated text analysis tool, Coh-Metrix, version 3.0. A refinement of the hypothesized model using 16 SCMs and the same four latent variables produced a good fit using CFA. The four latent variables were then used as input predictor variables together with a student-type indicator variable to examine the relationship with writing quality as reflected in writing scores of the eighth-grade, automatically scored formative assessment data for writing. A multiple linear regression (MLR) model was used to examine this relationship, and the findings indicated a modest positive relationship between each of the four latent variables and writing quality. Furthermore, this relationship varied significantly between at-risk and not-at-risk student type with increased use of the four latent variables having a greater impact on writing quality for at-risk students compared to not-at-risk students. The findings of this study will have important implications for methodology, writing assessment, and writing instructions on sentence-construction skills.

Analysis of Syntactic Complexity and Its Relationship to Writing Quality in Argumentative Essays

Thilagha Jagaiah

B.A., National University of Malaysia, 2001

M.A., National University of Malaysia, 2006

M.A., University of Connecticut, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Analysis of Syntactic Complexity and Its Relationship to Writing Quality in Argumentative Essays

Presented by

Thilagha Jagaiah, B.A., M.A.

Major Advisor

____

Natalie G. Olinghouse

Associate Advisor

____

Devin M. Kearns

Associate Advisor

____

Gilbert N. Andrada

University of Connecticut

2017

ACKNOWLEDGMENTS

The completion of this dissertation and all the work leading up to it (and, beyond) would have not have been possible without the phenomenal and consistent support from many.

I express my sincere gratitude to Dr. Natalie Olinghouse, my major advisor. She has provided me with an excellent opportunity to work with her and supported me all the way from when I first considered to apply to the Ph.D. program in Special Education, through completion of this degree. Her mentorship was paramount in providing me a well-rounded experience consistent with my long-term career goals. She encouraged me to not only grow as a researcher but also as an instructor and an independent thinker. Her invaluable advice has not only helped me to grow professionally but it also contributed to my growth as an individual. I am very grateful to her for showing confidence in my abilities and providing me the opportunity and guidance which allowed me to develop my dissertation project. Going forward, I strive to continue to uphold the same quality of work as she does, and I am thankful for the high standards to which she holds me.

I would like to thank my co-advisor, Dr. Devin Kearns, for his astute guidance, insightful comments, and encouragement throughout my doctoral program. I appreciate the time he spent on discussion of my dissertation project given his busy schedule. I am truly fortunate to have had the opportunity to work with him for he shared his vast experiences in both research and classroom instruction that have been instrumental in the success of this dissertation project. He has also nurtured my scholarly and professional identities over the course of my tenure as a doctoral student. I appreciate, as a mentor, how he has molded me into a professional in special education. I benefitted greatly from his expertise.

I am also grateful to Dr. Gilbert Andrada for giving generously of his time especially on weekends. His friendly guidance and detailed, insightful suggestions have been instrumental to the completing of dissertation. His thought-provoking questions and critical input have been vital to the progress of my research work. He has always been very approachable and eager to resolve and clarify any doubts I have. His general collegiality that he offered to me has been a source of inspiration in my professional growth. Working with him has been a great learning experience, and I consider myself fortunate for having this opportunity.

I offer sincerest thanks to Dr. Michael Coyne and Dr. Joshua Wilson for serving as readers for both my dissertation proposal and dissertation manuscript. Your suggestions and recommendations improved the quality of my dissertation. Special thanks to Dr. Wilson for guiding, encouraging, and nurturing me at countless points along the way throughout my doctoral program.

A very special gratitude goes to PARSACT and Dr. Hariharan Swaminathan as research partners who provided me access to the data used in this study. I am especially grateful to Dr. Swaminathan for all his extensive professional guidance, unfailing support, and assistance he provided me on research methods.

In a similar vein, I would like to recognize Dr. Chen Ming-Hui, Dr. Shyamala Nagaraj (University of Michigan), Yan Zhuang, and Jialin Han for their recommendations and assistance that guided the data analysis in my dissertation. Their involvement and constructive feedback helped me gain a better understanding of the complex statistical analysis required for my dissertation.

I would like to express my gratitude to graduate research assistants Jue Zhang, Shengyao Tang, Tongan Liu, and Yan Hang Wang, for their quick, accurate, and competent work using

Coh-Metrix to prepare the data for analysis. I thank them for their hard work and dedication in keeping to my deadlines because that allowed me to analyze the data in a timely manner.

I thank profusely to Dr. Gao Niu for all the great help he has given me throughout my doctoral program and my dissertation process. He is my best go-to person for anything related to technology, and I truly appreciate his patience in walking me through all the complex tables and diagrams I had to create for my dissertation. Because he has responded promptly and competently to every little and big request, he has made my dissertation experience so much more wonderful.

I express my deepest and heartfelt gratitude to my parents, brother, and Pam for supporting me spiritually and my life in general. They were always encouraging me with their best wishes.

A special thanks to Maya and Steve for their constant support and encouragement throughout my doctoral program. Their genuine interest in seeing me succeed has inspired me to strive to do my best and to continue the academic tradition in the family.

I would also like to thank my friends in the doctoral program for their time, encouragement, and support. On some occasions, it was words of wisdom from them that gave me the motivation to keep going. Deliberating over problems and happily talking about things other than research papers are some good memories I will cherish.

Finally, and most importantly, a tremendous thank you to my wonderfully supportive husband, Jay, for his unwavering source of inspiration, support, encouragement, and quiet patience throughout the dissertation journey. He has been a constant foundation of strength and support that has kept me going at my weakest moments. Through tears and laughter, he provided

the confidence I needed to push through the hardest moments. I would also like to acknowledge the innumerable sacrifices he made to ensure my success. Thank you for being there through it all and for understanding how to be the kind of partner and friend I need most. I could not have done this without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background of the Problem

Writing skills are central to academic achievement, to graduating from college, to gaining employment, and to communicating effectively. The single best predictor for college success and failure is one's ability to compose an extended text (ACT, 2005; Geiser & Studley, 2001; Noeth & Kobrin, 2007). Prospective employers select qualified candidates with proficient writing skills for both employment and promotions (National Commission on Writing in America's Schools and Colleges, 2005). Lack of writing proficiency not only affects employment opportunities but also involves a societal cost. The National Commission on Writing in America's Schools and Colleges (2004) reported that improving writing skills for hired employees requires the most remedial training, and industries spend an annual $3.1 billion to improve their employees' skills.

Writers of successful texts exhibit domain, discourse, and linguistic knowledge (Flower & Hayes, 1980). Domain knowledge enables writers to produce relevant ideas (Graham & Perry, 1993; Kellogg, 1987) that improve writing quality (Ericsson, Charness, Feltovich, & Hoffman, 2006; Simon & Chase, 1973). Discourse knowledge in relation to genres (e.g., narrative, descriptive, argumentative) provides students strategies on how to produce better-structured texts (Bereiter & Scardamalia, 1987). As noted by Applebee et al., (1990), writers who are linguistically proficient are able to produce grammatically correct complex sentence structures with appropriate punctuation, varied use of vocabulary, and accurate spelling. Writers who have domain, discourse, and linguistic knowledge can produce higher-quality texts

by generating and organizing ideas, using their knowledge to better revise their texts without over-imposing on their cognitive resources (Deane, 2013; Saddler & Graham 2005).

Writing skills are complex, and many students struggle with learning to write due to the multi-component nature of writing. To produce quality texts, writers have to be skillful in both higher-order skills (planning, drafting, revising, and editing) and lower-level skills (handwriting, spelling, vocabulary, and sentence construction; Hayes, 1996; Saddler & Graham, 2005). While all writing requires conscious effort and a great deal of practice in composing, developing and analyzing ideas, proficient and effective writing also hinges on the ability to craft formal, and well-constructed sentence structures. Sentence construction is not only a lower-level skill, but also a foundational skill that students acquire at lower elementary-grade levels. Lower-level skills are crucial building blocks of writing skills to successfully engage in higher-order skills. More specifically, lack of mastery in constructing syntactically complex sentences may hinder students' abilities to effectively translate thoughts and ideas into writing (Graham, 2006; Scardamalia & Bereiter, 1986; Strong, 1986). Knowing how to plan, for example, has ultimately little value if the writer is unable to construct effective sentences. Poor sentence-construction skills can be a serious inhibitor to successful writing (Saddler & Graham, 2005). While lack of sentence-construction skills impedes successful growth in writing for skilled writers, it is a significantly greater challenge for less-skilled writers in constructing simple sentence structures, let alone syntactically complex structures.

Mastery in both higher-order and lower-level skills enables skilled interaction between their writers' cognitive resources, the instructional context, and the demands of the writing task to produce a high-quality written text. Less skilled writers, on the contrary, lack automaticity in lower-level skills, which inhibits their ability to use higher-order skills to compose a text

(Saddler & Graham, 2005). This occurs because students have to concentrate on crafting sentences instead of focusing on the writing process (Graham, 2006; Scardamalia & Bereiter, 1986; Strong, 1986). Lack of proficiency in constructing sentences at varied levels of complexity causes some students to avoid writing or to give up writing altogether, further decelerating their writing development (Berninger, Mizokawa, & Bragg, 1991; Berninger & Swanson, 1994; McCutchen, 1996).

Development of lower-level skills facilitates the development of higher-order skills; sentence construction in particular is a lower-level skill that is critical for all grade levels. Specifically, it enables students at higher grade levels to express more complex thought processes in writing. Complex thought processes can be translated into coherent, cohesive, and well-argued text by using syntactically complex sentence structures. The ability to construct syntactically complex sentences at higher grades enables students to manipulate varied sentence structures to create different stylistic effects. Learning to construct complex sentences is an essential skill for all writers, but not all writers develop this skill at the same pace. For struggling writers and students with learning disabilities, this process may develop very slowly, and the ability to construct syntactically complex sentences may be one of the main differences between successful and struggling writers.

Consonant with this evidence on the importance of writing, the 2011 National Assessment of Educational Progress (NAEP) data clearly show that many students struggle with this important skill (NCES, 2012). The ability to construct varied sentence structures with varying levels of complexity was also used as a scoring criterion to determine a student's level of competency. The data show that only 24% of typically achieving eighth and twelfth graders, and 5% of 8th and 12th graders with learning disabilities wrote proficiently (NCES, 2012).

Overall, about one half of the eighth and twelfth graders perform at the basic level. In terms of race categories, about 89% of eighth-grade African American students, 86% of Hispanic students, 66% of White students, and 88% of the students eligible for the National School Lunch Program have not reached grade-level proficiency. The data suggest that while writing well is a challenge for skilled writers, it is even more so for less-skilled writers. This points to a need to better the understand writing skills that enhance students' writing quality.

Additionally, increased emphasis on sentence-level components of text complexity in the Common Core State Standards (CCSS) makes it essential that students write varied and more complex sentence structures. Writers must now construct sophisticated and complex sentences even in the early elementary grades to meet the CCSS for writing. This is a difficult task for many students because as researchers have found syntactic difficulties are a core feature in the profiles of many struggling writers and writers with learning disabilities (LD) (Houck & Billingsley, 1989; Kline, Schumaker, & Deshler, 1991; Morris & Crump, 1982; Myklebust, 1973; Newcomer & Barenbaum, 1991; Saddler & Graham, 2005). Despite the need to develop better sentence-construction skills, recent research studies have not paid as much attention to this; instead they have focused on higher-order skills such writing process (Graham & Harris, 2003, 2005; Graham, MacArthur, & Fitzgerald, 2013; Harris, Graham, Mason, & Friedlander, 2008; Myhill & Jones, 2009). The CCSS has raised writing standards by requiring students to construct syntactically complex sentences, which has led to increased attention on the relationship between syntactic complexity and writing quality. As mentioned earlier, if students do not have mastery in sentence construction skills, besides not meeting CCSS requirements, it is difficult for them to articulate increasingly complex ideas with clarity and confidence. Not being able to do this could impede performance in higher grade levels, postsecondary education,

and the workforce environment. To state it differently, students will not be college- and career -ready.

Sentence construction, along with planning, drafting, and revising, is a critical component of the writing process. Because a writer's ability to construct sentences is related to the working memory resources, constructing syntactically more complex sentences requires more effort from the writer. Lack of knowledge of complex sentence structures at the sentence level hinders a writer's ability to translate thoughts and ideas into dynamic sentences (Hayes & Flower, 1986). Therefore, struggling writers write simple sentences that provide information in its basic form without connecting or completing their thoughts. This in turn constrains other composing processes and produces similar structures throughout the text with limited variations to hold reader's interest (Morris & Crump, 1982; Mykelbust, 1973; Newcomer & Barenbaum, 1991).

A syntactically complex structure helps the writer convey ideas that tie together, sum up a series of thoughts, qualify a previous point, and transition between ideas to convey meaning effectively. For example, lack of syntactical complexity produces the following:
John is always punctual to school. John woke up late this morning. John was late for school (S1), whereas skill with syntactical complexity produces a more pleasing flow in the following sentence:
John, who is always punctual to school, woke up late this morning, and he was late for school (S2).
When ideas are presented as in the simple sentence (S1), relations between John being punctual to school and John waking up late this morning are unclear, and the individual sentences lack cohesion because they do not make references to the relations between these events. It is not

known that John waking up late was a one-time occurrence that cause him to be late for school. Each simple sentence conveys separate ideas, and the reader has to make the connections between them. Some readers may be able to make the connections due to already embedded knowledge while others may not, due to lack of familiarity with the events, and this impedes comprehension. These sentences lack connectives such as relative pronouns (who) and conjunctions (and) that contribute to cohesion by explicitly linking ideas at the clausal and sentence level (Halliday & Hasan, 1976; McNamara & Kintsch, 1996) as shown in the following sentence (S2). If students know how to construct sentences by connecting clauses and phrases, they are able to embed and lengthen sentences, which not only creates a complex structure, but also reduces the burden on cognitive resources of the interpreter. Sentence (S2) makes clear connections and reference between the subject (John) and the predicate (information after the subject that includes the verb). This complex sentence structure makes connections for the reader and conveys meaning effectively. According to Freedman (1979), if raters cannot decipher the connections, they may award a lower grade for an essay comprising several short, simple sentences. It is essential for students to have mastery in constructing varied sentence structures, including sentences that are syntactically complex, to produce quality texts.

Statement of the Problem

Numerous syntactic complexity measures (SCMs) have been proposed in various studies to examine writing development and fluency. Typically, the SCMs that have been examined quantified one or more of the following: length (e.g., mean T-unit length, sentence length, clause length), number of subordination or coordination (e.g., dependent clause, independent clause), types of syntactic structures (e.g., phrases, clauses), and sophisticated syntactic

structures (e.g., compound, subject and verb sentence pattern). Findings from these studies have important theoretical, practical, and educational implications. However, the validity of these results hinges upon three crucial factors. One is the validity of the SCMs or scales used to obtain these results; the other two are the size and representativeness of the writing samples analyzed. Human rating of syntactic complexity of large language samples is an extremely laborious process, requiring skilled raters to identify a range of relevant SCMs in the writing samples. This has posed a major challenge to researchers in the search for the most valid SCMs and the application of these SCMs to large writing samples. There is a clear need for text analysis tools that can automate the process with accuracy.

Syntactic complexity has been recognized as an important construct in writing by numerous studies in the past (see Jagaiah, 2016). In her systematic review, Jagaiah (2016) found at least 52 SCMs to examine syntactic complexity. Although researchers have assessed various SCMs, there is no consensus on which SCMs are appropriate measures of syntactic complexity.

Syntactic complexity is an abstract concept that cannot be defined or measured precisely. Therefore, researchers have used SCMs to characterize it. However, for an SCM to be considered an appropriate measure of syntactic complexity, it should show varying patterns by grade levels, student writing ability, and genre, or have an impact on writing quality. One reason previous studies were unable to find any consistent pattern with the SCMs that were examined was because the sample size and the number of measures examined in each study were small and varied from study to study. In addition, the various SCMs were defined differently in these studies, making it difficult to compare the results and to identify consistent patterns of interest. Furthermore, similar SCMs used in different studies produced inconsistent

results, in particular, for mean number of words per T-unit (T-unit length; see Hunt, 1970; Crowhurst, 1980a; Crowhurst, 1980b; Morris and Crump, 1982; Evans, 1979; Wagner et al., 2011). Consequently, it was difficult in the past for researchers and educators to decide on the best SCMs to reflect syntactic complexity.

It should be noted that few studies have examined the relationship between SCMs and writing quality (Jagaiah, 2016). Findings from these studies did not show consistent results (see Beers & Nagy, 2009; Crowhurst 1980a; Stewart & Grobe, 1979). Jagaiah (2016) found inconsistent relationships between syntactic complexity and writing quality by grade levels, genres, and SCMs, and this could have been a result of small sample sizes analyzed. Furthermore, no studies examined the relationship between syntactic complexity and writing quality based on students' writing abilities.

Previous studies have not attempted to simultaneously analyze several SCMs or group the myriad of SCMs into meaningful categories. One major challenge for past researchers was the lack of an automated text analysis tool to examine syntactic complexity. The labor-intensive task of a manual analysis made it difficult to search for the most valid SCMs. Consequently, most studies examined very few measures with a relatively small sample size (see Beers & Nagy, 2009; Belanger & Martin, 1984; Grobe, 1981; Stewart & Grobe, 1979). Additionally, skilled evaluators were required to identify and calculate the relevant SCMs in the writing samples as well as ensure high interrater reliability. In particular, only a few studies analyzed composite SCMs (e.g., syntactic density score; see Blair & Crump, 1984; Kagan, 1980; Morris & Crump, 1982) because this was more complex and more prone to error. There is a clear need to use automated text analysis tools such as Coh-Metrix that can automate the

process of analyzing large amounts of data to estimate numerous SCMs, including composite measures, with high accuracy and reduced interrater reliability issues.

To examine syntactic complexity holistically, it is important that the various SCMs that have been examined thus far be analyzed as groups of related SCMs instead of individual SCMs. Linguistic theory could provide guidance on how to create these groups of related SCMs. It would be easier to explain syntactic complexity to educators by analyzing a few groups of related SCMs rather than several individual SCMs. Using this information, educators can incorporate sentence-construction skills related to syntactic complexity in writing instruction and assessment.

The current study overcame the limitations of previous studies by (a) using Coh-Metrix, a reliable automated text analysis tool that has the ability to capture numerous, well-established individual and composite syntactic complexity measures in an automated manner; (b) using a large data set and simultaneously analyzing several SCMs; and (c) understanding the relationship between these SCMs and students' writing ability for a given grade level and genre.

Theoretical Framework

Syntactic theory is the theoretical framework that underlies the construction of syntactically complex sentences. Syntactic theory explains how a sentence is composed of constituents whether at the level of the word, phrase, clause, or sentence. These constituents are combined and arranged in grammatical ways to form potentially infinite sets of simple or complex sentences (Chomsky, 1957; Givon, 2009). As more phrases are embedded to the words, they form hierarchical structures (see Figure 1). Constituency and hierarchical structures make sentences become more complex. A sentence made up of several constituents is a resilient unit with no syntactic limits to its length or complexity once the minimal requirements

of subject and predicate have been met (Markels, 1984). For example, a minimal sentence such as *Mary laughed* contains a subject and a predicate which form the building block of sentences known as a clause. One way to increase complexity is to replace the subject and predicate with phrases of varying levels of complexity (Phillips, 2006). For example, *Mary, a quiet little girl, laughed loudly* will now be considered a syntactically complex sentence because the embedded structure (a quiet little girl) and the adverb (loudly) provide additional information that contained in the previous sentence *Mary laughed.*

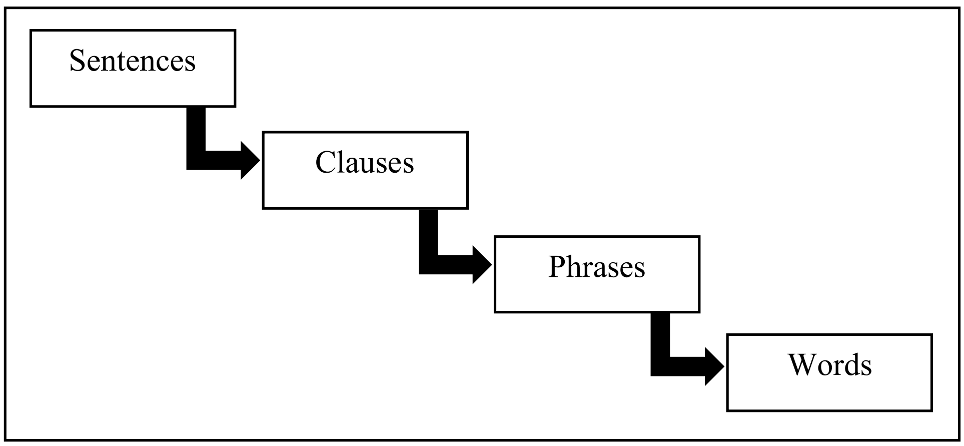

Figure 1. Hierarchical Structure of a Sentence.

Tree structure representation of syntactic theory. Syntactic complexity can be represented using a hierarchical tree structure as shown in Figure 2. The root of the tree is at the highest level, and it is the main sentence constituent or node. Represented by the symbol S, it has descending branch roots that point to its two constituents or phrases: a subject noun phrase (NP)[*Active children*] and a predicate or verb phrase (VP) [*like bright colors*]. These phrases are also nodes at the intermediate structural level. There may be many structural levels at the intermediate nodes. For example, the subject NP contains a noun (N) [*children*], and an adjective (ADJ)[*Active*]. Similarly, the VP contains a verb (V)[*like*], and an object NP [*bright colors*]. The object NP is further broken down into two individual nodes: an adjective

(ADJ)[*bright*] and a noun (N) [*colors*].   Figure 2 shows the representation of a three-level hierarchical structure of embedded constituents. The relations between the constituents are the connections within the nodes that form the hierarchical levels of complexity (Chomsky, 1957). As illustrated in Figure 2, a sentence comprises various levels of hierarchy that define whether it is simple or complex.

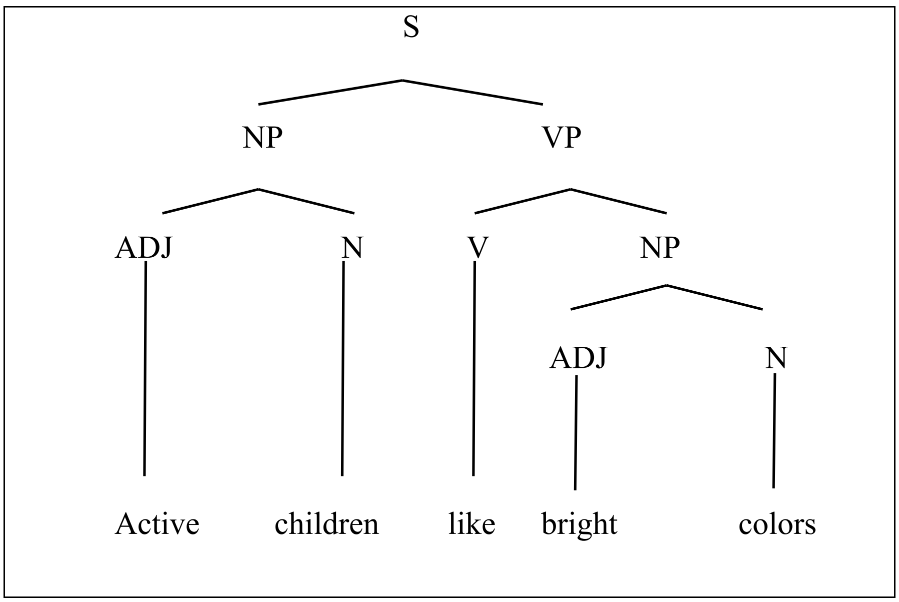

Figure 2. Two-constituent model of a sentence illustrating sentence components that contribute to the complexity of each constituent. S = Root of the tree; NP = Noun Phrase; VP = Verb Phrase; ADJ = Adjective; N = Noun; V = Verb.

Sentences with complex structures that comprise constituents of higher levels of complexity include conjunctions, clauses, and embedded clauses. Additionally, embedding clauses inside other clauses increases the syntactic complexity. The two most common types of such embedding are relative clauses in the noun phrase and verbal complements in the verb phrase. For example, the tree diagram in Figure 3 shows the embedding in the Noun Phrase (REL-clause). The main clause *Children are happy* has two hierarchical levels: NP (*Children*) and VP (*are happy*). However, when a relative clause is embedded, the number of hierarchical

levels increases to five, thus increasing the complexity level of the sentence. The second hierarchical level is the relative clause (REL) (*who like bright colors*). The third hierarchical level is constructed with a VP (*like bright colors*) and is followed by an adjective phrase that represents the fourth hierarchical level (*bright colors*). The fifth hierarchical level is represented by the noun phrase (*color)*]. To convey interrelationship of ideas used in higher levels of abstraction, writers employ even more complex structures such as subordinate clauses, which are a type of embedded structure.

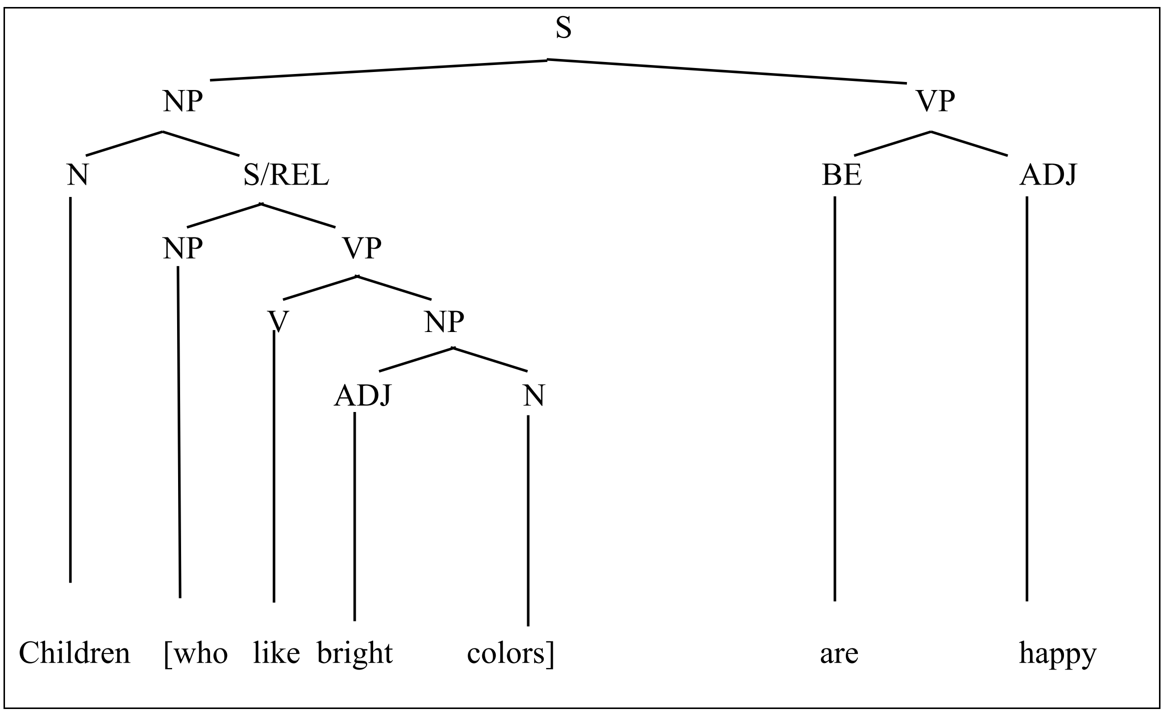

Figure 3. Two constituent hierarchical levels of a sentence illustrating sentence components that contribute to the complexity of each embedded clause. S = Root of the tree; NP = Noun Phrase; VP = Verb Phrase; N = Noun; S/REL = Relative Clause; BE = Auxiliary Verb; ADJ = Adjective; V = Verb.

Syntactic theory approaches will be used to examine the SCMs in relation to sentences, clauses, phrases, and words.

Purpose

The purpose of this study is to examine the fit of the hypothesized model based on 28 Coh-Metrix SCMs as indicators of four latent variables (*Sentence Pattern, Sentence Length, Sentence Connector, and Sentence Sophistication*). The hypothesized model was tested using the eighth-grade, automatically scored formative assessment data for writing. A multiple linear regression (MLR) model was developed to examine if the four latent variables with the associated Coh-Metrix SCMs confirmed by the Confirmatory Factor Analysis (CFA) showed a relationship with writing quality, and whether they varied between at-risk and non-at-risk eighth-grade students.

This study is unique because it tests a hypothesized model of four latent variables and 28 SCMs using CFA. The results from the MLR model could be used in future studies to examine the relationship between syntactic complexity and writing quality for different genres and grade levels.

Research Questions

The following research questions and hypotheses guide this study:

(RQ1) Is the hypothesized model based on 28 Coh-Metrix SCMs as indicators of four latent variables a good fit using the eighth-grade, automatically scored formative assessment data for argumentative writing*?* The four latent variables are Sentence Pattern, Sentence Length, Sentence Connector, and Sentence Sophistication.

(H1) The hypothesized model is a good fit for the eighth-grade automatically, scored formative assessment data for argumentative writing.

(RQ2) Do the scores of the four latent variables based on the 28 Coh-Metrix SCMs show a relationship with writing quality, and how does this relationship vary between at-risk and not-at-

risk students using the eighth-grade, automatically scored formative assessment data for writing?

(H2) The scores of the four latent variables based on the 28 Coh-Metrix SCMs show a relationship with writing quality and the relationship vary between at risk and not-at-risk students using the eighth grade automatically scored formative assessment data for writing.

Significance of the Study

In the search for appropriate SCMs, this study is beneficial for educators, students, and researchers. First, the study delineates important SCM categories as indicated by the four latent variables of Sentence Pattern, Sentence Length, Sentence Connector, and Sentence Sophistication. The findings of this study could become the basis for a follow-up intervention study to accomplish the following: (a) developing a practical translation of these four latent variables for instructors to use when teaching students sentence-construction skills, (b) developing rubrics to assess sentences where these four latent variables could be used as descriptors in the rubrics, and (c) incorporating the relevant latent variables in students' writing checklists. The follow-up intervention study could show that appropriate use of the latent variables in sentences may improve writing quality of texts produced by the students. If the follow-up intervention study shows encouraging findings, future researchers would be able to replicate and extend this study to include other grade levels and genres or other SCMs or latent variables that have not been included in this study.

Definition of Key Terms

Syntactic complexity. A sentence structure that connects pieces of information effectively and efficiently using sentence components with varying levels of hierarchy (Jagaiah, 2016).

Syntactic complexity measures (SCMs). Measurable sentence elements (e.g., sentence length, clause length, number of clauses, number of phrases) that are used to operationalize the construct of syntactic complexity.

At-risk. The 1992 National Center for Education Statistics (NCES) report defines the characteristics of at-risk eighth-grade students as failure to achieve proficiency in basic skills before high school graduation. These students are struggling writers who are likely to fail at school or drop out of school (Kaufman, Bradbury, & Owings, 1992). Consonant with the NCES report, the Response to Intervention (RTI) model defines students who do not achieve proficiency or do not meet benchmarks as being at some risk for academic failure (Fuchs & Fuchs, 2006).

Writing ability. Writing ability refers to the ability to navigate multiple aspects of the writing process including setting goals for writing, generating and organizing ideas, transforming ideas into varied sentence structures and transcribing these sentence structures, revising and editing text, and composing a full text. Writing ability also comprises mastery of both higher-order (planning, drafting, revising) and lower-level (spelling, handwriting, sentence construction, vocabulary) skills necessary for proficient or grade-level-appropriate writing.

Coh-Metrix. Coh-Metrix is an automated text analysis that compiles a number of computational linguistic measures. The current version, Coh-Metrix 3.0, which is available for public use over the Internet, includes 106 measures. Coh-Metrix can be freely accessed at www.cohmetrix.com. The indices are classified into eleven groups: descriptives, text easability principal components scores, referential cohesion, latent semantic analysis, lexical diversity, connectives, situation model, syntactic complexity, syntactic pattern density; word information; and readability.

Sentence. The Coh-Metrix analysis defines a sentence as a group of words that begins with the first word of a sentence (including sentence fragments) and is punctuated with an end punctuation mark, including a period, exclamation mark, or a question mark.

Main clause. A main or independent clause is a complete sentence that has at least a subject and a verb.

Dependent clause. A dependent clause has a subject and a verb, but it cannot stand on its own. The dependent clause provides additional information to the main clause.

CHAPTER 2

REVIEW OF LITERATURE

The evaluation criteria for current writing research have shifted from grammatical accuracy as the sole basis for grading to content, organization, style, vocabulary, and grammar (Schultz, 1994), increasing the importance of teaching the writing process. Current writing classroom practices devote considerable time to teaching students varied aspects of the writing process (planning, drafting, revising, editing; Beers & Nagy, 2009). Despite acknowledgment that content, organization, style, vocabulary, and grammar are essential to produce good quality texts, studies have shown that raters' evaluations of text quality are influenced by style (sentence complexity and syntax; Freedman, 1979; Schultz, 1994). Research suggests that raters generally perceive a written text as superior if it has syntactically more complex sentences when compared to a written text consisting primarily of simple sentences (see Beers & Nagy, 2009; Crowhurst, 1980a; Saddler & Graham, 2005; Schultz, 1994; Stewart & Grobe, 1979). The ability to construct syntactically more complex sentences is essential because writers need to transform and organize ideas that sometimes require them to integrate information into grammatically correct written sentences. Limited knowledge in conveying information using sentence structures that have varying levels of complexity may hinder a writer's ability to translate thoughts efficiently. This is especially important for students at higher grade levels who are expected to produce more sophisticated sentence structures.

Syntactic Complexity

Syntactic complexity has been actively investigated as an important construct in language development research for more than five decades (Jagaiah, 2016). The influence of studies on syntactic complexity peaked in the 1970s and 1980s, but virtually disappeared by the

1990s, when grammar instruction waned with the advent of the writing process instruction. Writing process instruction focused mainly on planning, drafting, revising, and editing of essays and did not address syntactic complexity. One possible reason is that the high-stakes tests required by the “No Child Left Behind” law did not assess grammar specifically. Consequently, sentence construction skills are no longer emphasized in writing instruction in K-12 classrooms.

Studies show that sentence construction skills are mastered at a relatively early age, and growth continues through college (Haswell, 2000; Stewart, 1978). Berninger, Yates, Cartwright, Rutberg, Remy, & Abbott (1992) claimed that basic mastery of sentence structures occurs by grade two. When writers become more proficient and advanced, they become skillful at generating varied complex structures (McCutchen et al., 1994) and longer sentences (Crossley, Weston, McLain Sullivan, & McNamara, 2011; Ferrari, Bouffard, & Rainville, 1998; Haswell, 2000; McNamara, Crossley, & McCarthy, 2010; McNamara, Graesser, McCarthy, & Cai, 2014). Using varied and complex structures in written texts is associated with writing quality (Crossley, Roscoe, McNamara, 2014). However, this association is not consistent, quantifiable or associated with specific syntactic complexity measures.

One of the three important elements of writing development in addition to fluency and accuracy is syntactic complexity (Hunt, 1965; 1970). Although numerous studies have examined syntactic complexity in the past, measures used to examine syntactic complexity have remained a challenge for researchers. Various sentence elements (syntactic complexity measures) can be used to measure a syntactically complex sentence.

A sentence is composed of various constituents whether at the level of the word, phrase, or clause. When these constituents are combined and arranged in grammatical ways, a potentially infinite set of simple or complex sentences can be formed (Chomsky, 1957; Givon,

2009). Consequently, researchers have proposed a wide range of SCMs to characterize syntactic complexity.

Past measures of syntactic complexity. One common goal of previous studies was to identify SCMs to facilitate precise characterization of a sentence that is syntactically complex. Consequently, this led to a fairly large number of SCMs. Jagaiah (2016) identified 52 SCMs that have been used across 36 studies. The set of measures discussed in Jagaiah's (2016) systematic review represents a fairly complete range of elements used to examine sentences. In general, most of the SCMs can be categorized into the following six clusters: T-units, sentences, clauses, phrases, words, and combined measures. Of these clusters, two different classes of measures were used. The first class comprises measures that examined length: T-units, sentences, clauses, and phrases. The second class comprises measures that examined frequency count: number of T-units, clauses, phrases, and words.

Because past studies differed from one another in a multitude of ways, and the numerous measures do not clearly indicate their property or association to grade levels, students' writing abilities, genres, and writing quality, it is difficult to determine if the measures examined truly reflected syntactic complexity. For example, some researchers examined both words per T-unit and words per clause (e.g., Beers & Nagy, 2009; Crowhurst & Piche, 1979; Grobe, 1981; Hunt, 1970; Rubin & Piche, 1979; Price & Graves, 1980; Rodrigues, 1980; Rousseau, Bottge, & Dy, 1993; Smith, 1974; Smith & Swan, 1978; Stewart, 1978). Both these measures are mathematically similar as they account for the length of a T-unit or a clause. A main clause is a T-unit, and a main clause with a subordinate clause is also a T-unit. Without a clear definition of which measures were being used, studies have inadvertently examined syntactic complexity using several different measures. This indicates some of the limitations of previous studies on

syntactic complexity. Also, it is unclear if all the SCMs studied in the past are important measures of syntactic complexity, and whether there are patterns that can be discerned regarding important characteristics such as grade level, students' writing abilities, and genre. Finally, the size and type of writing samples analyzed varied across studies as well. Because of the labor-intensive nature of manual analysis, the size of the samples analyzed tended to be small, and there was no information on interrater reliability.

While there was no consensus on the definition of syntactic complexity and which measures accurately captured syntactic complexity, a study of the accumulated body of research on syntactic complexity suggests that at the syntactical level, complexity can be examined in relation to four latent variables: Types of Sentence Patterns, Sentence Length, Use of Sentence Connectors, and Sentence Sophistication (Jagaiah, 2016).

The following section provides an overview of these four latent variables which have been examined by Jagaiah (2016) in her systematic review.

Sentence pattern. The Sentence Pattern latent variable reflects grammatical classes at clause, phrase, and word levels. Specifically, it incorporates the following four components: (a) sentence types (e.g., simple, compound, complex, compound-complex), (b) word order of main and subordinate clauses (e.g., subject + verb, subject + verb + object), (c) word classes (e.g., nouns, verbs, adjectives, adverbs, determiners, prepositions, conjunctions, modal auxiliaries, and have or be auxiliaries), and (d) phrases (e.g., noun phrases, verb phrases, adjective phrases, adverb phrases). Sentence types examined by Blair and Crump (1984) found that students used more simple sentences in the descriptive mode compared to the argumentative genre across the three grade levels. Complex sentences were found to be highly used in the argumentative genre; however, significant difference in use was only found at the tenth-grade level. Fewer

compound sentences were used in the argumentative genre across all three grade levels. Moran (1981) also examined the use of sentence types in the argumentative, descriptive, and explanatory genres written by students with learning disabilities (LD) and low-achieving students in grades seven through ten. Both groups of students used all sentence types. However, students with LD averaged fewer simple, compound, and compound-complex sentences, but more complex sentences. Both run-on sentences and fragments occurred more frequently in the samples from students with LD samples on the average.

Blair and Crump (1984) also examined word order of main and subordinate clauses. Two word-order patterns, subject-verb and subject-verb-object, revealed consistently higher proportions of use for all three grade levels in the descriptive genre. However, several word-order patterns showed consistent differences of use between genres across the grades. The proportions of subject-verb-complement (noun) patterns were sizably larger for the argumentative genre at all three grade levels.

Moran's (1981) study of word classes did not yield any significant differences in the use of word classes between students with LD and low-achieving students. The reason for this is that all words in a sentence belong to a specific word-class category. A string of words in a sentence matches a specific word class. Therefore, there were no significant differences in the use of word classes between students with LD and students who are otherwise low-achieving. It must be noted that the use of numerous adjectives, adverbs, or noun and verb phrases are likely to increase the complexity of sentences because they are packed with more information compared to sentences without additional use of word classes.

More recent studies tend to include phrasal complexity using the length of phrases as a measure to examine sentence pattern (see Crossley, Weston et al., 2011; Ravid & Berman,

2010). Both Crossley et al., (2011) and Ravid and Berman (2010) argued that the phrase measures are an important component of sentence patterns to examine syntactic complexity because sentences using more phrases were found to be syntactically more complex.

Sentence length. To examine Sentence Length, previous studies used the following sentence elements: T-units, sentences, clauses, and phrases (e.g., Beers & Nagy, 2009, 2011; Crossley et al., 2011; Hunt, 1970; Ravid & Berman, 2010). These elements were examined by calculating the number of words in the T-units, sentences, clauses, and phrases. The longer the elements, the more syntactically complex the sentences are.

Studies show that T-units, sentences, clauses, and phrase length increase with grade level (see Crossley et al., 2011; Crowhurst, 1980a; 1980b; Evans, 1979; Hunt, 1970; Morris & Crump, 1982; Ravid & Berman, 2010; Rodrigues, 1980; Rousseau, Bottge, & Dy, 1993; Rubin & Piche, 1979; Stewart, 1978; Stewart & Grobe, 1979; Wagner, et al., 2011). These studies also found that students in the early grades produced shorter and less syntactically complex sentences. Hunt (1970), however, argued that sentence length is not a good index of syntactic complexity, at least during early grade levels, because the differences are not significant. These inconsistent findings make these studies difficult to compare.

Sentence connector. Sentence Connector refers to the link between ideas and embedded clauses. Using different types of links or connectives such as causal (e.g., because, so), logical (e.g., and, so), contrastive (e.g., although, however), temporal (e.g., first, next) and additive (e.g., and, moreover, also) not only creates a connection between two clauses, but also creates sentences with varied sentence patterns (e.g., simple, compound, complex, compound-complex, subject-verb-object, subject-subject-verb-verb-complement; see Blair & Crump, 1984;

Morris & Crump, 1984). The more varied the sentence patterns, the more complex the sentences will be.

Text connectors are also an essential component of sentence construction skills. Use of connectives (e.g., although, unless, moreover, before) begins in children who are in first grade (King & Rentel, 1979) and continues until eighth grade (McCutchen & Perfetti, 1982). McCutchen (1986) found that eighth-grade students use more connectors compared to sixth-grade students. Crossley, Weston, et al. (2011), on the other hand, found that ninth-grade writers used greater incidence of connectors in their texts than did eleventh-grade writers and first-year college writers.

Sentence sophistication. Sentence Sophistication refers to instances of phrases (e.g., noun, verb, prepositional, adverb) and embedded clauses in a sentence. A sentence contains many structural levels that are dependent on the combination of various types of clauses (independent and dependent clauses) and phrases. The length of the clauses and phrases also contributes to the complexity of the structure (see Beers & Nagy, 2009; 2011; Crossley et al., 2011; Rubin & Piche, 1979). Studies found that more sophisticated structures were used in the argumentative texts than in descriptive, narrative, or compare- and- contrast genres (see Beers & Nagy, 2009; Blair & Crump, 1984; Crowhurst, 1980a; Crowhurst, 1980b; Crowhurst & Piche, 1979; Prater & Mayo, 1984), and in texts by students who are typically-achieving than in those by low-achieving students who are and students with LD (Lane & Lewandowski, 1994); Morris & Crump, 1982); Prater & Mayo, 1984); Hunt, 1970), and in higher grade levels (see Balioussis, Johnson, & Pascual-Leone, 2012; Crossley, Weston, et al., 2011; Crowhurst, 1980a; 1980b; Evans, 1979; Hunt, 1970; Morris & Crump, 1982; Ravid & Berman, 2010; Rodrigues,

1980; Rousseau, Bottge, & Dy, 1993; Rubin & Piche, 1979; Stewart, 1978; Stewart & Grobe, 1979; Wagner et al., 2011).

When examining the complexity of sentence structures, most previous studies used T-units, a traditional method of measuring sentence sophistication. A T-unit can be defined as the shortest allowable grammatical unit that comprises the main clause and subordinated clauses (Hunt, 1965, 1970). T-units break sentences that are long or are joined by a string of the conjunction *and* which make the sentences ungrammatical. Breaking them into T-units provides a better count of the sentence length. However, the use of T-units to examine syntactic complexity has provided mixed results, with some studies demonstrating no patterns between T-unit measures by grade level, students' writing abilities, or genre (Belanger & Martin, 1984; Crowhurst, 1980a; Hunt, 1970; Stewart & Grobe, 1979). Error-free T-units are a better measure because they are governed by the rules of sentence construction (Crossley & McNamara, 2014). However, it is difficult to identify error-free T-units using automated essay scoring, and to do it manually would require expert hand coding, which is susceptible to subjectivity and error.

Syntactic Complexity and Grade Levels

Most studies found that higher grade-level students wrote syntactically more complex sentences when compared to lower-grade level students. In his seminal study, Hunt (1970) examined sentences written by students in grades four, six, eight, ten, and twelve. He reported that as grade-level increases, students tend to write syntactically more complex sentences. This was further supported by subsequent studies (see Balioussis, Johnson, & Pascual-Leone, 2012; Crossley, Weston et al., 2011; Crowhurst, 1980a; 1980b; Evans, 1979; Hunt, 1970; Morris & Crump, 1982; Ravid & Berman, 2010; Rodrigues, 1980; Rousseau, Bottge, & Dy, 1993; Rubin

& Piche, 1979; Stewart, 1978; Stewart & Grobe, 1979; Wagner et al., 2011). However, these findings were dependent on the type of latent variables examined at each grade level.

Sentence length. Several studies showed a steady increase in Sentence Length (mean number of words per T-unit score) at advanced grade levels in every elementary, middle school, and high school (see Rousseau, Bottge, & Dy, 1993; Stewart, 1978) or grade-level interval (Crowhurst**,** 1980a; Crowhurst, 1980b; Crowhurst & Piche, 1979; Evans, 1979; Morris & Crump, 1982; Rubin & Piche, 1979; Stewart & Grobe, 1979; Wagner et al., 2011). At the postsecondary level, Haswell (2000) also reported that undergraduate, matriculating, and junior students produced longer Sentence Length, indicating syntactic growth over time. This suggests that as students mature, they use an increased number of words in sentences, which increases the length of the sentences.

Sentence sophistication. Studies that investigated Sentence Sophistication (mean number of clauses per T-unit) concluded that its use in the written texts increased with grade level (see Crowhurst & Piche, 1979; Smith, 1974; Stewart & Grobe, 1979; Wagner et al., 2011). These findings were also supported by Ravid and Berman (2010), Beers and Nagy (2011), and Rousseau, Bottge, and Dy (1993) which suggested as students mature, they tend to write more sophisticated sentences (subordination structures), which increases the sentence complexity. However, other studies did not find similar results. Hunt (1970) found that Sentence Sophistication increased from grades four to six and six to eight but not from grades eight to ten or ten to twelve. This suggests that upon mastery of complex structures, students tend to use complex sentence structures in their texts at the higher-grade levels. A later study, however, contradicted Hunt's findings. Stewart and Grobe (1979) found that Sentence Sophistication was higher than those reported by Hunt (1970) for similar grade levels (Grades 5, 8, and 11). The

contradictory findings could be a consequence of the higher grade-level intervals that were used in Stewart and Grobe's study (1979).

Higher syntactic complexity scores appear to be dependent on grade levels as well as the type of latent variables examined. Comparison between studies is difficult because individual studies investigated different latent variables, and each measure may indicate assorted levels of complexity that could be impacted by grade levels. However, most studies support Hunt's hypothesis that Sentence Length (mean number of words per T-unit, mean number of words per clause) and Sentence Sophistication (mean number of clauses per T-unit) are reliable measures of increasing maturity in writing.

Syntactic Complexity and Students' Writing Abilities

A few studies have used several latent variables to examine the relationship between syntactic complexity and students' writing abilities. These variables ranged from Sentence Length (mean number of words per T-unit and clauses), Sentence Sophistication, and Sentence Connector (frequency count of clauses and morphemes) to combinations of the four latent variables.

Sentence length. Hunt (1970) found that students who are typically achieving (TA) wrote more mean number of words per clause compared to low-achieving students. A clause can be as simple as *The cat ate a mouse* or as complex as *The cute little cat ate a huge black mouse*. The more words used in the clause increases Sentence Length and the level of sentence complexity. Other components of the Sentence Length variable (number of words per sentence, number of words per T-unit) were found to be highly used by students who are TA compared to students with LD and students who are low-achieving (Hunt, 1970; Moran, 1981).

Hunt (1970) and Houck and Billingsley (1989) also found that TA students produced

fewer T-units compared to students with lower writing abilities or students with LD. The reduced number of T-units is due to the increased number of complex sentences evident in the written texts of TA students. This is in contrast to students with LD who used more simple sentences (hence more T-units).

Taken together, these studies indicate that students who are TA produce more sophisticated structures (Sentence Sophistication) and longer sentences (Sentence Length) than students who are low-achieving and students with LD.

Sentence sophistication

Studies found that TA students used an increased number of clauses per T-units, sentence types (simple, compound, complex, and compound complex) and number of morphemes per word compared to students with learning disabilities or students who are low achieving (see Hunt, 1970; Moran, 1981; Prater & Mayo, 1984). This suggests that although constructing syntactically complex sentences is a difficult skill, students who are typically achieving are better able to manage this compared to struggling writers or students with learning disabilities.

Syntactic Complexity and Genre

Depending on the latent variables examined, the highest syntactic complexity scores occur in different genres. Generally, most studies have found the argumentative genre to have the highest syntactic complexity score compared to descriptive or narrative genres.

Sentence length. Hunt (1965, 1970), claimed that Sentence Length (mean number of words per clause) predicted syntactic complexity, and this was supported by later studies of Sentence Length (mean number of words per T-unit; see Beers & Nagy (2009); Blair & Crump (1984) and mean number of words per clause; see Beers & Nagy, 2009). They found longer

sentences were used in descriptive, argumentative, and expository genres. Two studies (Crowhurst and Piche, 1979; Beers and Nagy, 2011) found that Sentence Length (mean number of words per clause) indicated the descriptive genre to be more syntactically complex compared to one study (Beers & Nagy, 2009) for argumentative genre and another (Ravid & Berman, 2010) for expository genre. These findings suggest that all three genres require longer sentence structures to present information.

Sentence sophistication. Only one measure of the Sentence Sophistication latent variable, syntactic density score, was examined in the argumentative genre. Blair and Crump (1984) found syntactic density score, a combination of ten different measures that incorporate measures from the four latent variables, to be highly evident in argumentative texts. The combination of the four latent variables increases the complexity of a sentence structure.

In summary, most studies that examined syntactic complexity and genre found that for the argumentative genre, students tend to use longer and more sophisticated sentences.

Syntactic Complexity and Writing Quality

The relationship between SCMs and writing quality has not been clearly established by previous studies (Jagaiah, 2016). Only a few studies examined this relationship using two latent variables: Sentence Length (number of words per T-unit and number of words per clause) and Sentence Sophistication (number of clauses).

Sentence length. Beers and Nagy (2009) found positive, negative, and no significant correlations depending on the grade levels examined while Stewart and Grobe (1979) reported a weak relationship between Sentence Length and writing quality. Longer T-units and longer clauses did not increase the writing quality in the argumentative and narrative genre in Beers and Nagy's (2009) study. Similarly, Stewart and Grobe (1979) did not find any relationship

between Sentence Length and writing quality in the expository genre. Grobe (1981), on the other hand, found Sentence Length components to be poor predictors of writing quality.

Crowhurst (1980a) also examined the relationship between Sentence Length (mean number of words per T-unit) and writing quality of narrative and argumentative essays of students in grades six, ten, and twelve. She found that argumentative essays that had longer sentence length received significantly higher writing scores at both grades 10 and 12. She also found that as students mature, they tend to write longer sentences. In contrast, there was no significant difference between the high and low Sentence Length scores in either the narrative or argumentative texts for grade six.

Stewart and Grobe (1979) found that Sentence Length (mean number of words per T-unit and mean number of words per clause) in grade-five written texts correlated significantly with quality ratings, but there was no relationship for texts written by grades eight and eleven students. Belanger and Martin (1984), on the other hand, found a very weak negative correlation between Sentence Length (T-unit length) and writing quality in the writing samples of grade nine or grade ten students. The weak negative correlations between writing quality and Sentence Length (mean number of words per T-unit) across all grade levels and genres supported findings by Stewart and Grobe (1979), and Crowhurst (1980a), which suggest that mean number of words per T-unit was not a good indicator of writing quality. Although such findings give useful hints, they do not provide a clear picture of the relationship between Sentence Length and writing quality.

Sentence sophistication. Stewart and Grobe (1979) examined Sentence Sophistication using mean number of clauses per T-unit of texts written by grade-five students. They found that the relationship between Sentence Sophistication and writing quality correlated

significantly. However, a similar finding was not established for texts written by grade- eight and- eleven students. Similarly, Belanger and Martin (1984) did not find a significant relationship between Sentence Sophistication (syntactic density score) and writing quality in the writing samples of grade-nine or grade-ten students.

In conclusion, the latent variables that examined the relationship between syntactic complexity and writing quality did not appear to be good predictors of writing quality. Because only a few studies examined the relationship between syntactic complexity and writing quality, it was difficult to observe a specific pattern. Furthermore, relationships between writing quality and syntactic complexity between different genres may not be meaningful because each genre requires the use of distinct types of syntactic structures to present information precisely. This could also explain the lack of a consistent relationship between syntactic complexity and writing quality by grade levels as these studies used different genres and measures from each latent variable to investigate this relationship.

SCMs from Coh-Metrix

A few SCMs from Coh-Metrix were used to examine syntactically complex sentences produced by K-12 typically achieving and struggling writers who are native speakers of English. Studies using SCMs from Coh-Metrix for this population are limited. However, there are numerous studies using SCMs from Coh-Metrix to analyze essays written by English Language Learners.

One of the studies using SCMs from Coh-Metrix is Crossley, Weston et al., (2011), who examined the mean number of words before the main verb, the mean number of high-level constituents (defined as sentences and embedded sentence constituents) per word, and the average number of modifiers per noun phrase in essays written by students in ninth grade,

eleventh grade, and college freshman. Crossley et al., (2011) grouped these SCMs into broad measures that reflected general linguistic constructs, and they selected the variables that were significantly different as a function of the writers' grade levels. Of these three SCMs, only mean number of words before the main verb was the best predictor of essay grade level. Using a pairwise comparison, Crossley et al., (2011) found that all grade levels demonstrated significant differences from one another in terms of the mean number of modifiers per noun phrases used. More advanced writers at the freshman college level produced a greater number of syntactically complex sentences (as measured by the number of modifiers per noun phrase) than the ninth-grade writers.

Coh-Metrix has also been used to discriminate between low- and high-quality essays. McNamara, Crossley, and McCarthy (2010) examined argumentative essays (n = 120) from undergraduate students at Mississippi State University using two SCMs from Coh-Metrix (the mean number of higher-level constituents per word and number of words before the main verb) to identify linguistic features of writing quality in English as the first language context. The essay length was limited to 500-1,000 words and four essay topics. Findings showed that only mean number of words before the main verb showed the largest difference between high- and low-proficiency essays. McNamara et al. (2010) did a stepwise regression analysis and found that mean number of words before the main verb predicted essay ratings. Essays that had a greater number of words before the main verb were rated highly, indicating that more successful essays were more syntactically complex. However, studies using selected Coh-Metrix measures to examine syntactically complex sentences produced by K-12 typically achieving or struggling writers who are native speakers of English are limited. Therefore, it was difficult to determine if the two Coh-Metrix SCMs predicted writing quality.

Summary

A proliferation of studies in the past five decades investigated syntactic complexity, but they only used individual SCMs in contrast to clusters of SCMs. While collectively many SCMs were examined, each study only consistently employed one to three SCMs in their analysis. Despite the fact that there have been some positive relationships between individual SCMs (e.g., mean number of words per T-unit, mean clause length, mean number of words per phrase) and syntactic complexity, little is known about how closely related SCMs when combined together within categories or as latent variables emerge as a more effective method of examining syntactic complexity. Individual SCMs are not sensitive enough to provide this information. Hence, there is no consensus among researchers regarding what qualifies as the most appropriate SCMs or clusters of SCMs to examine syntactic complexity. The gap in the literature is the lack of any large comprehensive study that examines numerous SCMs and identifies significant clusters of SCMs or latent variables which best indicate syntactic complexity.

Therefore, the current study examined the fit of a hypothesized model that grouped 28 Coh-Metrix SCMs into four latent variables using the eighth-grade, automatically scored formative assessment data for argumentative writing. The four latent variables are Sentence Pattern, Sentence Length, Sentence Connector, and Sentence Sophistication. The current study also examined the relationship between the four latent variables with writing quality, and how the relationship varied between at-risk and non-at-risk students.

CHAPTER 3

METHOD

Study Design

A hypothesized model based on 28 Coh-Metrix SCMs as indicators of four latent variables (Sentence Pattern, Sentence Length, Sentence Connector, and Sentence Sophistication) was analyzed using Confirmatory Factor Analysis (CFA) to test the model fit of eighth-grade automatically scored formative assessment data for writing. A multiple regression approach was then used to test if the four latent variables based on the fitted Coh-Metrix SCMs had a relationship with writing quality, and whether the latent variables impacting writing quality varied between at-risk and not-at-risk eighth-grade students.

Data Source

Data for this study were drawn from the 2012-2013 spring semester of a Benchmark Writing Assessment System (BAS-Write), a web-based skills assessment tool. BAS-Write was a statewide computer-based, automatically scored benchmark writing assessment for students in grades three to eight. The BAS-Write provided classroom teachers an avenue to formatively screen students' writing abilities and plan writing instructions.

A State Department of Education in one of the Northeastern regions of the United States provided archived de-identified essays that included demographic information such as district name and District Reference Group (DRG), school name, sex, race, grade level, status of free/reduced lunch, status of English Language Learner (ELL), and number of students receiving special education services. These data included students' typed responses to a prompt, the length of responses (number of words), and scores for sentence structure, word choice, mechanics, number of spelling and grammar errors for each response, and the state

accountability assessment writing scores. The prompts were based on three different genres: argumentative, informative/explanatory, and narrative. In addition to prompts from Measurement Incorporated, teachers, school administrators, and district personnel also created their own writing prompts for all the three genres.

The essays from BAS-Write comprised responses to several independent prompts that required test-takers to compose an argumentative essay that asserted and defended an opinion on a particular topic. Test-takers typed their responses to an on-screen writing prompt that was prepackaged for each grade level, with a 60-minute time constraint. Once students had completed the task, they submitted their essays and received individualized quantitative feedback. Students had a choice to revise their essays as many times as they liked after each submission. For the purpose of this study, only the first submission was considered for analysis in order to capture students' abilities to construct syntactically complex sentences prior to any automated feedback.

Each essay was scored on a scale that ranged from six to thirty-six using an Automated Essay Scoring (AES) engine called Project Essay Grade (PEG; Page, 1966, 1994). The students received six sub-scores ranging from one to six related to writing quality: overall development, organization, support, sentence structure, word choice, and mechanics using both holistic and traits scores, which were similar to scores assigned by human raters (Chung & O'Neil, 1997; Shermis & Burstein, 2003). The essay score (PEG Sum Score) is the sum of the six individual sub-scores which was used as the measure of writing quality.

To provide the writing quality scores, PEG uses *trins,* an intrinsic variable (fluency, diction, grammar, punctuation), and *proxes* (the approximation correlation between the intrinsic variables). When computing *proxes*, the actual counts of establishing the correlation of fluency

or *trin* with the amount of vocabulary or *prox* in an essay are considered (Page, 1994). Using a two-stage scoring system (training stage and scoring stage), the proxy variables from the scoring stage were determined for each essay and entered into the prediction equation. The beta weights from the training stage were then computed and a score was assigned (Chung & O'Neil, 1997). The more current PEG model contains several parsers, various dictionaries, special collections, and classification schemes to compute the scores (Page, 2003; Shermis & Barrera, 2002).

Writing Samples

The current study examined writing samples of eighth-grade students who responded to argumentative prompts for an automatically scored formative writing prompt during the 2012 - 2013 spring semester. Eighth-grade essays were selected for this study because this is the age group where more sophisticated use of sentence structures typically emerges (Beers & Nagy, 2009, 2011; Blair & Crump, 1984; Hunt, 1965; 1970; Rousseau, Bottge, & Dy, 1993). When constructing syntactically complex sentences, students need to deal with a number of demands, including selections of appropriate clauses, phrases, and words, inter- and intra-sentence connections. This process is even difficult for college-age students (Evans, 1979; Kagan, 1980; Kellog, 1987; Ravid & Berman, 2010; Smith & Swan, 1978; Stewart, 1978) because high levels of cognitive resources are imposed by sentence construction. Therefore, students at elementary and lower middle-school grade levels may not be able to produce various sentences that are syntactically complex. Examining students' writing preparedness, specifically constructing syntactically complex sentences, for this eighth-grade age group is critical because they are most linked to success at high school. If eighth graders develop difficulties in constructing syntactically complex sentences, this impedes their mastery of the more complex writing

process that emerges at later stages. In particular, it is important to examine the use of sentence structures that are syntactically complex by eighth-grade struggling writers which may be very different from typically achieving students.

Argumentative essays were selected because previous studies found that effective argumentative essays tend to include complex sentences (Beers & Nagy, 2011; Blair & Crump, 1984; Crowhurst & Piche, 1979; Perron, 1977; San Jose, 1972). Writers need to establish close causal links between facts and their opinions, and a syntactically complex sentence allows them to make the connections clear. One major component evaluated in argumentative essays is the logical reasoning used to support the arguments, the quality of which can be increased by using more sophisticated sentence structures. Therefore, it is expected that variability in writing quality might be explained particularly well in this genre by the use of syntactically complex sentences.

In addition, state and national writing standards have a strong emphasis on argumentative writing. This is true of the state standards where the data are obtained from students in grades six, seven, and eight, and the state accountability assessment for grade six. The Common Core State Standards (CCSS), which are the recognized standards for most states, also emphasize the argumentative genre for students in grades six and higher.

Automated Text Analysis Tool (Coh-Metrix 3.0)

The writing samples were analyzed using Coh-Metrix, the afore-mentioned automated text analysis tool that provided a large array of sentence complexity indices (Graesser, McNamara, Louwerse, & Cai, 2004). Coh-Metrix was developed to analyze and measure text on five levels of discourse: words, syntax, textbase, situation model, and genre/rhetorical structure (Graesser & McNamara, 2011; Graesser et al., 2004; McNamara, Graesser, &

Louwerse, 2012; Graesser, Millis, & Zwaan, 1997; Kintsch, 1998; McNamara, Louwerse, McCarthy, & Graesser, 2010; McNamara, Crossley, & McCarthy, 2010; Pickering & Garrod, 2004). The situation model refers to the deeper meaning representations that involve much more than the explicit words (van Dijk & Kintsch, 1983; Graesser & McNamara, 2012; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; Zwaan & Radvansky, 1998). The situation model is the subject matter described in text genre that comprises a mental representation of the deeper meaning of the text (Kintsch, 1998; Singer & Leon, 2007). Mental representations are schemata that in narrative texts, for example, include causation, intentionality, time, space, and protagonists that create cohesion and coherence (McNamara, et al., 2014).

Hundreds of measures were used to examine the five levels of discourse in relation to words, sentences, and connections between sentences that were funneled into factors that were aligned with these levels (Graesser & McNamara, 2011; McNamara & Graesser, 2012). A sixth level, pragmatic communication between speaker and listener, or writer and reader, is part of this framework, but this level is not relevant to this study. These hundreds of measures have been reduced to the current 106 indices in Coh-Metrix version 3.0. The current study examined 28 of the 106 indices. The 28 Coh-Metrix SCMs were selected from among the 52 SCMs compiled in Jagaiah's (2016) systematic review and linguistic theory. For a full description of the entire Coh-Metrix indices, see Graesser et al. (2004), and McNamara and Graesser (2012).

At least 100 published studies have established Coh-Metrix as an extremely powerful text analysis tool that is capable of assessing and differentiating an enormous variety of text types from the genre level to the sentence level (McNamara, Crossley, & McCarthy, 2010; McNamara et al., 2014). In addition to the external validation of Coh-Metrix as a recognized and authoritative text analysis tool, other reasons led to choice of Coh-Metrix. First, Coh-

Metrix provides a range of SCMs at the text, paragraph, sentence, and word levels, and it has been used extensively to analyze texts in written and oral discourse. This was especially pertinent to the current study because sentence-level measures in writing samples were evaluated. These indices include both clausal and phrasal attributes as noted in the literature (e.g., Crossley et al., 2011; Ravid & Berman, 2010), and Coh-Metrix fulfills this requirement.

Second, Coh-Metrix was developed by researchers in the fields of computational linguistics (Jurafsky & Martin, 2008), corpus linguistics (Biber, Conrad & Reppen, 1998), discourse processes (Graesser, Gernsbacher, & Goldman, 2003; Sanford & Emmott, 2012), cognitive science (Kintsch, 1998; Landauer, McNamara, Dennis, & Kintsch, 2007), and psychology (McCarthy & Boonthum-Denecke, 2012) to capture numerous indices and differentiate between different types of clausal embedding. This is important because the analysis of syntactic complexity encompasses theories from multiple disciplines.

Finally, Coh-Metrix has the ability to facilitate a large-scale, empirical evaluation of a wide range of indices used to measure syntactic complexity. This is critical, given the large sample size used for this study. Closely related to this point are the speed and flexibility offered by Coh-Metrix in assessing syntactic complexity, in contrast to using human raters who could be subjective, have training requirements, require time to score, and may have poor inter-rater reliability, which consume time and resources. The Charniak parser, an integral component of the Coh-Metrix algorithm, reports the highest average accuracy for expository and narrative texts (with greater accuracy reported for narrative texts; Hempelmann, Rus, Graesser, & McNamara, 2006) compared to any other parser types. The parser identifies the syntactic tree structure to scale the syntactic ease or difficulty (McNamara et al., 2014).

The following section describes how the writing samples were screened and how the hypothesized model was constructed to analyze the SCMs.

Writing Sample Selection

The BAS-Write data were provided by the State Department of Education and comprised grade-eight students who responded to argumentative, descriptive, informative, and narrative genres. These essays provided a general representation of writing as found in middle schools in the state selected. There were 3,172 writing samples written by 1,244 students.

Preliminary analyses were used to check for missing data. The next step was to ensure that the database only contained argumentative writing samples based on the criteria used to identify argumentative prompts. To determine that the prompts given to the students were argumentative, two raters examined the prompts based on several criteria. First, the prompt had to contain specific language. For example, the argumentative prompts required students to support, defend, or argue (for or against) a position by providing details that substantiated their stand. Second, the prompt could not require the students to refer to any outside texts such as articles or literary texts because they could be qualitatively different. Prompts that did not meet any of these criteria were removed. A total of 16 prompts were identified as argumentative (See Appendix A). Using these prompts to sub-set the data yielded 1,053 qualifying writing samples. Subsequently, the contributing records were matched with the Spring 2012 State Accountability Assessment (SAA) to ensure that each selected essay had a writing score that allowed at-risk and not-at-risk student classification based on the test scores. One essay was removed because the SAA writing score indicated that the student was a seventh grader while the remaining 1,052 students were in grade eight.

A discrepancy in word count was noted between the BAS-Write word count score and eight of the 1,052 essays scanned using Coh-Metrix. These eight essays were removed, and the total number of essays was reduced to 1,044. The eight essays were scanned by a data analyst at the University of Memphis, who used the Coh-Metrix software available at the university instead of the Coh-Metrix software that is available online, and that may have caused the discrepancy in the word count between the two versions. To ensure that the scores were consistent, only essays scanned by the online version of the Coh-Metrix 3.0 were retained.

The subsequent step involved scanning the 1,044 writing samples using the online Coh-Metrix version 3.0 to obtain the scores for the selected 28 SCMs. Because there was a discrepancy between the eight essays scanned using the Coh-Metrix software at the University of Memphis and essays scanned by the online version of the Coh-Metrix 3.0, the data entry was checked for errors. Word count for every essay scanned by the online version of the Coh-Metrix 3.0 that was different from the word count given by the BAS-Write was examined, and it agreed with 99.80% of the essays. Small discrepancies in word count were created from differences in whether a hyphenated compound noun (e.g., well-being) or compound adjective (e.g., well-behaved) was viewed as one or two words. To be consistent, word count obtained from Coh-Metrix was used because the values obtained for SCMs were also from Coh-Metrix.

Initially, 30 Coh-Metrix SCMs were selected. However, the Coh-Metrix output for the two SCMs, incidence score of positive connectives (CNCpos) and incidence score of negative connectives (CNCneg), produced no values. At the time these essays were scanned, the online Coh-Metrix 3.0 version was not able to compute the scores for the CNCpos and CNCneg SCMs. Consequently, these two SCMs were removed, and this reduced the SCMs from 30 to 28.

Next, a descriptive analysis was calculated for the 28 SCM values to identify outliers. Removing outliers was necessary to ensure that the models developed for the two research questions would be representative of the majority of the data being analyzed. Essays were designated as outliers if three or more SCMs fell outside three standard deviations of the mean, and these essays were removed. This further reduced the number of essays to 1,029.

Finally, students were classified as at-risk or not-at-risk students based on writing quality. Writing quality was measured by the writing scores provided in the Spring 2012 SAA by the AES (PEG Scores) using a standardized scoring rubric. Students who achieved in Bands 1 and 2 (i.e., below basic and basic) were classified as at-risk while students who achieved in Bands 3, 4, and 5 (i.e., proficient, goal, or advanced levels) were classified as not-at-risk. Based on the guidelines stated in the State Board of Education (2010), students classified as at-risk produced writing samples that included underdeveloped or minimally developed ideas that resulted in little expansion of key ideas and construction of awkward sentence structures. Students in this category had limited or no ability to apply the conventions of standard English to edit and revise written work. Students who were not-at-risk generally had between adequate to exceptional ability to communicate ideas in writing. Not-at-risk students were able, to a large extent, to expand on key ideas and also to apply conventions of standard English to revise and edit their work.

A total of 115 students were identified as at-risk, and a total of 914 were identified as not-at-risk (see Table 1). About 11.18% students in this dataset were at risk, and this number approximately matched the state's 11.90% (ages of 6 – 21) of students who had been identified as children with disabilities (IDEA Data Center, 2012).

Demographics

Table 1 summarizes the demographic information for the eighth-grade BAS-Write students.

Table 1

Demographic Information for the Eighth-Grade Benchmark Assessment-Write Data

Variable	At-Risk[a]	%	Not-At-Risk[b]	%	Total
Number of Students	115	11.18	914	88.82	1,029
Gender					
Female	31	5.88	496	94.12	527
Male	84	16.73	418	418	83.27
Race					
White	66	9.57	624	90.43	690
Hispanic/Latino	30	19.11	127	80.89	157
African American	15	16.13	78	83.87	93
Asian	3	3.80	79	96.34	82
American Indian/ Native Alaskan	1	50	1	50	2
Native Hawaiian/ Pacific Islander	0	0	0	0	0
Two or more races	0	0	5	100	5
Free or Reduced Lunch	57	23.36	187	76.64	244
English Language Learners	9	42.86	12	57.14	21
Special Education	55	52.38	50	47.62	105
DRG[c]					
A-C	34	5.85	547	94.15	581
D-F	18	19.78	73	80.22	91
G-I	17	18.88	73	81.11	90
X-Y	46	17.23	221	82.77	267

Note. Using the writing scores obtained from the Spring 2012 Grade Eight State Accountability Assessment, At-risk[a] = students who received Band scores of 1, and 2 and Not-at-risk[b] = students who achieved Band scores of 3, 4, and 5. DRG[c] = District Reference Group, categorizes school districts based on similar socioeconomic status (SES). DRG A indicated school districts that are very affluent with low-need, while DRG I indicated school districts that have significantly lower SES with significantly high need. DRG X refers to charter school, and DRG Y refers to magnet schools.

A demographic breakdown of writing scores was done before analyzing the data. Table 2 reports the mean and standard deviation of the writing scores for at-risk and not-at-risk students. The writing scores showed a significant difference by Student Type.

Table 2

Mean and Standard Deviation of the Writing Scores for At-Risk and Not-At-Risk Students

Variable	Mean At-Risk	SD At-Risk	Mean Not-At-Risk	SD Not-At-Risk	t	p
Writing Score	16.79	3.36	22.23	3.70	24.73	0.00*

Note. $N = 1{,}029$. * $p < 0.05$

Table 3 reports the mean and standard deviation of the writing scores for female and male students. The writing scores showed a significant difference by sex.

Table 3

Mean and Standard Deviation of The Writing Scores for Female and Male Students

Variable	Mean Female	SD Female	Mean Male	SD Male	t	p
Writing Score	22.63	3.84	20.56	3.97	8.52	0.00*

Note. $N = 1{,}029$. * $p < 0.05$

Motivation for Current Hypothesized Model

Previous studies examined syntactic complexity using individual SCMs. Most studies only used one to three SCMs, and these SCMs varied from one study to another.

None of the studies focused on the factor structure of syntactic complexity when examining the SCMs. Kagan (1980) conducted the only study attempting to identify SCMs that explained syntactic complexity using six principal component factors. However, it should be noted that Kagan's (1980) study examined the SCMs to identify a relationship between syntactic complexity and analytic cognitive style, but these six factors were not confirmed using a specified model. The current dissertation study is the first study to use CFA to analyze several SCMs simultaneously by grouping them into four latent variables.

Selected Latent Variables and Syntactic Complexity Measures

Twenty-eight Coh-Metrix SCMs were selected by referencing the 52 SCMs compiled in Jagaiah's (2016) systematic review and linguistic theory. The 28 Coh-Metrix SCMs were theoretically grounded and validated, and were aligned with theories of discourse which operate at multiple levels of language related to words, sentences, and connections between sentences (McNamara et al., 2014).

To be selected as an appropriate SCM to measure syntactic complexity, the Coh-Metrix SCMs had to have a perfect or partial match with the 52 SCMs in the systematic review, or they had to be related to sentence elements that would indicate syntactic complexity in linguistic theory. To be perfectly matched, the SCMs had to measure the same sentence element. For example, the number of prepositional phrases in the systematic review is the same as incidence score of prepositional phrases in Coh-Metrix. To be partially matched, the SCMs had to reflect syntactic complexity by nature of its function in the structure of the sentence. For example, the SCM, number of adverbs of time (when, then, once, while), is closely related to the temporal connectives incidence in Coh-Metrix, which also measures adverbs of time. However, it is not

clear if adverbs of time in the literature were limited to only four adverbs of time (when, then, once, while), while Coh-Metrix calculated all the adverbs of time. An example of an SCM selected based on linguistic theory is the agentless passive voice in Coh-Metrix. This SCM was included because, according to linguistic theory, passive construction is more complex than the active construction (Chomsky, 1965; Bresnan, 1981; Gazdar, Klein, Pullum, & Sag, 1985). This is evident in the way it is constructed. Passive voice is formed by combining a form of the verb to be with the past participle of a transitive verb or modal auxiliary verbs, and this increases the level of complexity of a sentence structure.

Hypothesized Model

In the current study, the researcher specified a hypothesized model with four latent variables and 28 Coh-Metrix SCMs by referencing the SCMs compiled in Jagaiah's (2016) systematic review and linguistic theory. Only five SCMs from the Coh-Metrix SCMs perfectly matched (see Appendix C) the SCMs in the literature, and five more were partially matched (see Appendix D). The remaining 18 Coh-Metrix SCMs were selected because they are related to sentence elements that would indicate syntactic complexity in linguistic theory (See Appendix E).

The fit of the 28 SCMs as indicators for the four latent variables was estimated in the following manner. First, the 28 Coh-Metrix SCMs were specified as indicators for the syntactic complexity attribute they were purported to measure. These attributes represented the four hypothesized latent variables in the current study: Sentence Pattern (12 SCMs), Sentence Length (3 SCMs), Sentence Connector (7 SCMs), and Sentence Sophistication (6 SCMs). Table 4 lists the 28 SCMs corresponding to the four latent variables. Tables 5 – 8 provide a detailed description of each SCM separately for each latent variable.

Table 4

Initial Hypothesized Model with Four Latent Variables and 28 Coh-Metrix SCMs

	Latent Variables			
	Sentence Pattern	Sentence Length	Sentence Connector	Sentence Sophistication
Coh-Metrix SCMs	Noun phrase incidence (DRNP)	Mean number of words (DESSL)	All connectives incidence (CNCAll)	Mean number of modifiers per noun phrase (SYNNP)
	Verb phrase incidence (DRVP)	Standard deviation of mean number of words (DESSLd)	Causal Connectives incidence (CNCCaus)	Minimal edit distance, part of speech (SYNMEDpos)
	Adverbial phrase incidence (DRAP)	Mean number of words before main verb (SYNLE)	Logical connectives incidence (CNCLogic)	Minimal edit distance, all words (SYNMEDwrd)
	Preposition phrase incidence (DRPP)		Adversative/contras tive connectives incidence (CNCADC)	Minimal edit distance, lemmas (SYNMEDlem)
	Agentless passive voice forms incidence (DRPVAL)		Temporal connectives incidence (CNCTemp)	Mean adjacent sentence structure similarity (SYNSTRUTa)
	Negation expression incidence (DRNEG)		Expanded temporal connectives incidence (CNCTempx)	Mean all sentence structure similarity (SYNSTRUTt)
	Gerund incidence (DRGERUND)		Additive connectives incidence CNCAdd)	
	Infinitive incidence (DRINF)			
	Noun incidence (WRDNOUN)			

Latent Variables			
Sentence Pattern	Sentence Length	Sentence Connector	Sentence Sophistication
Verb incidence (WRDVERB)			
Adjective incidence (WRDADJ)			
Adverb incidence (WRDADV)			

Note. Four latent variables: Sentence Pattern (12 SCMs), Sentence Length (3 SCMs), Sentence Connector (7 SCMs), Sentence Sophistication (6 SCMs) SCMs= 28

Sentence pattern indices in Coh-Metrix. Twelve SCMs from two different Coh-Metrix categories (syntactic pattern density and word information) were hypothesized to indicate the Sentence Pattern latent variable. These SCMs reflect grammatical classes at phrase and word levels. The word-level SCMs were included because previous studies found that students with reflective and articulated styles wrote longer sentences with increased numbers of nouns, verbs, adjectives, and adverbs (Kagan, 1980; Moran, 1981).

The Sentence Pattern latent variable also indicates the varied structures found within sentences based on the incidence score of the SCMs. It is informed by the density of specific syntactic patterns that reflect grammatical classes at phrase and word levels. As described by McNamara et al., (2014), an incidence score is computed for each part of speech category and for different sets of part-of-speech categories. An incidence score is defined as the number of occurrences of a particular category per 1,000 words, and these scores can be manually reproduced. For example, to compute the incidence score of noun phrase density, count the total number of noun phrases, divide this by the total number of words in the essay, and multiply it by 1,000. Therefore, if a sentence has a higher incidence of noun and verb phrases, it is packed with more information, thus making the sentence more complex. Table 5 provides

the definition, example, and variables of each SCM in the Sentence Pattern latent variable used in the analysis.

Table 5

Syntactic Complexity Measures in Coh-Metrix that Indicate the Sentence Pattern Latent Variable

Sentence Pattern SCMs	Definition	Example of Structure	Variable Name
Incidence score of noun phrase	A noun phrase comprises a noun (person, place, or thing) and modifiers (phrases and clauses that describe the noun)	I enjoy watching at *the glistening snow.*	DRNP
Incidence score of verb phrase	A verb phrase comprises an auxiliary or helping verb, and the main verb.	You *should have listened* to your teacher.	DRVP
Incidence score of adverbial phrase	Words that modifies the verb, adjective or an adverb. Prepositional phrases and infinitive phrases can function as an adverb phrase.	Walk *very carefully* across the wet floor.	DRAP
Incidence score of prepositional phrase	Begins with a preposition (e.g., on, at, in, with) and ends with a noun, pronoun, gerund, or clause.	I will visit you *in the evening*.	DRPP
Incidence score of agentless passive voice	A passive clause with no by-phrase or agent (doer).	*The old books were packed and stored in the garage.*	DRPVAL
Incidence score of negation	Refers to statement that is not true, or it is not the case.	neither, neither…nor, not, never *Neither* of us bought the books although we were expected to buy at least one. Using prefixes: dis-, un- and suffixes -less.	DRNEG

Sentence Pattern SCMs	Definition	Example of Structure	Variable Name
		The student was *disrespectful* to the teacher.	
Incidence score of gerund	Gerunds function as nouns, and every gerund ends in *ing*	*Reading* is my favorite pastime.	DRGERUND
Incidence score of infinitive	Always begin with *to* followed by a verb.	I wanted *to write* a poem	DRINF
Incidence score of nouns	A noun refers to people, places, things, or animals	The *cat* caught the *bird*.	WRDNOUN
Incidence score of verbs	A verb refers to an action or state.	The boy *opened* the door and *walked* into a dark room.	WRDVERB
Incidence score of adjectives	An adjective refers to a word that describes the noun	I brought home a *big* and *heavy* sofa.	WRDADJ
Incidence score of adverbs	A word that describes or modifies an adjective, a verb, or other adverb.	The kids ran *quickly* in the park.	WRDADV

Note. Hypothesized latent variable: Sentence Pattern. SCMs=12

Sentence length. Sentence length can be captured by the number of words in a sentence, which makes sentence length a measurable variable. However, number of words in a sentence is not the only SCM that is captured by sentence length which also includes the standard deviation of the length of a sentence in a text. Thus, the Sentence Length latent variable is a combination of characteristics of sentence length that represent the syntactic complexity of a sentence.

Three SCMs from Coh-Metrix were selected from two different categories (descriptive and syntactic complexity) and hypothesized to indicate the Sentence Length latent variable. Sentences that are grammatically constructed with more words are longer, and they may be more complex (McNamara et al., 2014). The SCMs hypothesized to indicate the Sentence

Length latent variable were included because previous studies have shown some correspondence between sentence length and syntactic complexity (see Beers & Nagy, 2009; 2011; Crosley et al., 2011; Hunt, 1970; McNamara, et al., 2014; Ravid & Berman, 2010).

The first SCM hypothesized to indicate Sentence Length is the mean number of words per sentence or is literally the number of words per sentence. To compute the mean number of words per sentence, count the total number of sentences in and the total number of words in the essay. Then, divide the total number of words by the number of sentences in the essay. This computation can be manually reproduced.

Examining the standard deviation of sentence length (the second SCM) is essential because a large standard deviation indicates variety in sentence length, which could be an indicator of syntactic complexity. To compute the standard deviation of sentence length, one counts the number of words for each sentence in an essay and calculates the sample standard deviation.

Mean number of words before main verb is the third SCM that was included in the sentence length latent variable. It includes phrase or dependent clause length. Longer phrases and clauses indicate the use of more words, which increases the density of the information in the phrase or clause. A sentence that has a complex subject due to embedded phrases or clauses (e.g., adverbial clauses) before the main verb would receive a high SCM value. On the other hand, if a sentence has a less complex subject because it lacks a phrase or a clause embedded before the main verb, it would receive a low SCM value. For example, *Before the day ended in a horrific manner, the gracious and concerned teacher* (13 words) *managed* (main verb) *to calm down all her students* has a higher SCM value than the following sentence *Before the day ended, the teacher* (6 words) *managed* (main verb) *to calm down all her students.* To compute

the mean number of words before main verb one counts the number of words before each main verb and divides it by the total number of main verbs in the essay. This computation can be done manually. Table 6 provides the definition, example, and variables of each SCM in the Sentence Length latent variable used in the analysis.

Table 6

Syntactic Complexity Measures in Coh-Metrix that Indicate the Sentence Length Latent Variable

Sentence Length SCMs	Definition	Example of Structure	Variable Name
Mean number of words per sentence	Refers to the average number of words in each sentence in a text. A word in this context refers to anything that is tagged as a part-of-speech as indicated by the Charniak Parser.	I was late because I had to complete my task.	DESSL
Standard deviation of mean number of words per sentence	Refers to the standard deviation of the measure for the mean length of sentences in a text.	It is important to check your bag before you leave the class. Make sure your homework is in your bag.	DESSLd
Mean number of words before the main verb	Main verb is operationalized as the main verb in the first independent clause in sentence.	*The little girl* ate the pizza.	SYNLE

Note. Hypothesized latent variable: Sentence Length. SCMs=3

Sentence connector. Nine SCMs from a single Coh-Metrix category (connectives) were hypothesized to indicate the Sentence Connector latent variable. Connectors are important to create cohesive links between two independent clauses, or an independent clause with a dependent clause within a sentence or between sentences. The link provides clues for how the text was organized (Cain & Nash, 2011). Dichotomous in nature, connectors comprise both positive and negative features. Halliday and Hasan (1976) and Louwerse (2001) state that

connectors are represented by a specific cohesion feature, and are associated with positive additive (e.g., also, moreover), negative additive (e.g., however, but), positive temporal (e.g., after, before), negative temporal (e.g., until), and causal (e.g., because, so) measures.

Connectors play an important role in the creation of cohesive links between ideas (Crismore, Markkanen, & Steffensen, 1993; Longo, 1994) and provide clues about text organization (van de Kopple, 1985). Connectors also add or contrast information within a sentence increasing the structural complexity of sentences (Blair & Crump, 1984; Moran, 1981) because they link ideas and clauses in a sentence or between sentences (McNamara et al., 2014).

Previous studies have shown a relationship between connectors and syntactic complexity. Moran (1981) found students with learning disabilities (LD) and students who are low achieving (LA) were able to construct sentences that were syntactically complex using connectors. The findings revealed that these students used both complex and compound sentences, of which require the use of connectors. Similarly, Blair and Crump (1984) found increased use of compound complex sentences in argumentative essays written by students with LD in grades six, eight, and ten. These essays were found to be syntactically more complex.

Connectors were calculated based on an incidence score defined as the number of occurrences of a particular connector per 1,000 words. For example, to compute the incidence score of causal connectors, count the total number of causal connectors, divide by the total number of words in the essay and multiply it by 1, 000 (Crossley & McNamara, 2011). These scores can be manually reproduced. Table 7 provides the definition, example, and variables of each SCMs in the Sentence Connector latent variable used in the analysis.

Table 7

Syntactic Complexity Measures in Coh-Metrix that Indicate the Sentence Connector Latent Variable

Sentence Connector SCMs	Definition	Example of Structure	Variable Name
Incidence score of all connectives	Connectors create cohesive links between ideas and clauses and provide clues about text organization. Five general classes of connective measures: Causal Logical Adversative/contrastive Temporal Additive. Positive and negative connectives can be found within the five general classes of connective measures.	Specific example for each type of connector is provided in the respective connector.	CNCAll
Incidence score of causal connectives	A sentence that denotes cause and requires the use of causal connectives.	'because', 'so', 'therefore'. Sentence: I was late this morning because it rained heavily.	CNCCaus
Incidence score of logical connectives	Two sentences joined by a grammatical conjunction to form a grammatically compound sentence.	variants of 'and', 'or', 'not' and 'if-then' Sentence: Jack went to the bookstore, and he bought a book.	CNCLogic
Incidence score of adversative and contrastive connectives	Words that are used to joining two ideas that are considered to be different	'although', 'whereas' Sentence: Although I was tired, I completed my task.	CNCADC

Sentence Connector SCMs	Definition	Example of Structure	Variable Name
Incidence score of temporal connectives	Words or phrases that tells when something is happening.	"first", "until' Sentence: First, you have to clean the potatoes.	CNCTemp
Incidence score of expanded temporal connectives	Words or phrases that tells when something is happening.	"first", "until' Sentence: I have until May to finish my project.	CNCTempx
Incidence score of additive connectives	Words used to add information or connect ideas.	"and", "moreover" Sentence: Jilla forgot to complete her assignment. Moreover, she forgot to prepare for her quiz.	CNCAdd

Note. Hypothesized latent variable: Sentence Connector. SCMs = 7

Sentence sophistication. Syntactic complexity also can be measured by examining the SCMs that indicate Sentence Sophistication. Six SCMs from Coh-Metrix were hypothesized to indicate the Sentence Sophistication latent variable. Some sentences have complex and embedded structures, and they increase the hierarchical levels in the structure of the sentence. Increased numbers of hierarchical structures indicate an increased level of complexity. Sentences that have an increased number of hierarchical levels are usually structurally dense with information (Graesser et al., 2004).

As seen in Table 6, six SCMs were used to examine Sentence Sophistication. Three of the SCMs have semantic and dissimilar sentence-structure properties. They measure how closely these sentences use similar structures by using the Minimal Edit Distance (MED) method of computation. The three variations of the MED (SYNMEDpos, SYNMEDwrd, and SYNMEDlem) were calculated by using the average of the distance between each of the SCMs

from one another between adjacent sentences in the essay. Coh-Metrix does not provide clear steps on how to calculate this measure, and it is not as straight-forward as it appears to interpret the scores. Table 8 provides the definition, example, and variables of each SCM in the Sentence Sophistication latent variable used in the analysis.

Table 8

Syntactic Complexity Measures in Coh-Metrix that Indicate the Sentence Sophistication Latent Variable

Sentence Sophistication SCMs	Definition	Example of Structure	Variable Name
Mean number of modifiers per noun phrase	Modifiers can be articles, possessive nouns, noun phrases, adjectives, participles, adjective clauses, and prepositional phrases and infinitives in a noun phrase. The number of modifiers in each noun phrase is counted. The total is divided by the total number of the words in the text. This computation can be manually reproduced.	It was a big, blue house.	SYNNP
Minimal edit distance of part of speech	Parts of speech refers to nouns, pronouns, adjectives, determiners, verbs, adverbs, prepositions, conjunctions, and interjections. It calculates the extent to which one sentence needs to be modified (edited) to make it have the same syntactic composition as a second sentence. These scores will indicate if the students have varied their sentence structures. To compute manually will be a laborious task. The algorithm in Coh-Metrix has built-in rules that will compute the scores.	The boy runs after the girl. The girl runs after the boy.	SYNMEDpos

Sentence Sophistication SCMs	Definition	Example of Structure	Variable Name
Minimal edit distance of all words	SYNMEDwrd calculates the extent to which one sentence needs to be modified (edited) to make it have the same syntactic composition as a second sentence. These scores will indicate if an essay has varied sentence structures. Because Coh-Metrix has built-in rules, it is difficult to reproduce these scores.	Similar sentence pattern: The cat took the ball from the rat. The rat took the ball from the cat. Dissimilar sentence pattern: The cat took the ball from the rat. The rat caught the ball and ran away.	SYNMEDword
Minimal edit distance of lemmas	SYNMEDlem calculates the extent to which one sentence needs to be modified (edited) to make it have the same syntactic composition as a second sentence. These scores will indicate if an essay has varied sentence structures. Because Coh-Metrix has built-in rules, it is difficult to reproduce these scores.	The position for the noun cat and rat are different. The cat took the ball from the rat. (The rat is an object) The rat took the ball from the cat. (The rat is the subject)	SYNMEDlem
Mean number of sentence syntax similarity between adjacent sentences	Proportion of intersection tree nodes between all sentences and across paragraphs. Measures the uniformity and consistency of the syntactic constructions in the text or similarity (Sim) between all combinations of sentence pairs across paragraphs. This SCM is measured by removing uncommon subtrees found between two adjacent sentences. Known as Sim, the SYNSTUTt is calculated the following way: Sim = nodes in the common tree/(the sum of the nodes in the two sentence trees – nodes in common tree)	Similar sentence pattern: The cat took the ball from the rat. The rat took the ball from the cat.	SYNSTRUTt

Sentence Sophistication SCMs	Definition	Example of Structure	Variable Name
	Example: The first tree sentence has 8 nodes and 6 nodes with 4 common nodes. The similarity is Sim = 4/((8 +6)-4) = 4/10 = 0.4		
Mean number of all combinations of sentence syntax similarity across paragraphs	Proportion of syntactic structures between all adjacent sentences. It examined syntactic similarity at the phrasal level and the parts of speech. Example 1: The dog (noun phrase) ran (verb). Example 2: It (pronoun) jumped (verb) into (preposition) the pond (noun phrase). Sim = nodes in the common tree/(the sum of the nodes in the two sentence trees – nodes in common tree) Example: The first tree sentence has 8 nodes and 6 nodes with 4 common nodes. The similarity is Sim = 4/((8 + 6) -4) = 4/10 = 0.4	The cat was under the chair. It saw a rat with a ball. The cat took the ball from the rat. The rat took the ball from the cat. The rat ran away. The cat took the ball from the rat. The rat squeaked loudly. The rat took the ball from the cat.	SYNSTRUTa

Note. Hypothesized latent variable: Sentence Sophistication. SCMs=9

Data Analysis

Statistical analysis. CFA and multiple linear regression (MLR) were used to analyze the data.

Confirmatory factor analysis. A CFA was performed using the statistical analysis package, Statistical Analysis System (SAS) 9.4 CALIS procedure, to analyze the data to determine if the hypothesized model based on four latent variables and 28 Coh-Metrix SCMs was a good fit using the eighth-grade, automatically scored formative assessment data for writing. Figure 4 provides a schematic depiction of the hypothesized model, which was driven by a combination of *a priori* and discourse theories. The CFA was carried out to assess which

SCMs indicated each latent variable and whether the hypothesized model appropriately fitted the data.

Four commonly reported indices were used to assess how well the model fitted the data: goodness-of-fit index (GFI), root mean square error of approximation (RMSEA, Comparative Fit Index (CFI), and standardized root mean square residual (SRMR), an index that is sensitive to misspecified factor covariances or latent structures (Hu & Bentler, 1999). RMSEA values of 0.06 or less, in conjunction with GFI values of more than 0.95, SRMR of less than 0.08, and CFI more than 0.95 were considered indicative of good fit (Hu & Bentler, 1999). Traditionally, chi-square value has been used as evidence of good model-data fit, but the chi-square value is sensitive to model size and non-normality (Bollen, 1989), and it is no longer widely used or reported.

Because each SCM uses different units of measurement, the SCM values were converted into a standardized *z* score measure before the CFA was performed. The standardized *z* score is given by the following formula:

$$\text{standardized z score} = (\text{SCM value} - \text{SCM mean}) / (\text{SCM standard deviation}) \qquad (1)$$

The CFA is an iterative process to determine how well the hypothesized model fits based on several CFA test metrics. The CFA test metrics did not support the initial hypothesized model, so the model was modified to achieve a better fit based on theoretical perspectives.

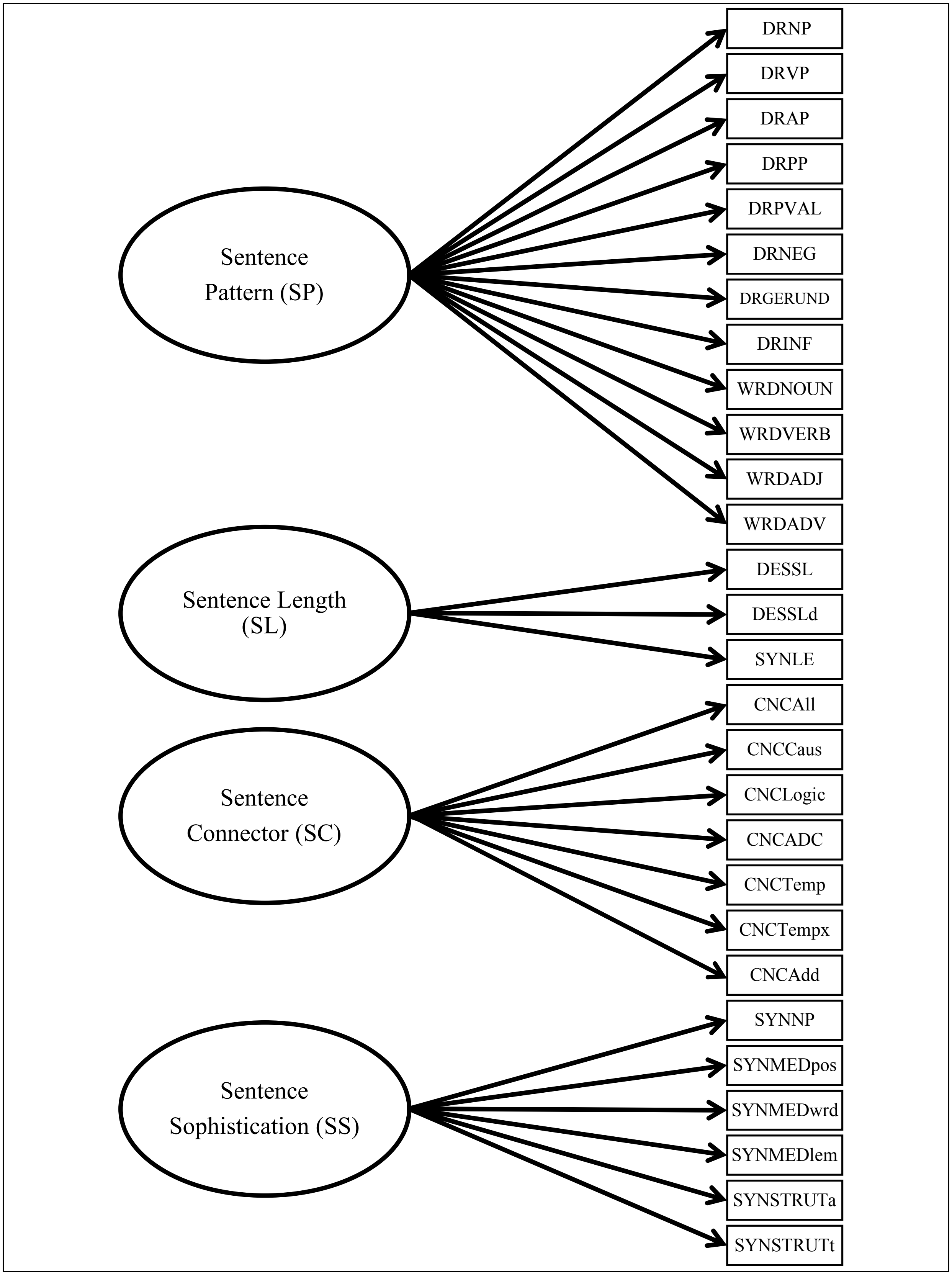

Figure 4. A confirmatory factor analysis with four latent variables and 28 SCMs.

Multiple linear regression. Scores of the four latent variables were computed for each essay using the factor score regression coefficients produced by the CFA. An MLR model was developed using scores of the four latent variables produced by the CFA as predictor variables, Student Type (at-risk or not-at-risk) as indicator variable, and writing score from the Spring 2012 State Accountability Assessment as the dependent variable to analyze research question two. MLR was an appropriate model to analyze RQ2 for several reasons. An MLR model structure was selected because, after the hypothesized model was found to be a good model fit using CFA, the data structure was not complex in terms of the relationships between the latent variables, student type, and writing quality. Second, the MLR was easy to understand and interpret. Finally, by incorporating interactions in the model, the MLR effectively captured the relationship of the four latent variables and student type on writing quality.

The multiple regression model had the following functional form:

$$y = \beta_0 + \beta_1 SP + \beta_2 SL + \beta_3 SC + \beta_4 SS + \beta_5 ST + \beta_6 ST*SP + \beta_7 ST*SL + \beta_8 ST*SC + \beta_9 ST*SS + \beta_{10} SP*SL + \beta_{11} SP*SC + \beta_{12} SP*ST\ \beta_{13} SP*SS + \beta_{14} SL*SC + \beta_{15} SL*SS + \beta_{16} SC*SS + e \quad (2)$$

where y is writing quality score, SP is Sentence Pattern, SL is Sentence Length, SC is Sentence Connector, SS is Sentence Sophistication, and ST is Student Type which was coded as zero for at-risk student and one for not-at-risk student. The regression model incorporated interactions between the latent variables and between the latent variables and Student Type to fully understand the impact of the latent variables and student type on writing quality of grade-eight students. The regression model determined the strength of the relationship between the latent variable scores resulting from the CFA with writing quality using standard regression metrics like R^2 and p values of the beta coefficients.

CHAPTER 4

RESULTS

Prior to testing the hypothesized model in RQ1 using CFA, a descriptive analysis of the unstandardized SCM scores between at-risk and not-at-risk students and between females and males based on the 28 SCMs was performed for each latent variable. A two-sample *t*-test of unequal variances was calculated to test whether the difference in mean scores were significant for each SCM at $\alpha = .05$.

Tables 9 to 12 show the results of the analysis between at-risk and not-at-risk students. For the Sentence Pattern latent variable, of the 12 SCMs only DRNP, DRPVAL, WRDADV showed significant differences between at-risk and not-at-risk students. For the Sentence Length latent variable, of the three SCMS only DESSL showed a significant difference between at-risk and not-at-risk students. For the Sentence Connector latent variable, of the seven SCMS only CNCAll, CNCCaus, and CNCTemp showed a significant difference between at-risk and not-at-risk students. For the Sentence Sophistication latent variable, all the six SCMs showed a significant difference between at-risk and not-at-risk students except for SYNTSTRUTa and SYNSTRUTt.

Tables 13 to 16 show the results of the analysis between females and males. For the Sentence Pattern latent variable, of the 12 SCMs DRNP, DRNEG, DRGERUND, DRINF, WRDNOUN, and WRDADV showed significant differences between females and males. For the Sentence Length latent variable, none of the three SCMS showed a significant difference between females and males. For the Sentence Connector latent variable, none of the seven SCMS showed a significant difference between females and males. For the Sentence

Sophistication latent variable, all the six SCMs showed a significant difference between females and males students except for SYNNP and SYNTSTRUTa.

Table 9

*Participant Unstandardized SCM Scores Between At-Risk and Not-At-Risk Students for Sentence Pattern Latent Variable (*N *= 1,029)*

SCMs	At-Risk		Not-At-Risk			
	M	*SD*	*M*	*SD*	*t*	*p*
DRNP	339.57	36.24	326.96	31.44	3.57	.001*
DRVP	291.63	37.04	286.27	32.36	1.48	.141
DRAP	35.90	16.42	38.35	12.13	-1.55	.122
DRPP	73.19	24.74	74.82	19.15	-0.68	.497
DRPVAL	3.19	5.61	4.48	4.56	-2.35	.020*
DRNEG	13.70	10.97	15.29	10.09	-1.49	.138
DRGERUND	19.20	15.51	21.14	11.28	-1.30	.198
DRINF	27.13	14.45	25.29	10.98	1.31	.191
WRDNOUN	198.64	40.42	196.53	34.44	0.54	.589
WRDVERB	137.59	28.44	138.39	21.19	-0.31	.757
WRDADJ	60.45	21.27	60.33	17.88	0.06	.952
WRDADV	61.61	22.81	67.07	17.87	-2.48	.014*

Note: SCMs = Sentence Complexity Measures; DRNP = Noun Phrase Incidence: DRVP Verb Phrase Incidence = DRAP Adverbial Phrase Incidence; = DRPP = Preposition Phrase Incidence; DRPVAL = Agentless Passive Voice Forms Incidence; DRNEG = Negation Expression Incidence; DRGERUND = Gerund Incidence; DRINF = Infinitive Incidence; WRDNOUN = Noun Incidence; WRDVERB = Verb Incidence; WRDADJ = Adjective Incidence; WRDADV = Adverb Incidence; * = $p < .05$

Table 10

Participant Unstandardized SCM Scores Between At-Risk and Not-At-Risk Students for Sentence Length Latent Variable (N = 1,029)

	At-Risk		Not-at-Risk			
SCMs	*M*	*SD*	*M*	*SD*	*T*	*p*
DESSL	24.00	11.33	11.33	7.20	2.67	0.008*
DESSLd	11.41	5.70	5.70	5.57	1.85	0.065
SYNLE	4.28	3.75	3.75	1.88	-0.17	0.859

Note. DESSL = Mean Number of Words; DESSLd = Standard Deviation of Mean Number of Words; SYNLE = Mean Number of Words Before Main Verb; * = $p < .05$

Table 11

Participant Unstandardized SCM Scores Between At-Risk and Not-At-Risk Students for Sentence Connector Latent Variable (*N* = 1,029)

SCMs	At-Risk		Not-At-Risk			
	M	*SD*	*M*	*SD*	*T*	*p*
CNCAll	106.15	25.37	100.82	19.17	2.18	.003*
CNCCaus	37.72	17.09	30.76	11.02	1.76	.000*
CNCLogic	62.28	19.81	58.92	16.88	-1.39	.008
CNCADC	14.50	10.36	15.93	8.80	-2.25	.165
CNCTemp	12.58	11.26	15.03	7.94	0.85	.026*
CNCTempx	15.28	9.30	14.46	9.22	1.88	.399
CNCAdd	53.46	18.97	50.04	13.94	1.76	.062

Note. CNCAll = All Connectives Incidence; CNCCaus = Causal Connectives Incidence; CNCLogic = Logical Connectives Incidence; CNCADC = Adversative/Contrastive Connectives Incidence; CNCTemp = Temporal Connectives Incidence; CNCTempx = Expanded Temporal Connectives Incidence; CNCAdd = Additive Connectives Incidence; * = $p < .05$.

Table 12

Participant Unstandardized SCM Scores Between At-Risk and Not-At-Risk Students for Sentence Sophistication Latent Variable (N = 1,029)

SCMs	At-Risk *M*	At-Risk *SD*	Not-at-Risk *M*	Not-at-Risk *SD*	*T*	*p*
SYNNP	0.56	0.14	0.63	0.13	-4.42	.001*
SYNMEDpos	0.64	0.05	0.65	0.04	-2.26	.026*
SYNMEDwrd	0.84	0.05	10.87	0.04	-4.25	.001*
SYNMEDlem	0.82	0.05	0.84	0.04	-3.57	.001*
SYNSTRUTa	0.09	0.04	0.09	0.03	-0.30	.765
SYNSTRUTt	0.09	0.04	0.09	0.03	-0.002	.998

Note. SYNNP = Mean Number of Modifiers Per Noun Phrase; SYNMEDpos = Minimal Edit Distance, Part of Speech; SYNMEDwrd = Minimal Edit Distance, All Words; SYNMEDlem = Minimal Edit Distance, Lemmas; SYNSTRUTa = Mean Adjacent Sentence Structure Similarity; SYNSTRUTt = Mean All Sentence Structure Similarity; * = $p < .05$.

The descriptive analysis of standardized SCMs by sex showed that all the SCMS for the four latent variables showed significant differences by sex at $\alpha = .05$. Tables 13 - 16 provide the detailed results of the descriptive analysis of the standardized SCM scores by sex for the four latent variables.

Table 13

Participant Unstandardized SCM Scores Between Females and Males for Sentence Pattern Latent Variable (N = 1,029)

SCMs	Female		Male		*T*	*p*
	M	*SD*	*M*	*SD*		
DRNP	323.90	31.33	331.36	32.90	-3.72	.001*
DRVP	287.21	32.41	286.52	33.51	0.34	.736
DRAP	38.46	11.78	37.70	13.59	0.95	.343
DRPP	74.06	18.38	75.23	21.29	-0.94	.347
DRPVAL	4.12	4.28	4.55	5.11	-1.45	.146
DRNEG	16.31	10.30	13.88	9.96	3.85	.001*
DRGERUND	21.82	11.17	19.98	12.43	2.50	.013*
DRINF	26.20	10.76	24.77	12.03	2.01	.045*
WRDNOUN	193.35	34.66	200.30	35.28	-3.18	.002*
WRDVERB	138.77	21.52	137.91	22.76	0.62	.536
WRDADJ	61.02	17.77	59.63	18.80	1.22	.223
WRDADV	68.50	17.74	64.34	19.16	3.61	.001*

Note. Female students (n = 527); Male Students (n = 502); DRNP = Noun Phrase Incidence: DRVP Verb Phrase Incidence = DRAP Adverbial Phrase Incidence; = DRPP = Preposition Phrase Incidence; DRPVAL = Agentless Passive Voice Forms Incidence; DRNEG = Negation Expression Incidence; DRGERUND = Gerund Incidence; DRINF = Infinitive Incidence; WRDNOUN = Noun Incidence; WRDVERB = Verb Incidence; WRDADJ = Adjective Incidence; WRDADV = Adverb Incidence; * = p < .05.

Table 14

Participant Unstandardized SCM Scores Between Females and Males for Sentence Length Latent Variable (N = 1,029)

SCMs	Female *M*	Female *SD*	Male *M*	Male *SD*	*T*	*p*
DESSL	21.31	7.42	21.56	8.24	-0.53	.598
DESSLd	10.47	5.33	10.50	5.85	0.09	.925
SYNLE	4.42	2.25	4.25	2.07	1.33	.083

Note. N = 1,029. Female students (n = 527); Male Students (n = 502); DESSL = Mean Number of Words; DESSLd = Standard Deviation of Mean Number of Words; SYNLE = Mean Number of Words Before Main Verb; * = $p < .05$.

Table 15

Participant Unstandardized SCM Scores Between Females and Males for Sentence Connector Latent Variable (N = 1,029)

SCMs	Female *M*	Female *SD*	Male *M*	Male *SD*	*T*	*p*
CNCAll	101.33	18.48	101.48	21.53	-0.12	.906*
CNCCaus	31.32	11.51	31.83	12.58	-0.67	.502
CNCLogic	58.61	15.82	59.97	18.62	-1.25	.210
CNCADC	16.27	8.51	15.22	9.47	1.87	.062
CNCTemp	14.81	7.68	14.69	9.12	0.24	.813
CNCTempx	14.37	8.50	14.82	10.00	-0.78	.436
CNCAdd	51.19	13.80	49.56	15.43	1.79	.074

Note. N = 1,029. Female students (n = 527); Male Students (n = 502); CNCAll = All Connectives Incidence; CNCCaus = Causal Connectives Incidence; CNCLogic = Logical Connectives Incidence; CNCADC = Adversative/Contrastive Connectives Incidence; CNCTemp = Temporal Connectives Incidence; CNCTempx = Expanded Temporal Connectives Incidence; CNCAdd = Additive Connectives Incidence * = $p < .05$.

Table 16

Participant Unstandardized SCM Scores Between Females and Males for Sentence Sophistication Latent Variable (N = 1,029)

SCMs	Female		Male		T	p
	M	*SD*	*M*	*SD*		
SYNNP	0.62	0.13	0.62	0.13	-0.63	.528
SYNMEDpos	0.66	0.03	0.65	0.04	5.30	.001*
SYNMEDwrd	0.87	0.04	0.86	0.04	3.99	.001*
SYNMEDlem	0.84	0.04	0.83	0.04	3.75	.001*
SYNSTRUTa	0.09	0.03	0.09	0.03	-1.72	.085
SYNSTRUTt	0.08	0.03	0.09	0.03	-2.04	.042*

Note. Female students (n = 527); Male Students (n = 502); SD = Standard Deviation; SYNNP = Mean Number of Modifiers Per Noun Phrase; SYNMEDpos = Minimal Edit Distance, Part of Speech; SYNMEDwrd = Minimal Edit Distance, All Words; SYNMEDlem = Minimal Edit Distance, Lemmas; SYNSTRUTa = Mean Adjacent Sentence Structure Similarity; SYNSTRUTt = Mean All Sentence Structure Similarity; * = $p < .05$.

The following section presents the findings for RQ1 and RQ2.

RQ1: Is the hypothesized model based on 28 Coh-Metrix SCMs as indicators of four latent variables a good fit using the eighth-grade, automatically scored formative assessment data for argumentative writing*?* The four latent variables are Sentence Pattern, Sentence Length, Sentence Connector, and Sentence Sophistication.

Initial Hypothesized Model

Using the entire sample of 1,029 essays, four latent variables and 28 SCMs (see Table 1), a Confirmatory Factor Analysis (CFA) was performed to test the goodness-of-fit of the hypothesized model. CFA is an objective test of a theoretical model that tests the hypothesis if a relationship exists between the four latent variables and the observed variables (28 SCMs).

The relationship pattern was postulated *a priori* before the hypothesis was tested statistically (Perry, Nicholls, Clough, Crust, 2015). Because of the *a priori* specification that must be made, CFA is a deductive process that allows hypothesized models to be tested (Meyers, & Guarino, 2006). Figure 4 provides a visual depiction of the initial hypothesized model.

Four key CFA fit statistics were used to test whether the model was a good fit. One commonly used CFA metric, chi-square value, was not reported because this statistic is sensitive to sample size. For models based on large sample sizes (400 or more), the chi-square value is almost always statistically significant (Bentler & Bonnet, 1980; Jöreskog & Sörbom, 1993) and would reject the hypothesis of a good model fit (Perry et al., 2015). For this reason, other CFA fit statistics were used.

The goodness of fit index (GFI) is a measure of fit between the hypothesized model and the observed covariance matrix. The GFI ranges between 0 and 1, with a value of over 0.90 generally indicating an acceptable model fit. The root mean square error of approximation (RMSEA) avoids issues of sample size by analyzing the discrepancy between the hypothesized model with optimally chosen parameter estimates and the population covariance matrix. The RMSEA ranges from 0 to 1, with smaller values indicating better model fit. A value of 0.06 or less is indicative of acceptable model fit. The standardized root mean square residual (SRMR) is the square root of the discrepancy between the sample covariance matrix and the model covariance matrix. The SRMR ranges from 0 to 1, with a value of 0.08 or less being indicative of an acceptable model. The comparative fit index (CFI) analyzes the model fit by examining the discrepancy between the data and the hypothesized model, while adjusting for the issues of sample size inherent in the chi-squared test of model fit, and the normed fit index. CFI values

range from 0 to 1, with larger values indicating better fit. A CFI value of 0.95 or higher is presently accepted as an indicator of good fit (Hu & Bentler, 1999).

The results in Table 17 based on the four key CFA fit statistics clearly demonstrated that all the four CFA fit statistics (GFI, RMSEA, SRMR and CFI) did not meet the criteria of a good fit.

Table 17

Key Fit Statistics of the Initial Hypothesized CFA Model (N = 1,029)

CFA Fit Statistics	Minimum Criteria	Value
Goodness of Fit Index (GFI)	>0.90	0.75
Root Mean Square Error of Approximation (RMSEA	<0.06	0.12
Standardized Root Mean Square Residual (SRMR)	<0.08	0.11
Comparative Fit Index (CFI)	>0.95	0.51

Note. N = 1,029. Initial model = 28 SCMs. Criteria for a well-fitted model: GFI > 0.95. CFI > 0.95. RMSEA < 0.06. SRMR < 0.08.

The CFA did not show a good fit for two reasons. First, SCMs in one latent variable may have been highly correlated with SCMs in other latent variables. The correlation matrix for all 28 SCMs in Appendix B shows that several SCMs from different latent variables were highly correlated with each other. Second, some SCMs within a latent variable could improve the CFA model fit when they were combined and not analyzed separately. Thus, several revisions were made to the initial hypothesized model.

Revision to Initial Hypothesized Model

The initial model was revised by examining the relationships between SCMs from different latent variables using the SCM correlation matrix and factor loadings in the CFA

model. The revised model involved either removing or combining SCMs in the initial model to obtain a better fit.

Removed SCMs. First, the DESSL (number of words in a sentence) SCM in the Sentence Length latent variable was removed because it was highly correlated with two SCMs, SYNSTRUTa (similar sentence structures between adjacent sentences) and SYNSTRUTt (similar sentence structures between all sentences) in the Sentence Sophistication latent variable with correlation coefficients of r = -0.56 and -0.62 respectively. A good CFA model would require SCM measures within a latent variable to be correlated, but SCMs between latent variables to have low correlations. This would ensure that the latent variables are distinct and that each latent variable incorporates and describes appropriate measures. Also, Hunt (1970) claimed that counting the number of words in a sentence is inconsequential because it will only provide information on the length of the text and not its syntactic complexity. Next, the DRNEG (incidence score of negation) SCM in the Sentence Pattern was removed because it did not contribute to explaining Sentence Pattern latent variable due to a low negative factor loading of -0.16.

One more SCM that was removed from the Sentence Sophistication latent variable was SYNNP (mean number of modifiers per noun phrase) because it had a low factor loading of 0.01 in the CFA model. Two SCMs, CNCADC (incidence score of adversative and contrastive connectives) and CNCTempx (incidence score of expanded temporal connectives) in the Sentence Connector latent variable were removed because of low factor loadings (-0.05 and 0.02 respectively). Another SCM, CNCADD (incidence score of additive connectives), in the Sentence Connector latent variable was removed because of a low factor loading of -0.05 in the CFA model. Because the initial hypothesized model was not an acceptable fit, the SCMs were

examined to discover if any of the SCMs captured similar measures of syntactic complexity. Based on this, two SCMs, DRNP (incidence score of noun phrase) and DRVP (incidence score of verb phrase) in the Sentence Pattern latent variable were removed because they can be captured by the SCMs, WRDNOUN (incidence score of nouns) and WORDVERB (incidence score of verbs) in the Sentence Pattern latent variable.

Combined SCMs. Four SCMs from the Sentence Pattern latent variable, WRDNOUN, WRDVERB, WRDADJ, and WRDADV, were combined into a single SCM labelled WORD. Because these four SCMs of the Sentence Pattern latent variable were all related measures, WORD was created by taking the average of these four SCMs instead of analyzing them separately. The WORD SCM was important to explain complex sentence structures because more mature and skillful writers produce sentences that contain a greater number of linguistic features such as grammatical word classes that are related to complex sentence structures (Hunt, 1970; McNamara et al., 2011). By averaging the four Sentence Pattern SCMs related to WORD instead of eliminating any of the SCMs, the impact of all the four SCMs was captured in the single combined WORD SCM. Similarly, two related SCMs from the Sentence Pattern latent variable, DRAP (adverb phrase) and DRPP (prepositional phrase) were combined into a single SCM labeled PHRASE. These two SCMs can be combined because they are closely related. Also, prepositional phrases can function as adverb phrases. Adverb phrases alone will not capture propositional phrases that do not function as adverb phrases. Therefore, it made sense to combine the two.

After the targeted SCMs were either removed or combined, CFA was used to estimate a model with 16 indicators to create the final hypothesized model. Table 18 shows the final hypothesized CFA model with the 16 SCMs.

Table 18

Final Hypothesized CFA Model

Sentence Pattern	Sentence Length	Sentence Connector	Sentence Sophistication
WORD WRDNOUN WRDVERB WRDADJ WRDADV	Standard deviation of mean number of words (DESSLd)	All connectives incidence (CNCAll)	Minimal edit distance, part of speech (SYNMEDpos)
Agentless passive voice forms incidence (DRPVAL)	Mean number of words before main verb (SYNLE)	Causal Connectives incidence (CNCCaus)	Minimal edit distance, all words (SYNMEDwrd)
Gerund incidence (DRGERUND)		Logical connectives incidence (CNCLogic)	Minimal edit distance, lemmas (SYNMEDlem)
Infinitive incidence (DRINF)		Temporal connectives incidence (CNCTemp)	Mean adjacent sentence structure similarity (SYNSTRUTa)
PHRASE DRAP DRPP			Mean all sentence structure similarity (SYNSTRUTt)

Note. A total of 16 SCMS. New SCMs: WORD = WRDNOUN + WRDVERB + WRDADJ + WRDADV; PHRASE = DRAP + DRPP.

Final Hypothesized Model

The final CFA model reduced the initial 28 SCMs into 16 SCMs by combining or removing the SCMs. The four latent variables in the initial hypothesized model remained the same. The factor loadings for the final model are shown in Table 19. Most of the SCMs showed significant factor loadings (p = .05) with the exception of DRPVAL which had been retained because passive voice sentence construction indicates a varied and more complex sentence structure (McNamara et al., 2011).

Table 19

Standardized Factor Loading Matrix for Final Hypothesized Model (N = 1,029)

	Estimate	SE	*T*	*p*
Sentence Pattern				
DRPVAL	-0.04	0.05	-0.86	.39
DRGERUND	0.20	0.05	4.33	<.001
DRINF	-0.46	0.05	-9.50	<.001
WORD	0.52	0.05	10.38	<.001
PHRASE	0.43	0.05	9.25	<.001
Sentence Length				
DESSLd	0.61	0.05	12.81	<.001
SYNLE	0.54	0.05	11.96	<.001
Sentence Connectors				
CNCAll	1.03	0.02	43.74	<.001
CNCCaus	0.38	0.03	13.02	<.001
CNCLogic	0.70	0.02	29.95	<.001
CNCTemp	0.36	0.03	12.27	<.001
Sentence Sophistication				
SYNMEDpos	-0.09	0.04	-1.98	.048
SYNMEDwrd	-0.27	0.04	-6.29	<.001
SYNMEDlem	0.40	0.04	9.63	<.001
SYNSTRUTa	0.64	0.05	12.94	<.001
SYNSTRUTt	-0.48	0.04	-10.74	<.001

Note. DRPVAL = Agentless passive voice forms incidence; DRGERUND = Gerund incidence; DRINF = Infinitive Incidence; WORD = WRDNOUN (Noun Incidence) + WRDVERB (Verb Incidence) + WRDADJ (Adjective Incidence) + WRDADV (Adverb Incidence); PHRASE = DRAP (Adverb Phrase) + DRPP (Prepositional Phrase); DESSLd = Standard deviation of mean number of words; SYNLE = Mean number of words before main verb; CNCAll = All connectives incidence; CNCCaus = Causal Connectives incidence; CNCLogic Logical connectives incidence; CNCTemp = Temporal connectives incidence; SYNMEDpos = Minimal Edit Distance, Part of Speech; SYNMEDwrd = Minimal Edit Distance, All Words; SYNMEDlem = Minimal Edit Distance, Lemmas; SYNSTRUTa = Mean Adjacent Sentence Structure Similarity; SYNSTRUTt = Mean All Sentence Structure Similarity

The CFA metrics, GFI, RMSEA, and SRMR, all met the minimum criteria for a good model fit. The only index value that did not exceed the corresponding criterion was CFI with the value of 0.70, which was below the minimum criteria of 0.95. The lower than acceptable CFI score can be explained if most of the correlations between SCMs are close to zero (Fan, Thompson, & Wang, 1999). This implied less covariance to explain, which reduced the effectiveness of the CFI in the eighth-grade data automatically scored formative assessment data for argumentative writing. The other three CFA metrics were less impacted by the low correlations between SCMs. The distribution of the SCM correlations shown in Table 20 indicated that 85% of the SCMs had correlations between $r = -0.2$ and 0.2. These low correlations explain the lower than acceptable CFI value.

It is important to note that the criteria for a good CFA model is an aggregate criterion and not based on individual factor loading criteria for each SCM. Even if some factor loadings are close to zero, they can be included in the model as long as the criteria for the aggregate CFA metrics were met.

Table 20

Distribution of the Correlation Coefficients of the 16 SCMs (*N* = 1,029)

Correlations Coefficient	Frequency	Percentage
< -0.4	0	0%
-0.4 – -0.2	7	6%
-0.2 – 0	61	51%
0 – 0.2	41	34%
0.2 – 0.4	10	8%
0.4 – 0.8	1	1%
> 0.8	0	0%
Total	120	100%

Note. 85% of the correlation coefficients are between -0.2 and 0.2.

Taken together, these results suggest that the final model provided a reasonable fit for the data; therefore, the revised model was selected as the final CFA model. Table 21 summarizes the key metrics of the final CFA model.

Table 21

Key Fit Statistics of the Final CFA Model (*N* = 1,029)

CFA Fit Statistics	Minimum Criteria	Value
Goodness of Fit Index (GFI)	>0.90	0.95
Root Mean Square Error of Approximation (RMSEA	<0.06	0.05
Standardized Root Mean Square Residual (SRMR)	<0.08	0.07
Comparative Fit Index (CFI)	>0.95	0.70

The four latent variables correlation matrix for the final CFA model in Table 22 shows only moderate correlations between the latent variables with values ranging from -0.5 to 0.5. This demonstrates that the latent variables were distinct from each other.

Table 22

Latent Factor Correlation Matrix (N = 1,029)

	1	2	3	4
1. Sentence Pattern	-			
2. Sentence Length	-0.32*	-		
3. Sentence Connector	-0.01	0.40*	-	
4. Sentence Sophistication	-0.06	-0.03	-0.07	-

Note: Final CFA Model = 16 SCMs. All correlations indicated by * are statistically significant at $p < .001$.

Figure 5 shows the final hypothesized CFA model with the factor loadings and correlations between the latent variables. All the SCMs were significantly correlated at $\alpha = 0.05$ with their respective latent variables except for DRPVAL with Sentence Pattern.

The final hypothesized model with 16 SCMs and the same four latent variables significantly improved the goodness of fit compared to the initial hypothesized model with 28 SCMs. Therefore, the final hypothesized model supported RQ1 that the four latent variables and 16 SCMs was a good fit for the eighth-grade, automatically scored formative assessment data for argumentative writing.

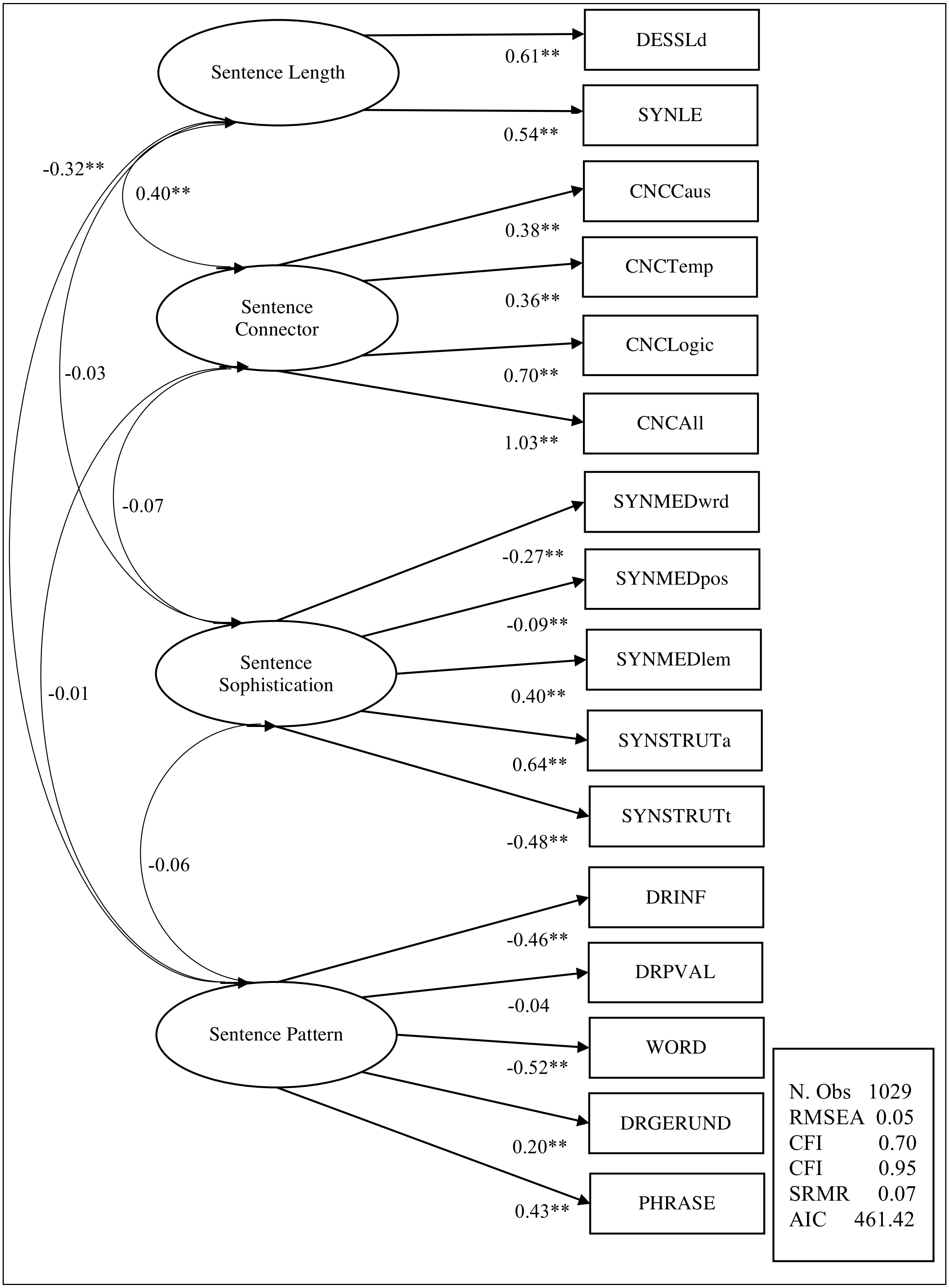
Sentence Length
Sentence Connector
Sentence Sophistication
Sentence Pattern
DESSLd
SYNLE
CNCCaus
CNCTemp
CNCLogic
CNCAll
SYNMEDwrd
SYNMEDpos
SYNMEDlem
SYNSTRUTa
SYNSTRUTt
DRINF
DRPVAL
WORD
DRGERUND
PHRASE
0.61**
0.54**
0.38**
0.36**
0.70**
1.03**
-0.27**
-0.09**
0.40**
0.64**
-0.48**
-0.46**
-0.04
-0.52**
0.20**
0.43**
-0.32**
0.40**
-0.03
-0.07
-0.01
-0.06
N. Obs 1029
RMSEA 0.05
CFI 0.70
CFI 0.95
SRMR 0.07
AIC 461.42

Figure 5. A graphical represenation of four latent variables and 16 SCMs using CFA. CFA = confirmatory factor analysis; N. Obs = number of observations; RMSEA = root mean square error of approximation; CFI = comparative fit index; GFI = goodness-of-fit index; SRMR = standardized root mean square residual. All factor $\rightarrow$ indicator paths are significant at $\alpha = 0.05$ with the exception of DRPVAL. ** refers to factor loadings that were significant.

The scores of the four latent variables for each essay were computed using the factor score regression coefficients in Table 23 obtained from the CFA. The score for a latent variable is simply a linear combination of the product of the SCM value with the associated factor score regression coefficient.

The four latent variables comprising the 16 SCMs in the final hypothesized model using CFA in RQ1 were the predictor variables included in RQ2.

(RQ2) Do the four latent variables using the 16 Coh-Metrix SCMs show a relationship with writing quality, and how does the relationship vary between Student Type (at risk and not-at-risk students) using the eighth-grade, automatically scored formative assessment data for writing?

Initial Multiple Linear Regression Model

A multiple linear regression (MLR) model was developed to analyze the second research question. The dependent variable in the MLR was the writing score for the 1,029 essays. The four independent variables in the MLR comprised the standardized scores of the latent variables: Sentence Pattern, Sentence Length, Sentence Connector, and Sentence Sophistication for each essay. In addition, an indicator variable was created, with zero representing at-risk students, and one representing not-at-risk students. The initial MLR model incorporated all possible two-way interactions between the four latent variables and all possible two-way interactions between the latent variables and Student Type (at-risk and not-at-risk). Table 24 shows results from the analysis of variance (ANOVA) of the initial MLR model.

Table 23

Factor Scores Regression Coefficients (*N* = 1,029)

	Latent Variables			
SCMs	Sentence Pattern	Sentence Length	Sentence Connector	Sentence Sophistication
DESSLd	-0.10	0.47	-0.01	-0.00
CNCAll	0.04	0.26	1.20	-0.04
CNCCaus	0.00	-0.01	-0.03	0.00
CNCLogic	0.00	-0,02	-0.09	0.00
CNCTemp	0.00	0.01	-0.03	0.00
SYNLE	-0.07	0.35	-0.01	-0.00
DRPVAL	-0.02	0.00	0.00	0.00
DRGERUND	0.11	-0.02	-0.00	-0.00
DRINF	-0.31	0.06	0.00	0.00
SYNSTRUTa	-0.02	0.00	0.00	0.48
SYNSTRUTt	0.01	0.00	-0.00	-0.29
SYNMEDlem	0.01	0.00	0.00	0.22
SYNMEDwrd	0.00	0.00	-0.00	-0.14
SYNMEDpos	0.00	0.00	-0.00	-0.04
WORD	1.01	-0.18	-0.00	-0.03
PHRASE	0.43	-0.08	-0.00	-0.01

Note. *N* =1,029. Four Latent Variables with 16 SCMs.

Table 24

Analysis of Variance of the Initial Multiple Regression Model (*N* = 1,029)

Source	*df*	*SS*	*MS*	*F*	*p*
Model	15	5,291.63	352.78	30.95	<.0001
Error	1,013	11,544.82	11.40		
Corrected Total	1,028	16,836.45			

The significant *p* value for the *F* statistics ANOVA result implied that the regression model using the four latent variables, the Student Type indicator variable, and the interactions between these variables produced a better fit to explain writing quality over the mean of the writing scores.

In the initial MLR model, which considered all possible two-way interactions between the four latent variables and the latent variables with Student Type, several of the regression coefficients turned out to be not significant at $\alpha = 0.05$. Table 25 shows the regression coefficients and *p* values for the initial multiple regression model.

The coefficients of the three latent variables, Sentence Pattern, Sentence Length, and Sentence Connector, were positive. However, only one latent variable, Sentence Length, in the MLR model, was significant at $\alpha = 0.01$ which implied that it contributed positively toward increasing writing scores. Sentence Sophistication had a negative coefficient, but it was not significant. Therefore, no implications on writing quality can be made. The Student Type indicator variable was highly significant with $p < .001$, which implied that the impact of the latent variables on writing scores in the MLR model varied by Student Type. The MLR model produced an R^2 of 0.31, and this implied that the four latent variables, the indicator variables, and all the two-way interactions between the variables explain 31% of the variability in the writing scores. This was a reasonable R^2 value because syntactic complexity is just one of many factors, including overall content and organizational structure that impact writing quality. The initial MLR model supported the RQ2 hypothesis that the four latent variables of syntactic complexity did not show a clear relationship to writing quality, and that the relationship varied by Student Type. In particular, the initial model had several two-way interaction variables that were not significant (e.g., Sentence Length*Not-At-Risk, Sentence Connector*Sentence

Sophistication). Based on these limitations, revisions were made to the initial multiple linear regression model.

Table 25

Regression Coefficients for the Initial Multiple Regression Model (*N* = 1,029)

Parameter	Estimate	Standard Error	*T*	*p*
Intercept	17.18	0.34	50.17	.001
Not-At-Risk	5.25	0.35	15.16	.001
Sentence Length	0.78	0.47	1.65	.010
Sentence Connector	0.29	0.30	0.98	.328
Sentence Pattern	0.83	0.38	2.16	.031
Sentence Sophistication	-0.09	0.41	-0.22	.826
Sentence Length*Sentence Pattern	0.33	0.21	1.57	.117
Sentence Length*Sentence Connector	-0.39	0.13	-3.01	.003
Sentence Length*Sentence Sophistication	-0.23	0.27	-0.85	.395
Sentence Connector*Sentence Pattern	-0.17	0.15	-1.16	.247
Sentence Connector*Sentence Sophistication	0.09	0.16	0.56	.576
Sentence Pattern*Sentence Sophistication	0.24	0.22	1.07	.285
Sentence Length*Not-At-Risk	-0.30	0.50	-0.59	.553
Sentence Pattern*Not-At-Risk	1.17	0.42	2.76	.006
Sentence Connector*Not-At-Risk	-0.91	0.32	-2.81	.005
Sentence Sophistication*Not-At-Risk	0.51	0.44	1.15	.250

Note. t = t-statistic value. p = p value of t-statistic. $R^2 = 0.31$.

Final Multiple Linear Regression Model

A final MLR model was developed to refine the initial model by removing several non-significant two-way interaction variables and only analyzed the three interactions (Sentence

Length*Sentence Connector, Sentence Pattern*Student Type, Sentence Connector*Student Type). Table 26 shows results from the analysis of variance (ANOVA) of the final MLR model.

Table 26

Analysis of Variance of the Final Multiple Regression Model (N = 1,029)

Source	*df*	*SS*	*MS*	*F*	*p*
Model	9	4975.52	552.84	47.50	<.0001
Error	1019	11860.92	11.64		
Corrected Total	1028	16836.45			

The significant p value for the F statistics indicated that using the independent variables in the MLR model provided a better prediction of the essay-writing quality of eighth-grade students compared to the mean of the writing scores.

The large sample size exceeding 1,000 and the relatively few (eight) predictor variables and interactions generated a power of one for the test. This implied there was a strong probability that the conclusion reached from the F test was correct. The results of the final MLR model are shown in Table 27.

Table 27

Regression Coefficients for the Final Multiple Regression Model (N = 1,029)

Parameter	Estimate	Standard Error	*t*	*p*
Intercept	17.03	0.33	51.40	<0.001
Student Type	5.31	0.34	15.44	<0.001
Sentence Length	0.43	0.20	2.14	0.030
Sentence Connector	0.44	0.25	1.79	0.070
Sentence Pattern	0.73	0.33	2.19	0.030
Sentence Sophistication	0.35	0.14	2.46	0.010

Sentence Length*Sentence Connector	-0.41	0.11	-3.71	<0.001
Sentence Pattern*Student Type	1.22	0.36	3.42	<0.001
Sentence Connector*Student Type	-1.08	0.26	-4.10	<0.001

Note. $R^2 = 0.31$

In the final model, all four latent variables had positive regression coefficients and all were significant at $\alpha = 0.1$. Three of the four latent variables had $p < .05$ with Sentence Connector having $p = .07$. All the two-way interactions had p-values less than .001. The negative regression coefficient of 0.41 for the two-way interaction of (Sentence Length) * (Sentence Connector) had an interesting interpretation. While Sentence Length and Sentence Connector individually contributed to increasing writing quality, long sentences with several connectors could have a negative impact (Hunt, 1970; McNamara et al., 2011). The interactions between student type and latent variables indicated that increased use of Sentence Pattern by not-at-risk students had a positive impact on writing quality compared to at-risk students. However, increased use of Sentence Connector by not-at-risk students had a negative impact on writing quality compared to that of at-risk students.

All the other key metrics in the MLR final model were similar to the MLR initial model. The R^2 for the final model remained at 31%.

Impact of Student Type on Writing Quality. The not-at-risk Student Type indicator was highly significant at $p < .001$, which implied that the impact of the latent variables on writing scores varied by Student Type. If the Student Type indicator was removed from the MLR model, the R^2 reduced to 0.14, which indicated that Student Type was a highly significant variable. It was interesting to see how R^2 changed when writing quality was analyzed using different sets of predictor variables. Table 28 shows these differences. The results indicated

that the final model based on four latent variables, Student Type, and interactions provided the best fit (in terms of R^2) to explaining writing quality compared to any subsets of predictor variables.

Table 28

Differences in R^2 Based on Different Sets of Predictor Variables (N = 1,029)

Predictor Variables	R^2
Only Student Type	0.18
Only Four Latent Variables	0.13
Only Four Latent Variables and Interactions (Sentence Length* Sentence Connector)	0.14
Final Model	0.31

A detailed analysis on the impact of changes in the latent variables on writing quality between students who are at-risk and not-at-risk was performed. The change in writing scores by Student Type was calculated when a single latent variable was changed and all the other latent variables were measured at their mean values. The changes ranged from two standard deviations below the mean to two standard deviations above the mean in increments of 0.5 standard deviations. The mean and standard deviations for each of the latent variables are reported in Table 29.

Table 29

Mean and Standard Deviations for Latent Variables (N = 1,029)

Latent Variable	Mean	SD
Sentence Pattern	0.00	0.73
Sentence Length	0.00	0.77
Sentence Connector	0.00	1.09
Sentence Sophistication	0.00	0.79

Tables 30 to 31 and Figures 6 to 7 show the change in writing scores by Student Type when a single latent variable was changed and all the other latent variables were measured at their mean values. The changes ranged from two standard deviations below the mean to two standard deviations above the mean in increments of 0.5 standard deviations. Table 32 and Figure 8 show the change in writing scores for the interaction of Sentence Length * Sentence Connector with Sentence Length fixed at three levels and Sentence Connector varying in increments of 0.5 standard deviations.

Sentence Pattern. To interpret Table 30, if the Sentence Pattern score increased by one standard deviation and the other latent variables stayed unchanged at their mean value, the writing score of at-risk students was predicted to increase by 3%. On the other hand, the writing score of not-at-risk students was predicted to increase by 6%.

Table 30

Impact of Changes in Sentence Pattern on Writing Quality (N = 1,029)

Standard Deviation of Sentence Pattern	At-Risk Writing Score	Ratio of Adjusted Score to Mean Score (%)	Not-At-Risk Writing Score	Ratio of Adjusted Score to Mean Score (%)
-2	15.96	94	19.48	87
-1.5	16.23	95	20.20	90
-1	16.50	97	20.91	94
-0.5	16.76	98	21.63	97
0	17.03	100	22.34	100
0.5	17.30	102	23.05	103
1	17.56	103	23.77	106
1.5	17.83	105	24.48	110
2	18.10	106	25.20	113

Note: The mean at-risk writing score of 17.03 represents all latent variables having a mean score of zero. The ratio of adjusted score to mean score of 94 for at-risk students equals 15.96/17.03.

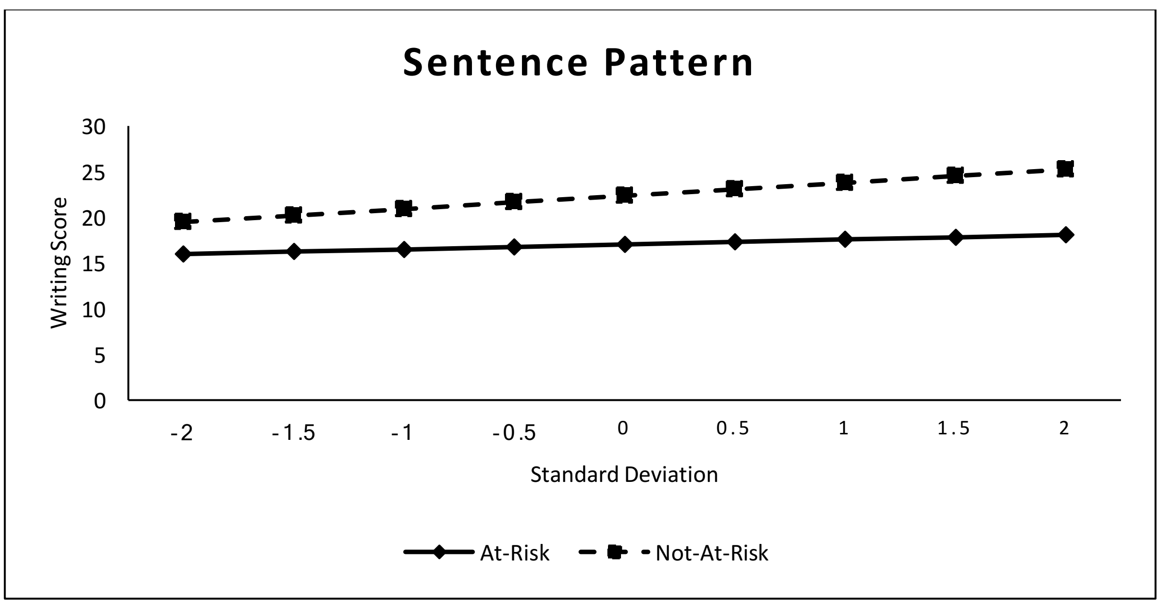

Figure 6. Impact of Changes in Sentence Pattern on Writing Score by Student Type

Sentence Connector.

Table 31

Impact of Changes in Sentence Connector on Writing Quality

Standard Deviation of Sentence Connector	At-Risk	Ratio of Adjusted Score to Mean Score (%)	Not-At-Risk	Ratio of Adjusted Score to Mean Score (%)
-2 SD	16.07	94	23.74	106
-1.5 SD	16.31	96	23.39	105
-1 SD	16.55	97	23.04	103
-0.5 SD	16.79	99	22.69	102
0 SD	17.03	100	22.34	100
0.5 SD	17.27	101	21.99	98
1 SD	17.51	103	21.64	97
1.5 SD	17.75	104	21.29	95
2 SD	17.99	106	20.94	94

Note: The mean at-risk writing score of 17.03 represents all latent variables having a mean score of zero. The ratio of adjusted score to mean score of 94 for at-risk students equals 16.07/17.03.

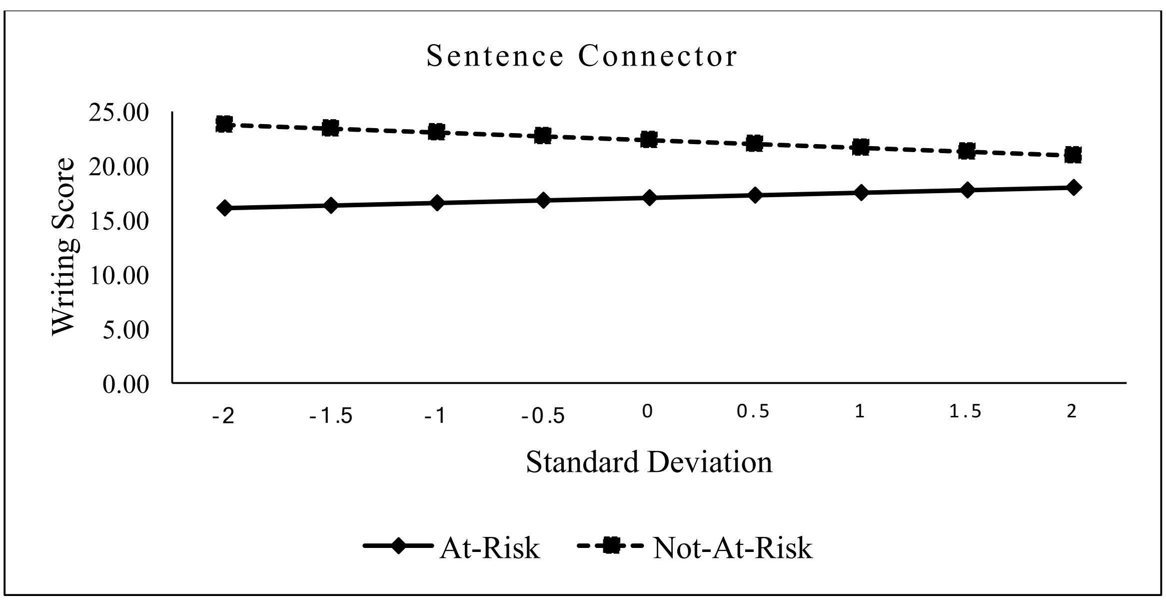

Figure 7. Impact of Changes in Sentence Connector on Writing Score by Student Type

To interpret Table 31, increased use of Sentence Connector increased writing score for at-risk students but decreased writing scores for not-at-risk students. Specifically, if the Sentence Connector score increased by one standard deviation, and the other latent variables stayed unchanged at their mean value, then the writing score of at-risk students was predicted to increase by 3%. On the other hand, the writing score of not-at-risk students was predicted to decrease by 3%.

Sentence Length * Sentence Connector interaction. To interpret the interaction effect between Sentence Length * Sentence Connector, the value of Sentence Length was fixed at three-levels: high, medium, and low. The values for high Sentence Length was fixed at one standard deviation above the mean, medium Sentence Length was fixed at the mean, and low Sentence Length was fixed at one standard deviation below the mean. For each fixed level of Sentence Length, a graph of writing scores was plotted with Sentence Connector ranging from below two standard deviations to above two standard deviations in increments of half a standard deviation. Table 32 and Figure 8 show the graphs of writing scores for all students with Sentence Length fixed at three levels.

To interpret Figure 8, for a fixed value of Sentence Length, writing scores generally increased as Sentence Connector scores increased. Specifically, for students with low Sentence Connector use (e.g., -2 SD for Sentence Connector), high Sentence Length with low Sentence Connector generated higher writing scores than low Sentence Length with low Sentence Connector. However, as the use of Sentence Connector increased (e.g., +2 SD), then there was no impact of Sentence Length on writing score.

Table 32

Impact of Changes in Writing Score for a fixed Sentence Length and Varying Values of Sentence Connector (N = 1,029)

Standard Deviation of Sentence Length	High Length	Ratio of Adjusted Score to Mean Score (%)	Medium Length	Ratio of Adjusted Score to Mean Score (%)	Low Length	Ratio of Adjusted Score to Mean Score (%)
-2	16.70	0.96	16.07	0.94	15.44	0.92
-1.5	16.87	0.97	16.31	0.96	15.76	0.94
-1	17.03	0.98	16.55	0.97	16.07	0.96
-0.5	17.20	0.99	16.79	0.99	16.39	0.98
0	17.36	1.00	17.03	1.00	16.70	1.00
0.5	17.52	1.01	17.27	1.01	17.02	1.02
1	17.69	1.02	17.51	1.03	17.33	1.04
1.5	17.85	1.03	17.75	1.04	17.65	1.06
2	18.02	1.04	17.99	1.06	17.96	1.08

Note. The mean writing score of 17.36 for all students represents all latent variables having a mean score of zero. The ratio of adjusted score to mean score of 1.02 = 17.69/17.36 represents high length and standard deviation of Sentence Connector equal to one.

Overall, an increase in the four latent variables had a greater impact on at-risk eight grade students compared to not-at-risk students. While an increase in the use of an individual latent variable generated only a modest increase in writing scores, the combined effect of increasing all the latent variables by one standard deviation was predicted to increase writing scores for at-risk students by 8%. However, the same result for not-at-risk students generated an increase of only 4% in writing scores. On the other hand, a decline in the use of latent variables by one standard deviation from the average generated a 12% decline in writing scores for at-risk students, but only an 8% decline in writing scores for not at-risk students.

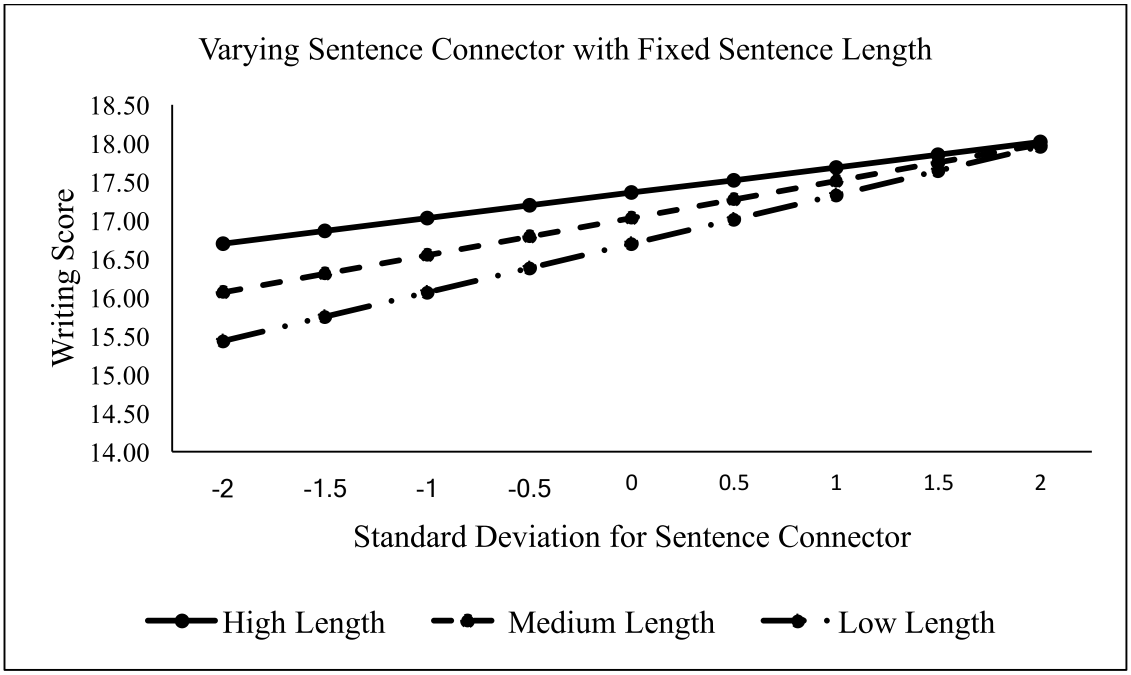

Figure 8. Impact of Changes in Writing Score When Sentence Length is Fixed with Varying Values of Sentence Connector

The final MLR model supported the hypothesis in RQ2 that the four latent variables that were confirmed using CFA showed a relationship with writing quality and the values of the predicted variables varied significantly by Student Type.

CHAPTER 5

DISCUSSION

The purpose of this study is twofold. First, the researcher wanted to develop a deeper understanding of SCMs by hypothesizing 28 selected SCMs to indicate four meaningful latent variables: Sentence Pattern, Sentence Length, Sentence Connectors, and Sentence Sophistication. Second, the relationship between the four latent variables and Student Type (at-risk and not-at-risk) with writing quality was analyzed.

To accomplish this, data on 1,029 eighth-grade, automatically scored argumentative essays with associated writing scores from a Benchmark Writing Assessment (BAS-Write) were analyzed. Eighth graders were selected because at this age students are expected to produce more sophisticated sentence structures and use several of the latent variables modeled in this dissertation study. The argumentative genre was chosen because it requires the use of syntactically more complex sentences to connect ideas and information in a logical manner. The use of a large dataset comprising at-risk and not-at-risk students allowed the analysis of several SCMs simultaneously and by Student Type. This overcame many of the limitations of earlier studies that examined fewer measures with small sample sizes.

The 28 SCMs from Coh-Metrix were selected from Jagaiah's (2016) systematic review on syntactic complexity and linguistic theory. To obtain the 28 SCM values, the writing samples were scanned using the automated text analysis tool, Coh-Metrix version 3.0. The SCMs from the systematic review were matched against the Coh-Metrix indices related to syntactic complexity to obtain the 28 selected SCMs. Five SCMs from the Coh-Metrix SCMs perfectly matched the SCMs in the literature, and five more were partially matched. The remaining 18 Coh-Metrix SCMs were selected because they were related to sentence elements

that would indicate syntactic complexity in linguistic theory. Four latent variables were hypothesized to incorporate the individual characteristics of the 28 selected SCMs. For example, the Sentence Length latent variable included mean number of words (DESSL), standard deviation of mean number of words (DESSLd), and mean number of words before the main verb (SYNLE).

The following section discusses the findings from the two research questions.

Final Hypothesized Model

The final hypothesized model was tested using the four latent variables and 16 SCMs. The 16 SCMs were reduced from the original 28 SCMs by either removing or combining SCMs to achieve a good fit in the CFA model. Accordingly, this indicated that the four latent variables, Sentence Pattern, Sentence Length, Sentence Connectors, and Sentence Sophistication, incorporating the 16 SCMs, could be used to examine whether a sentence was syntactically complex. Previous studies showed patterns in the use of various SCMs which could be associated with these four latent variables. For example, Hunt (1970) examined high- and low-achieving students, and he found that Sentence Length was an important measure to distinguish between the two groups of students.

The various SCMs that indicated Sentence Pattern in the current study have also been established in previous studies as an important measure of syntactic complexity. Previous studies have used word classes, phrases, gerunds, and infinitives to examine syntactic complexity (see Blair & Crump, 1984; Kagan, 1980; Moran, 1981; Morris & Crump, 1982; Ravid & Berman, 2010). They found that to construct syntactically complex sentences, writers needed to include varied word classes (nouns, verbs, adjectives, adverbs), phrases, passive voice, gerunds, and infinitives. Identifying the incidence score of these SCMs in Sentence

Pattern was important because previous studies found that both at-risk and not-at-risk students used these sentence elements. However, they differed in frequency of use.

While previous studies examined some of the individual SCMs associated with the latent variables in the current study, they did not evaluate them as groups of SCMs measuring similar characteristics. Therefore, it was difficult to directly compare the findings from this study to previous studies because this is the first study to group SCMs into meaningful latent variables and use CFA to confirm the model fit.

The good fit achieved by the CFA for the eighth-grade argumentative essays implied that the four latent variables were relevant and were commonly used measures by this age group in this genre. Therefore, incorporating syntactically more complex sentences in this group can be achieved by mastery of these four latent variables.

The raw data (writing samples) can be used to illustrate the use of the four latent variables confirmed by the CFA approach.

Sentence Pattern. Sentence Pattern can be characterized by word classes, phrases, passive voice forms, gerunds, and infinitives. The following are excerpts of an at-risk student with low use (one standard deviation below average) and a not-at-risk student with high use (one standard deviation above average) of Sentence Pattern elements.

An excerpt of a writing sample from an at-risk student.

I would be against this rule because some people are really dedicated to sports and dont wanna get kicked off the sports team. What I would suggest (verb) is to try and get students the help they need to bring there grade up. Theres no need to kick the student off the sports team for a grade below a c to me thats just nonsense (adjective).

An excerpt of a writing sample from a not-at-risk student.

Another reason why i think parents should not buy the tracking device (adjective phrase) is that it shows you do no want them to grow up. If your child is less than ten then maybe its a good idea. That helps the parents (noun) know that they are safe and where they are supoose to be. when your child hits the age of about twelve then it is time to let your child to have alittle bit of freedom.

The writing sample of the at-risk student has very limited use of Sentence Pattern in terms of passive voice forms, gerunds, and infinitives. However, the writing sample does contain word classes (nouns, verbs, adjectives, adverbs), phrases (adjective and preposition) which may have contributed to the Sentence Pattern score. In contrast, the not-at-risk student has high use of various elements of Sentence Pattern except for passive voice and gerunds that could have further increased the Sentence Pattern score.

Sentence Length. Sentence Length is the number of words in a sentence, clause, or phrase and its variation in an essay. The following are excerpts of an at-risk student with low use (one standard deviation below average) and a not-at-risk student with high use (one standard deviation above average) of Sentence Length elements.

***An excerpt of a writing sample from an at-risk student*.**

I dont think its fair at all if parents buy this device. There are many reasons why it is not fair. I will explain three reasons why I dont think parents shouild buy this device. I dont think parents should buy this device because it is an invasion of privacy, kids would get mad at there parents, and parents would use it to much.

***An excerpt of a writing sample from a not-at-risk student*.**

I do argee because if say a child is missing then they could look at the signal and then you would be able to find the missing child. It would be the best way for parents to look for their childs if they went somewhere without their parents. They will always have that tracker on them incase something happened. would you ever want your child to be missing and have no idea where they are. It is the ultimate way for people who are crazy and take kids for them to know where their own son/doughter is. If they went to the mall and didnt come back in time the parent can tell where they are and if they are right out the door or on there way home. This is the best way for parents to insure their childs safty.

This writing sample of the at-risk student with low use of Sentence Length is characterized by short sentences. In comparison, the not-at-risk student with high use of Sentence Length incorporates varied sentence length, which may have contributed to the Sentence Length score.

Sentence Connector. Sentence Connector can be characterized by connectives (e.g., and, but, because, however, etc.). The following are excerpts of an at-risk student with low use (one standard deviation below average) and a not-at-risk student with high use (one standard deviation above average) of Sentence Connector elements.

***An excerpt of a writing sample from an at-risk student*.**

I do not agree with this idea. I believe it's wrong and a violation of ones privacy. To have a parent be able to track their child is not right. We children have our rights to go places we want without having our parents tracking our every move. How would they like it if we tracked where they went every day?

***An excerpt of a writing sample from an at-risk student*.**

I think that parents having a tracking device on their kids is a good idea because what if the child says that they are going to the mall with their friends but are actually going somewhere else with someone other then their friends. Another reason this would be a good idea is because what if the kid is grounded and decides to sneak out to a party, then the parents will be able to track them and go get them. This would also be a good idea becasue what if your child was walking home from school and got kidnapped, then the parents could track their child and report where they are to the police.

The excerpt selected from the writing sample of the at-risk student does not contain any connectives. Without the use of connectives, the sentences in the excerpt are short and choppy. In contrast, the not-at-risk student with high use of Sentence Connector increased the sentence length and connected ideas more cohesively. However, the same connectors were used repeatedly.

Sentence Sophistication. Sentence Sophistication can be characterized by use of parts-of-speech, varied words, and varied sentence structures. The following are excerpts of an at-

risk student with low use (one standard deviation below average) and a not-at-risk student with high use (one standard deviation above average) of Sentence Sophistication elements.

***An excerpt of a writing sample from an at-risk student*.**

Finally, I feel that parents who get a global positioning device are way to overprotective. Parents who get a tracker device will make their kids rebel against them and make them do things they wouldn't usally do. For example, if you tell them not to go to a party way past their curphew, they will go because they know you are watching them. They will no longer listen to you because they will feel that you don't trust them and they will think that you are being overprotected towards them. Thats why I think that parents should not be able to get a global positionig device.

***An excerpt of a writing sample from a not-at-risk student*.**

Third, by tracking your child, you will know if they are getting to a certain place on time, and they aren't late. For example, if you asked you child to go pick something up at school at 4:00 and they don't get thier until 4:30, then you could miss out on an important event. Also, if your child is staying after school, and they come home later than you expected, then you will know to be home for them, rather than not knowing when they would be coming home and leaving them alone for a certain amount of time. To add, this can also help with teaching responsibility to your child, by telling them what the outcome will be if they are not on time. To end, it is important to keep track of where your child is so you know when and where to be.

Both excerpts demonstrate the use of parts-of-speech, varied words, and varied sentence structures. However, the not-at-risk student used more Sentence Sophistication elements that the at-risk student.

While the use of the four latent variables is evident in the raw data, the relationship with writing score cannot be matched against the MLR model directly because the values of the individual latent variables cannot be controlled. For example, the MLR model showed that increased use of the Sentence Pattern latent variable would have a positive relationship with writing quality. However, a specific writing sample with a high Sentence Pattern value could have a lower-writing score than a different writing sample with a low Sentence Pattern value if the values of the other latent variables are different between the two writing samples.

Conclusion. Using CFA, the four latent variables comprising 16 SCMs were a good representation of syntactic complexity for the sample eighth-grade, automatically scored formative assessment data for argumentative writing analyzed in this study. This supported the hypothesis of the first research question.

Relationship Between the Four Latent Variables and Student Type with Writing Quality

A multiple linear regression (MLR) was used to analyze the relationship between writing quality as reflected in the writing score (dependent variable) and the four latent variables (independent variable). Student Type was used as an indicator variable.

Final MLR model. In the final MLR model, after removing several non-significant two-way interaction variables, all the regression coefficients for the four latent variables were positive and significant at $\alpha = 0.1$. Three of the four latent variables had $p < .05$ with only Sentence Connector having $p = .074$. Sentence Pattern (word classes, passive sentences, gerunds, infinitives, and phrases) had the highest regression coefficient of 0.73 ($p = .029$) when compared to the other latent variables, which implied that it had the greatest impact on writing quality. It must be noted that the regression coefficients for the Sentence Length, Sentence Connector, and Sentence Sophistication were 0.43, 0.44, and 0.35 respectively, which indicated that all four latent variables had a positive impact on writing quality. Recognizing this positive impact, a follow-up intervention study should be done to determine if grade eight students should construct complex sentences based on the four latent variables to improve their writing scores. These latent variables are abstract concepts, and they need descriptors to enable educators to incorporate them in writing instruction. Specifically, Sentence Pattern refers to word classes, phrases, passive voice forms, gerunds, and infinitives. Sentence Length, on the other hand, refers to the number of words in a sentence, clause, or phrase and its variation in an

essay. Sentence Connector corresponds to varied connectives such as and, but, because, however, etc., while Sentence Sophistication includes parts-of-speech, varied words, and varied sentence structures that differ with adjacent sentences. All of these sentence elements can be emphasized in writing instruction to increase sentence complexity and to improve writing scores.

This is the first study to demonstrate a positive relationship between the four latent variables and writing quality. In contrast, findings from previous studies showed weak or inconsistent relationships between the SCMs measured and writing quality. Belanger and Martin (1984) examined the relationship between syntactic complexity and writing quality using syntactic density score, which incorporates characteristics of Sentence Pattern. However, the syntactic density score measure included 10 different SCMs, and the findings indicated that there was no relationship between syntactic density score and writing quality. These differing conclusions could be attributed to the fact that the SCMs in the current study for Sentence Pattern were not directly comparable with syntactic density score. Findings from this study are more conclusive because it analyzed a significantly larger sample size and used CFA to determine the best fitting SCMs for the four hypothesized latent variables.

The not-at-risk indicator was highly significant at $p < .001$ which implied that the students who are not-risk had a higher average writing score than the at-risk students. If the Student Type indicator was removed from the MLR model, the R^2 reduced from 0.31 to 0.14, which indicated that Student Type was a highly significant variable and contributed to 17% of the variance in student writing scores. If all the latent variables were kept constant at their mean value for both at-risk and not-at-risk students, on the average, at-risk students would have a writing score of 17.03 while not-at-risk students would have a writing score of 22.34. This

means that, independent of the use of syntactically complex sentences in writing, not-at-risk students on the average have 31% higher writing scores than at-risk students.

The contribution of syntactic complexity to writing scores can be seen if the four latent variables were increased beyond their mean values. For at-risk students, if each of the four latent variables was increased to one standard deviation above the mean, the writing scores would correspondingly increase by 8% above the mean score of 17.03 to 18.31. On the other hand, for not-at-risk students, if each of the four latent variables was increased to one standard deviation above the mean, the writing scores would correspondingly increase by 4% above the mean score of 22.34 to 23.34. If each of the four latent variables was decreased to one standard deviation below the mean, the writing scores would correspondingly decrease by 12% below the mean score of 17.03 to 15.07 for not-at-risk students and decreased by 8% below the mean score of 22.34 to 20.66 for at-risk students.

To determine the impact of writing score by changes in each individual latent variable, separate estimates for at-risk and not-at-risk students were determined for latent variables which interacted with Student Type. For latent variables which did not interact with Student Type, the impact of writing score by changes in the latent variables was the same for at-risk and not-at-risk students. Specifically, for at-risk students, a one standard deviation increase in the latent variables increased the writing score by 0.33, 0.48, 0.53, and 0.28 for Sentence Length, Sentence Connector, Sentence Pattern, and Sentence Sophistication, respectively. To compare, for not-at-risk students, a one standard deviation increase in the latent variables increased the writing score by 0.33, 0.70, 1.42 and 0.28 for Sentence Length, Sentence Connector, Sentence Pattern, and Sentence Sophistication, respectively. Unlike at-risk students who showed positive

results in writing scores for increases in each of the latent variables, not-at-risk students demonstrated lower writing scores with increased use of Sentence Connectors.

The analysis of the relationship between syntactic complexity and writing quality by Student Type implies that explicit instruction on the use of word classes, adverbial and preposition phrases, gerunds, infinitives, passive voice forms, longer phrases and clauses, word choice and varied sentence structures may benefit all students. For at-risk students, increased use of connectives may improve their writing score. In contrast, not emphasizing these sentence elements would have a negative impact on writing scores for all students. Both these findings suggest the importance of incorporating syntactic complexity as an integral component of writing instruction, particularly for at-risk students.

Conclusion. Based on the findings of the MLR model, it can be concluded that syntactic complexity as manifested in the four latent variables showed a modest positive relationship to writing quality. Furthermore, this relationship varied significantly by Student Type. This supported the hypothesis of the second research question.

Implications of Study Findings

The findings of this study have contributed to the field of education in the areas of methodology, writing assessment, and writing instruction.

Methodology. The research methods used in this study were new from several perspectives. No study has compiled the great number of SCMs that have been used to examine syntactic complexity and combined them into meaningful groupings. The current study compiled the 28 SCMs and grouped them separately into four latent variables. Based on this compilation, the hypothesized model was developed and tested using CFA. Only one previous study, Kagan (1980), attempted to use 17 SCMs that explained syntactic complexity based on

a principal component factor analysis. However, Kagan (1980) did not hypothesize a model *a priori* and did not test that the 17 SCMs and the resultant six latent variables were a good fit. In this study, a new method of explaining syntactic complexity was developed. This was done using a CFA approach to combine 16 SCMs (reduced from an original set of 28 SCMs) into four latent variables. The four latent variables could be used instead of individual SCMs to assess writing quality in argumentative essays. Descriptions about the latent variables could be used in checklists and rubrics to evaluate good sentence-construction skills.

Examining the four latent variables comprising 16 SCMs has allowed important questions to be answered on how these four latent variables perform as objective indices of eighth-grade, automatically scored argumentative essays, their relationships to writing quality, and their interactions with student type. A major advantage of this research over previous studies was the investigation of a large number of SCMs simultaneously using one large data set. The use of automated essay scoring ensured that the calculated SCM values for each essay was accurate with no interrater reliability issues. The ability to analyze several SCMs accurately avoided the inconsistency and variability found among previous studies in terms of choice and definition of measures, writing task used, sample size, and student type.

Writing assessment. The findings also have important implications for writing assessment. In particular, the results suggest that SCMs comprising the four latent variables such as phrases (preposition, adverb) word classes (nouns, verbs, adjectives, adverbs), connectors, word choice, and varied sentence structures may elevate writing scores. Because a modest positive relationship between the four latent variables and writing quality was observed,

this finding has a basis in future studies to include descriptors related to these four latent variables as objective measures that can be used to assess sentences.

When assessing sentences, educators should identify if the sentences comprise various elements of the four latent variables: Sentence Pattern, Sentence Length, Sentence Connector, Sentence Sophistication. To ensure sentence elements from Sentence Pattern are used, students should construct sentences using varied nouns, verbs, adjectives, and adverbs that would provide basic information in the sentence. A sentence with only these word classes may result in simple sentence structures and these may be commonly used by students who are-at-risk. However, the use of adverbial and prepositional phrases in a single sentence with the combination of the word classes increases the complexity of the sentence. Other elements of Sentence Pattern that increase the complexity of sentences are passive voice forms, gerunds (e.g., dreaming, swimming, etc.), and infinitives (e.g., to play, to see, etc.). The following example shows how using a gerund in Sentence 3 makes it syntactically more complex compared to Sentence 1 and 2.

I swim on a hot day. (S1)

It is pleasant. (S2)

Swimming on a hot day is pleasant. (S3)

It is important for educators to recognize that some of the sentence elements such as adverbial and prepositional phrases, and passive voice forms may be difficult for both at-risk and not-at-risk students. Therefore, descriptors in the rubric should be aligned with grade-level expectations.

Varying sentence length is essential when composing a text. Findings of this study show that use of varied sentence length increased writing score. Varied clause and phrase

length are components of the Sentence Length latent variable. Varied sentence length avoids monotony, creates emphasis where needed, and helps the reader understand connections between different points. Descriptors in the rubric should be specific on what defines sentence length. Use of simple, compound, complex, and compound complex sentences may be appropriate descriptors to define sentence length. Using these sentence structures may increase writing score. A long sentence with a list of connectives that forms a paragraph may not be considered an appropriate sentence length, and this may reduce writing score.

The Sentence Connector latent variable reflects the use of connectives such as *and, but, because, however*, etc., to combine short and choppy sentences into longer, syntactically complex sentences. Descriptors such as causal, logical, contrastive, temporal, and additive connectives should be listed in the rubric. Use of these connectors may enable students to get a higher writing score. The following example shows how combining S1, S2, S3 could increase sentence complexity in S4.

I do not agree with this idea. (S1)

It is wrong. (S2)

It is a violation of one’s privacy. (S3)

I do not agree with this idea because it is wrong, and it is a violation of one’ privacy. (S4)

Sentence Sophistication, which refers to the use of varied parts-of-speech (nouns, verbs prepositions, adverbs, gerunds, etc.), word choice and varied sentence structures, increases the sentence complexity. Adjacent sentences should use different words and sentence structures to increase syntactic complexity and to receive a higher writing score. For example, the following excerpt shows different sentence elements used in adjacent sentences. S1 begins with a prepositional phrase in a simple sentence, and S2 begins with a noun phrase in a simple

sentence, while S3 begins with a conditional clause in a complex sentence. The three sentences have a variety of words that are not repeated in adjacent sentences. When words are not repeated in adjacent sentences, they increase the level of syntactic complexity.

In my opinion for success, you need intelligence and good looks. (S1)

This combination will help you become successful. (S2)

If you are trying to make it good in life, you need a foundation of both knowledge and looks. (S3)

Educators when developing rubrics to assess sentence-level skills should be mindful of grade-level expectations. Grade two students may not have learned how to construct passive voice forms, so it would not be realistic to assess this sentence element.

The current study is not an intervention study, but there is evidence in the literature from intervention studies to suggest that there is a relationship between sentences that are syntactically complex and writing quality. Saddler, Asoro, and Behforooz (2008), Saddler,Behforooz, and Asoro (2008), and Saddler and Graham (2005) used sentence combining skills as an intervention strategy to increase sentence-level complexity. All three studies found that when students constructed sentences that are syntactically complex, their writing scores were higher.

Writing instruction. The findings also have potential implications for writing instruction. Common Core State Standards or other state standards require students to master various sentence types (simple, compound, complex, compound-complex), and these types of sentence structures are related to Sentence Pattern, Sentence Length, Sentence Connector, and Sentence Sophistication latent variables. The MLR model predicting writing quality using the four latent variables and Student Type produced an R^2 of 31%. Recognizing that writing quality

is impacted by other factors besides syntactic complexity, this R^2 value is sufficiently high to indicate that construction of complex sentences could have an impact on writing scores. Future intervention studies should be conducted to test whether explicit instruction on increased use of word classes (e.g., nouns, verbs, adjectives, and adverbs), passive voice forms, length of phrases and clauses, connectives, word choices, and varied sentence structures for eighth-grade students would positively impact writing quality for at-risk and not-at-risk students.

Future intervention studies could include lessons on various sentence elements that describe the four latent variables by sequencing them based on level of complexity. For example, students can be taught sentence elements in the Sentence Pattern latent variable using the following order: word classes, phrases, gerunds, infinitives, and passive voice forms. It is important that students are provided with numerous examples on how to use various sentence elements to increase syntactic complexity. The following passive voice sentences could be written in varying levels of complexity:

The door was opened. (S1)

The door was opened by the little boy. (S2)

The door was opened by the little boy who was crying. (S3)

It should be made known to students that when using word classes or passive forms, sentences may become increasingly long. Long sentences with redundant words do not convey information succinctly and will not receive higher writing scores. Therefore, students should be taught to balance between length and appropriate number of other sentence elements to convey meaning effectively.

Writing long sentences (e.g., simple, compound, complex, and compound-complex sentences) includes the use of word classes, phrases, gerunds, infinitives, passive voice forms,

connectives, word choices, and varied sentence structures. Increasing sentence length without considering effective use of other sentence elements may reduce the writing score. Therefore, it is important, especially for students who are at-risk, to be taught how to combine sentences strategically to increase the length of a sentence without distorting the meaning. Hunt (1970) noted that less-skilled writers tend to combine short sentences using connectives such as *and* or *but* frequently. Repeated use of these connectives in a sentence is not an effective way to present ideas because ideas between sentences are not appropriately connected, and this may confuse the reader. Educators should expose students to various use of connectives to encourage them to construct effective sentences that are syntactically complex. Students should be taught to use coordinating (e.g., for, nor, yet) and subordinating (e.g., while, moreover, before) conjunctions to construct grammatically correct complex sentences. This may prevent students from writing run-on sentences. In the following run-on sentence, three sentences were combined with several connectors *and* with one sentence having a missing connector. It is a compound-complex sentence with a potential high score for syntactic complexity. However, lack of appropriate connectives may reduce the writing score.

Buying this product is a really bad idea because it will invade your child's privacy, it will destroy your child's self-esteem and make them feel like they're not trusted, and if the item gets into somebody else's hands it could put your child in danger.

Constructing sentences that are sophisticated requires instruction on sentence elements that describe Sentence Sophistication. Students should be taught to vary the use of words, and varied sentence structures (simple, compound, complex, and compound-complex) between adjacent sentences. To increase the sophistication level of a sentence, students should be taught to construct sentences with varying levels of hierarchy by embedding a subordinate clause to an

independent clause. The following excerpt is an example of repeated words and sentence structures between adjacent sentences. The writer has repeatedly used the prepositional phrase *To succeed in your life* to begin the first two sentences. The word *important* was used four times in a short paragraph. Repeated words and sentence structures will reduce the level of syntactic complexity which in turn could reduce the writing score.

To succeed in your life, there are many attributes you have to have. To succeed in your life, you need money, intelligence, and good looks. Most think that Money is the objective that is most important, but it really is intelligence. Money can bring you everything in the world that you want, but intelligence is more important. Good looks surely may be important to your average person, but intelligence is the most important.

Limitations

While there are many strengths in the current study, there are however, some limitations to acknowledge. The main limitation of this study is that it only examined exclusively the relationship between syntactic complexity and writing quality, and it did not take into account other factors such as overall content, organizational structure, vocabulary, mechanics, and length which could also impact writing quality. Sentence construction skills were viewed as one aspect of writing quality. Considering the multi-componential nature of writing, other components may also contribute to writing quality. It is therefore important to conduct another study to understand the role of syntactic complexity contextualized within other writing components.

Only one genre (argumentative) and only one grade level (Grade 8) was examined, and these findings do not necessarily translate to other genres and grade levels. Results may differ with other genres such as informative, descriptive, or narrative, and different grade levels.

Another limitation is that the study relied only on SCMs that were available on Coh-Metrix. Although these measures have been validated by numerous studies as an extremely

powerful text analysis tool, only five Coh-Metrix SCMs had a perfect correspondence with the SCMs used in previous studies. Other automated text analysis tools are Biber Tagger and Syntactic Complexity Analyzer. Biber Tagger analyzes 67 linguistic features, while Syntactic Complexity Analyzer counts instances of eight structures (e.g., clauses, dependent clauses, verb phrases, etc.) to produce the 14 indices of syntactic complexity. None of these tools meet all of the ideal criteria; however, Syntactic Complexity Analyzer has a higher parser rate accuracy than does Coh-Metrix. Results may differ when different text analysis tools are used and different SCMs are selected.

Automated text analysis tools (PEG and Coh-Metrix) were used to obtain the outcome and predictor variables. These computer-based tools use powerful algorithms to convert text into numbers. Because the algorithms used are not publicly available, it is not known if both these tools use similar algorithm to arrive at the writing scores and scores on SCMs. If the algorithms were similar, a natural bias could be created where high SCM scores in Coh-Metrix are associated with high writing scores in PEG. However, there are benefits to using these tools because they are both highly reliable and there is no measurement error. Additionally, PEG scoring is modeled on human trait scoring while Coh-Metrix is an authoritative text analysis tool and reports the highest average accuracy for expository texts (Hempelmann, Rus, Graesser, & McNamara, 2006), suggesting it unlikely they are modeled with similar algorithms.

The four latent variables and associated SCMs were examined using a single state's writing assessment data; the results may be different using data from other states. Some states may continue to use state standards, and complex sentence construction skills may only be introduced at higher grades (e.g., Grades 9 -12). Consequently, students may not be familiar with the varied sentence structures that they could use to translate their ideas into writing. In

addition, the fact that the current study analyzed eighth-grade writing samples alone may affect the interpretation and generalizability of the results.

In any kind of writing study, depending on how writing quality is measured, the relationship between writing quality and the latent variables examined may cause the results to vary from study to study. This relationship is dependent upon the specific measures being used and the genres being examined (Beers & Nagy, 2009). Also, the SCMs selected for this study were analyzed using the Coh-Metrix Automated Essay Scoring tool and its underlying algorithm to calculate the SCM values for each essay. Studies using different SCMs and tools to calculate the SCM values may come up with results that are not comparable with the findings of this dissertation study.

No previous studies have pooled various SCMs into different latent variables to examine syntactic complexity with the exception of Kagan (1980). It was difficult to corroborate findings from this study with Kagan's (1980) because the SCMs in her study were used to examine the relationship between syntactic complexity and analytic cognitive style. Also, Kagan (1980) did not confirm her specified model that she obtained from principal components factor analysis.

Areas of Future Research

To build on this study, future research should undertake principled replications of the analyses conducted using other genres and grade levels to expand on these findings. This might facilitate the understanding of syntactic complexity use and its relationship to writing quality in different genres and grade levels. For example, certain complex sentence structures (e.g., mean number of words before the main verb, mean number of modifiers per noun phrase, mean number of sentence syntax similarity between adjacent sentences) may not be reflected in the

data of earlier grades because they may not be developmentally appropriate, or they may not have formally learned the complex sentence structures. Also, the choice of SCMs may vary by different genres. It is likely that the modeling approach used in this study could confirm findings from previous studies that the argumentative genre uses more syntactically complex sentences as compared to other genres and that the impact of SCMs on writing quality is greater for the argumentative genre.

In addition, it would be interesting for future studies to examine the relationship between syntactic complexity and writing quality with other components of writing quality such as organization, content, and vocabulary. This will provide a more holistic and complete analysis of predictors of writing quality.

Only data from one state was used in the current study; therefore, it would be interesting to see if the results differ with the eighth-grade data from other states. These results might differ because in some states such as New Mexico, the complex sentence structures such as gerund phrases, infinitives as nouns are only introduced at the high school level, so students may or may not be able to construct sentences using syntactically complex structures.

Summary

In the current study, 28 SCMs from Coh-Metrix were selected using two criteria: Jagaiah's (2016) systematic review on syntactic complexity and linguistic theory. A hypothesized model of four latent variables and 28 SCMs was developed and tested using CFA. The model was refined into 16 SCMs with the same four latent variables in order to get a good fit. These four latent variables and a student-type indicator variable were used as predictor variables in an MLR model to examine the relationship with writing quality. The findings indicated that a well-constructed set of SCMs that were logically classified into four latent

variables was a good barometer for explaining writing quality for the eighth-grade dataset that was analyzed. This study has two major contributions to the writing literature. First, it is the only study of its kind to simultaneously analyze several SCMs and group them into latent variables using CFA to test the hypothesized model fit. This was accomplished by using a large dataset of more than 1,000 essays and using an automated text analysis tool to calculate the SCMs that were being analyzed. Previous studies analyzed only a few SCMs at a time using a manual approach with a small dataset. The use of CFA to test the fit of the 28 SCMs and the four latent variables is also a new approach in the literature on syntactic complexity. Second, the researcher developed an explicit model using MLR to study the relationship between the latent variables and student type with writing quality. For the first time, syntactic complexity as manifested in the four latent variables clearly showed a modest positive relationship to writing quality for each latent variable, and the relationship varied by Student Type. The findings have implications for methodology, writing assessment, and writing instruction on sentence construction skills.

APPENDIX A

ARGUMENTATIVE PROMPTS PROVIDED IN BENCHMARK WRITING ASSESSMENT

Argumentative Prompt

1. A major research study has been done that indicates that a majority of accidents occur when drivers are under the age of 18. The Governor is considering increasing the legal driving age so that no one under the age of 18 will be able to get a permit or a license. Do you think this is a good or bad idea? Write a letter to Governor Rell convincing her of your point of view. When you write your letter, be sure to:

   •State your position.
   •Provide support and details that your reader will find persuasive; and
   •Organize your ideas and present your argument clearly.

2 A very large store that sells a variety of merchandise is planning to open in your community. This will mean more choices and lower prices. The opening of this store may also result, however, in several small, family-owned stores going out of business. Are your for or against the building of the new store? Be sure to develop your response fully.

3. Do you think that athletes and entertainers are often paid huge sums of money for the work they do? How does society justify the difference between their salaries and those of people who make much less money doing other jobs? Be sure to develop your response fully.

4. Imagine you have a choice between being schooled at home full time or attending school with others. Think of the positive and negative aspects of each of these types of schooling. Choose whether home schooling or attending school with others is better. Be sure to develop your response fully.

5. In this country, many people are thinking about ways to change schools. Some people think that the school day should be longer. Take a position for or against changing the length of the school day, and support your reasons. Be sure to develop your response fully.

6. More and more people use computers, but not everyone agrees that this benefits society. Those who support advances in technology believe that computers have a positive effect on people. They teach hand-eye coordination, give people the ability to learn about faraway places and people, and even allow people to talk online with other people. Others have different ideas. Some experts are concerned that people are spending too much time on their computers and less time exercising, enjoying nature, and interacting with family and friends.

   When you write your paper, be sure to:
   1. State your opinion about the effects of computers.
   2. Give detailed reasons that will persuade the readers of the local newspaper to

agree with your position.
3. Organize your ideas well and present them clearly.

7. On the whole, would you say that indoor activities or outdoor activities are more enjoyable? Explain your choice, and be sure to develop your response fully.

8. Out of all the holidays that occur during this time of the year, which one is your favorite?

9. Parents can now buy a global positioning device that can let them know exactly where their child is at any moment. Decide whether or not you agree that it is acceptable to track their child's whereabouts. Give reasons in support of your stance, and be sure to develop your response fully.

10. Persuade your audience to watch the film, Forrest Gump. You must use persuasive language/transitions and at least one of the techniques learned in class.

11. Some people say that adults forget what it's like to be young after they reach a certain age. Do you agree or disagree with this idea? Write an essay stating your position, and give persuasive examples that support your view. Be sure to develop your response fully.

12. Suppose that your school is considering revising the academic requirements for its student athletes. The new policy will require students to maintain a minimum grade of “C” or “Average” in all subjects in order to participate in a sport. Would you be for or against this new policy? Provide reasons, and be sure to develop your response fully.

13. The Board of Education is considering a change to the school calendar. It has to decide if Columbus Day is a day school should be in session or a holiday. Some people say Columbus was a bold navigator who advanced civilizations. Others say he was a reckless adventurer seeking personal gain while causing trouble for Native Americans and advancing slavery. What is your opinion about celebrating Columbus Day?

14. The Internet offers us many great opportunities. There are, however, also disadvantages to consider. Do you think the internet is a positive or negative influence on our lives? Be sure to develop your response fully.

15. Which would best help you succeed in life as an adult: money, intelligence, or good looks? Be sure to develop your response fully.

16. Write a developed and logically argued essay on the topic of your choice.

APPENDIX B

CORRELATION BETWEEN SYNTACTIC COMPLEXITY MEASURES

SCMs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	4	25	26	27	28
1. DESSL																												
2. DESSLd	0.78																											
3. CNCAll	0.32	0.25																										
4. CNCCaus	0.14	0.11	0.39																									
5. CNCLogic	0.30	0.20	0.71	0.33																								
6. CNCADC	0.11	0.11	0.30	-0.09	0.33																							
7. CNCTemp	0.02	0.01	0.34	0.03	0.13	0.02																						
8. CNCTempx	-0.10	-0.05	0.13	0.32	0.06	0.02	0.02																					
9. CNCAdd	0.19	0.18	0.67	-0.04	0.30	0.48	0.02	0.06																				
10. SYNLE	0.53	0.32	0.23	0.00	0.21	0.07	0.09	-0.09	0.09																			
11. SYNNP	0.05	-0.01	-0.12	-0.18	-0.12	0.08	-0.01	0.03	-0.03	0.24																		
12. DRNP	-0.25	-0.14	-0.13	0.09	-0.26	-0.01	-0.03	0.26	0.06	-0.24	-0.25																	
13. DRVP	0.18	0.10	0.03	0.01	0.18	-0.16	-0.14	-0.30	-0.11	0.04	-0.34	-0.53																
14. DRAP	0.02	0.01	0.11	-0.09	0.13	0.05	0.20	0.02	0.08	0.00	-0.14	-0.22	-0.02															
15. DRPP	-0.16	-0.12	-0.15	-0.03	-0.28	0.10	-0.05	0.24	0.07	-0.10	0.08	0.47	-0.57	-0.10														
16. DRPVAL	-0.07	-0.10	-0.09	-0.12	-0.04	-0.04	-0.04	-0.08	-0.11	0.02	0.21	-0.19	0.12	-0.04	-0.05													
17. DRNEG	-0.14	-0.10	-0.12	0.09	-0.04	-0.02	0.05	-0.03	-0.17	-0.09	-0.11	-0.11	0.04	-0.01	-0.22	-0.01												
18. DRGERUND	0.03	0.01	-0.08	-0.10	-0.07	0.10	-0.01	-0.10	0.02	0.01	0.01	-0.13	0.08	0.10	0.14	-0.04	-0.02											
19. DRINF	0.12	0.10	-0.05	0.12	0.01	-0.10	-0.06	0.09	-0.13	0.03	-0.10	-0.24	0.53	-0.09	-0.18	-0.02	-0.03	-0.09										
20. WRDNOUN	-0.24	-0.18	-0.22	-0.11	-0.35	0.07	-0.09	0.23	0.06	-0.10	0.40	0.53	-0.57	-0.23	0.53	0.03	-0.21	-0.02	-0.21									
21. WRDVERB	-0.02	-0.02	0.03	-0.07	0.03	-0.10	0.05	-0.09	-0.02	0.02	-0.05	-0.19	0.33	0.01	-0.30	0.13	0.01	0.21	-0.02	-0.23								
22. WRDADJ	-0.14	-0.13	-0.01	0.03	-0.18	0.07	0.04	0.08	0.07	-0.02	0.24	0.13	-0.30	-0.16	0.14	-0.08	-0.07	-0.14	-0.07	0.23	-0.24							
23. WRDADV	-0.06	-0.04	0.08	0.00	0.09	0.12	0.23	0.02	0.02	-0.04	-0.19	-0.19	-0.09	0.73	-0.13	-0.04	0.46	0.07	-0.14	-0.26	-0.06	-0.06						
24. SYNSTRUTa	-0.02	-0.04	-0.06	-0.04	-0.06	-0.01	-0.02	-0.09	-0.02	-0.01	0.06	-0.02	-0.02	-0.03	0.01	-0.02	0.01	-0.01	0.02	0.05	-0.03	0.07	-0.02					
25. SYNSTRUTt	-0.01	0.02	0.05	-0.01	0.03	0.07	-0.05	0.05	0.10	0.03	0.00	0.04	-0.03	0.01	-0.01	-0.04	-0.04	-0.01	-0.03	0.04	-0.01	0.00	0.01	-0.25				
26. SYNMEDlem	0.01	-0.02	-0.02	0.01	-0.01	0.01	0.00	-0.04	-0.03	0.01	0.04	-0.02	0.01	-0.01	0.02	0.06	-0.06	-0.02	0.03	0.00	0.00	0.04	-0.07	0.21	-0.23			
27. SYNMEDwrd	-0.03	0.00	-0.01	0.00	0.02	-0.02	0.00	0.03	0.01	-0.04	-0.05	0.06	-0.04	0.06	0.03	0.03	0.02	-0.02	-0.03	0.04	-0.04	0.04	0.03	-0.25	-0.04	-0.07		
28. SYNMEDpos	-0.03	-0.02	0.02	-0.05	0.04	0.08	0.03	0.03	0.03	-0.02	0.00	0.00	-0.02	0.06	-0.03	0.02	0.03	-0.02	-0.03	0.00	-0.03	0.03	0.07	-0.03	0.13	0.11	0.08	

Note. N = 1,029. All correlations are statistically significant at $\rho < .001$.

APPENDIX C

PERFECT MATCH OF LITERATURE REVIEW MEASURES AND COH-METRIX MEASURES

Coh-Metrix Measures	Definition	Syntactic Complexity Measures	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measures
Mean number of words (length) of sentences (DESSL)	Average number of words in a sentence	Mean number of words per sentence	Length of a sentence: a group of words punctuated at the end of a sentence (Hunt, 1965)	Both SCMs are measured the same way.	Direct and different interpretation: Number of words are counted based on the first word that begins the sentence until the last word which has the end punctuation. So, in Coh-Metrix, a whole paragraph with several sentences with no end punctuation at each grammatical sentence is considered as a single sentence.
Incidence score of adverbial phrases (DRAP)	Incidence score of adverbial phrases. Examples: in silence, like a hawk	Number of adverbs of time	Frequency of an action.	The density of particular word types (adverbial phrases) indicates the text is informationally	Direct: Number of adverbial phrases divided by the number of words

Coh-Metrix Measures	Definition	Syntactic Complexity Measures	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measures
				dense (McNamara, Graesser, McCarthy, Cai, 2014).	multiplied by a 1,000.
Mean number of words before main verb: (SYNLE)	Length of a clause or a phrase	Mean number of words before the main verb	Left embeddedness of the main clause in sentences.	Both measures are Coh-Metrix indices.	Direct: The main verb (e.g., I think) think is considered as the main verb and there is just one word before, so SYNLE = 1.
Incidence score of preposition phrases (DRPP)	Incidence score of prepositional phrases.	[1]Number of prepositional phrases	Number of incidence score of prepositional phrases.	Both measures examine the incidence score of prepositional phrases.	Direct: Number of prepositional phrases divided by the number of words multiplied by a 1,000.
Incidence score of adverbs (WRDADV)	Incidence score of adverbs Examples: quickly, happily	Number of adverbs of time (when, then, once, while)	A word to describe a verb, adjective or an adverb.	The density of particular word types (adverbs) indicates the text is informationally dense.	Direct: Number of adverbs divided by the number of words multiplied by a 1,000.

APPENDIX D

PARTIAL MATCH OF LITERATURE REVIEW MEASURES AND COH-METRIX MEASURES

Coh-Metrix Indices	Definition	Syntactic Complexity Measures	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
Temporal Connectives Incidence CNCTemp	Incidence score of temporal connectives. Examples: (before," "after," "then")	Number of adverbs of time (when, then, once, while),	Refers to words that modifies a verb, adjective or an adverb in relation to time (when, then, once, while)	Partial match to temporal connectives because both these measures are adverbs of time.	Direct: Number of temporal connectives divided by the number of words multiplied by a 1,000.
Mean number of modifiers per NP: SYNNP	Frequency count of words, phrases, or clauses, which functions as an adjective or an adverb to provide a more specific description or meaning in a noun phrase	Total number of instances of free modifiers (initial + medial + final positions) (phrases and clauses)	Frequency count of words, phrases, or clauses, which functions as an adjective or an adverb in the initial, medial, and final position of a sentence to provide a more specific description or meaning	Measures all types of modifiers that modifies the whole sentence instead of specifically measuring noun phrases. However, in an essay, most modifiers do modify noun phrases, thus making this measure a partial match to the Coh-Metrix index.	Direct: The number of modifiers (words, phrases, or clauses, which functions as an adjective or an adverb to provide a more specific description or meaning) is counted and divided by number of noun phrases. Text with higher number of modifiers have higher scores and text with fewer modifiers have lower scores.

Coh-Metrix Indices	Definition	Syntactic Complexity Measures	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
Incidence score of verb phrases (DRVP)	This is the incidence score of verb phrases.	Number of "be" and "have" auxiliaries	Helps the main verb. For example, "It was written by a girl". The verb "was" provides further information and commonly used in passive sentence structures.	When a verb is used as an auxiliary verb (e.g., have, do, be) it will always team up with another verb to create a complete verb phrase. Therefore, the scores obtained for verb phrases indicate either the number of "be" and "have" auxiliaries.	Direct: Number of verb phrases divided by the number of words multiplied by a 1,000.
Incidence score of infinitives (DRINF)	Incidence score of infinitives Examples: be, have, has, read	Number of "be" and "have" auxiliaries	Helps the main verb. For example, "It was written by a girl". The verb "was" provides further information and commonly used in passive sentence structures.	Infinitives are prevalent with a high density of intentional content, where there are two parts to a sentence (subject and predicate). Subject and predicate length indicate syntactic complexity (McNamara, Graesser,	Direct: Number of infinitives divided by the number of words multiplied by a 1,000. Examples such as "to have", "to get" are counted as infinitives.

Coh-Metrix Indices	Definition	Syntactic Complexity Measures	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
				McCarthy, Cai, 2014).	

APPENDIX E

COH-METRIX MEASURES RELATED TO SYNTACTIC COMPLEXITY BASED ON LINGUISTIC THEORY

Coh-Metrix Indices	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
Incidence score of additive connectives: CNCAdd	Frequency count of additive connectives per 1,000 words.	Frequency count of the co-ordinations, will provide information on the incidence score of additive connectives because ‘and’ and ‘plus’ are additives.	Direct: Number of additives divided by the number of words multiplied by a 1,000.
Standard deviation of the mean length of sentences (DESSLd)	Standard deviation of mean length of sentences in a text.	A large standard deviation indicates that the text has large variation in terms of the lengths of its sentences, such that it may have some very short and some very long sentences. Length of sentence is an attribute of syntactic complexity.	Direct: Number of words in the essay is divided by the number of sentences.
All connectives: CNCAII	Incidence score of all connectives. Five general classes of connectives are examined.	Connectives that function as coordinating or subordinating conjunctions combine sentences, thus increasing the complexity of a sentence structure.	Direct: Number of all additives divided by the number of words multiplied by a 1,000.
Causal connectives: CNCCaus	Incidence score of causal connectives. Examples: ‘because’, ‘so’, ‘therefore’, ‘in order to’.	Connectives that function as coordinating or subordinating conjunctions combine sentences, thus increasing the	Direct: Number of causal connectives divided by the number of words multiplied by a 1,000.

Coh-Metrix Indices	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
		complexity of a sentence structure.	
Incidence score of logic connectives: CNCLogic	Incidence score of logic connectives. Examples: variants of ‘and’, ‘or’, ‘not’, and ‘if-then’	Connectives that function as coordinating or subordinating conjunctions combine sentences, thus increasing the complexity of a sentence structure.	Direct: Number of logic connectives divided by the number of words multiplied by a 1,000.
Incidence score of adversative/contrastive connectives: CNCADC	Incidence score of adversative/contrastive connectives. Examples: ‘although’, whereas’, ‘however’, ‘nevertheless’	Connectives that function as coordinating or subordinating conjunctions combine sentences, and this increases the complexity of a sentence structure.	Direct: Number of adversative/contrastive divided by the number of words multiplied by a 1,000.
Expanded Temporal Connectives Incidence CNCTempx	Incidence score of expanded temporal connectives Examples: first, until	Connectives that function as coordinating or subordinating conjunctions combine sentences, and this increases the complexity of a sentence structure.	Direct: The definition of expanded temporal connectives is not clear. So, it is unclear which words are counted as expanded temporal connectives. Number of expanded temporal connectives divided by the number of words multiplied by a 1,000.
SYNMEDpos Minimum editorial distance score for part of speech tags	Measures the minimum editorial distance score for part of speech tags	Important to know if students are able to use all the parts of speech. This measure in Coh-Metrix refers to content words	Indirect: SYNMEDpos calculates the extent to which one sentence needs to be modified (edited)

Coh-Metrix Indices	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
		(e.g., nouns, verbs, adjectives, adverbs) and function words (e.g., prepositions, determiners, pronouns).	to make it have the same syntactic composition as a second sentence. These scores will indicate if the students have varied their sentence structures. The calculation is not clearly defined for this measure, and it is not straight-forward to interpret.
SYNMEDwrd semantic and syntactic dissimilarity	Minimum editorial distance score for words. SYNMEDwrd considers the words but not the parts of speech Example: the, book	Examines a combination of semantic and syntactic dissimilarity by measuring the uniformity and consistency sentence construction between consecutive sentences in a text. Lack of similarity will indicate higher level of complexity because readers have to process words from different grammatical classes to understand the text (McNamara, Graesser, McCarthy, Cai, 2014).	Indirect: SYNMEDwrd calculates the extent to which one sentence needs to be modified (edited) to make it have the same syntactic composition as a second sentence. These scores will indicate if the students have varied their sentence structures. The calculation is not clearly defined for this measures, and it is not straight-forward to interpret.
SYNMEDlem semantic and syntactic dissimilarity)	Minimum editorial distance score for lemmas. SYNMEDlem considers the words but not the parts of	Examines a combination of semantic and syntactic dissimilarity by measuring the uniformity and consistency sentence construction between	Indirect: SYNMEDlem calculates the extent to which one sentence needs to be modified (edited) to make it have the same

Coh-Metrix Indices	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
	speech. Examples: book, run, the	consecutive sentences in a text. Lack of similarity will indicate higher level of complexity because readers have to process lemmas from different grammatical classes to understand the text (McNamara, Graesser, McCarthy, Cai, 2014).	syntactic composition as a second sentence. These scores will indicate if the students have varied their sentence structures. The calculation is not clearly defined for this measures, and it is not straight-forward to interpret.
Syntactic structure similarity SYNSTRUTt	Proportion of intersection tree nodes between all sentences and across paragraphs. Measures the uniformity and consistency of the syntactic constructions in the text or similarity (Sim) between all combinations of sentence pairs across paragraphs.	The syntactic structure similarity SYNSTRUTt index does account for similarity between all combinations of sentence pairs across paragraph, but this measure does not explicitly compute a subject and a verb pattern. It is possible sentence pattern is taken into account, but there are no measures that are specific to the measures used in previous studies.	Direct: This SCM is measured by removing uncommon subtrees found between two adjacent sentences. Known as Sim, the SYNSTRUTt is calculated the following way: Sim = nodes in the common tree/(the sum of the nodes in the two sentence trees – nodes in common tree) Example: The first tree sentence has 8 nodes and 6 nodes with 4 common nodes. The similarity is Sim = 4/((8 +6)-4) = 4/10 = 0.4.
Syntactic structure similarity adjacent: SYNSTRUTa	Proportion of intersection tree nodes between all adjacent sentences. Measures the	Higher scores in similar sentence structures indicate lower syntactic complexity	Direct: Measured by removing uncommon subtrees found

Coh-Metrix Indices	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
	uniformity and consistency of the syntactic constructions between all adjacent sentences *similarity (Sim) between adjacent sentence pairs in a text*. Looks at syntactic similarity at the phrasal level and the parts of speech. Example 1: The dog (noun phrase) ran (verb). Example 2: It (pronoun) jumped (verb) into (preposition) the pond (noun phrase).	(McNamara, Graesser, McCarthy, Cai, 2014).	between two adjacent sentences. Known as Sim, the SYNSTUTt is calculated the following way: Sim = nodes in the common tree/ (the sum of the nodes in the two sentence trees – nodes in common tree) Example: The first tree sentence has 8 nodes and 6 nodes with 4 common nodes. The similarity is Sim = 4/ (8 +6)-4) = 4/10 = 0.4
DRNP Incidence score of noun phrases	Incidence score of noun phrases. Examples: The big book, the little girl	The density of particular word types (noun phrases) indicates the text is informationally dense, and this indicates complexity (McNamara, Graesser, McCarthy, Cai, 2014).	Direct: Number of noun phrases divided by the number of words multiplied by a 1,000.
Incidence score of agentless passive voice forms. (DRPVAL)	Incidence score of agentless passive voice forms. Examples: A goal was scored in the half time.	Passive construction is more complex than the active sentence. Linguists laid the groundwork for this assumption by assigning a more complex structure to passive sentences (e.g. Chomsky. 1965; Bresnan,	Direct: Number of agentless passive voice divided by the number of words multiplied by a 1,000.

Coh-Metrix Indices	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
		1981: Gazdar. Klein. Pullum, & Sag, 1985). Passive voice is formed by combining a form of the verb to be with the past participle of a transitive verb or modal auxiliary verbs. This increases the level of complexity.	
Negations: Incidence score for negation expressions (DRNEG)	Incidence score for negation expressions Examples: does not, will not, without, none	Negation increases processing difficulty. The use of negation is formed by principal auxiliary or modal verb in a verbal structure. Use of 'not' 'without', 'none' or a combination of a negative word combined with a noun or a pronoun (No girls) increases structural complexity.	Direct: Number of negation expression divided by the number of words multiplied by a 1,000. It is unclear how negative expressions are counted. Is it counted by a single word or the whole phrase. However, the incidence scores are divided by the number of words in text and multiplied by 1,000.
WRDNOUN Incidence score of nouns	Incidence score of nouns Examples: tree, table	The density of particular word types (nouns) indicates the text is informationally dense, and this indicates the sentence is syntactically complex.	Direct: Number of nouns divided by the number of words multiplied by a 1,000. It is unclear how the number of nouns are counted in the essays.

Coh-Metrix Indices	Definition	Rationale	Direct/Indirect Interpretation of Coh-Metrix Measure
WRDVERB Incidence score of verbs	Incidence score of verbs Examples sleep, drink	The density of particular word types (verbs) indicates the text is informationally dense, and this indicates the sentence is syntactically complex.	Direct: Number of verbs divided by the number of words multiplied by a 1,000.
WRDADJ Incidence score of adjectives	Incidence score of adjectives Examples: big, angry	The density of particular word types (adjectives) indicates the text is informationally dense, and this indicates the sentence is syntactically complex.	Direct: Number of adjectives divided by the number of words multiplied by a 1,000.

References

ACT (2005). Crisis at the core: Preparing all students for college and work. Iowa City, IA: ACT. Retrieved from: http://www.act.org/content/act/en/research.html

Anderson, J.C. & D.W. Gerbing (1984). The effect of sampling error on convergence: Improper solutions, and goodness of fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155-73.

Applebee, A. N., Langer, J. A., Jenkins, L. B., Mullis, I., & Foertsch, M. A. (1990). Learning to write in our nation's schools. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Bagozzi, Richard P., & Todd F. Heatherton (1994). A general approach to representing multifaceted personality constructs: Application to self-esteem, *Structural Equation Modeling*, 1 (1), 35-67.

Balioussis, C., Johnson, J., & Pascual-Leone, J. (2012). Fluency and complexity in children's writing: The role of mental attention and executive function. *Rivista Di Psicolinguistica Applicata*, *12* (3), 33-45.

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing: An Interdisciplinary Journal*, *22* (2), 185-200.

Beers, S. F., & Nagy, W.E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing: An Interdisciplinary Journal*, *24* (2), 183-202.

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

Bentler, P.M. and Bonnet, D.C. (1980), Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychological Bulletin*, *88* (3), 588-606.

Bereiter, C., & Scardmalia, M. (1987). The psychology of written composition. Hillsdale, NJ: Erlbaum.

Berninger, V. W., & Swanson, H. L. (1994). Modification of the Hayes and Flower model to explain beginning, developing writing. In E. Butterfield (Ed.), *Advances in cognition and educational practice: Children's writing: Toward a process theory of development of skilled writing* (Vol. 2, pp. 57–82). Greenwich, CT: JAI Press.

Berninger, V. W., Mizokowa, D., & Bragg, R. (1991). Theory-based diagnosis and remediation of writing disabilities. *Journal of School Psychology*, 29, 57–79.

Berninger, V. W., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writers. *Reading and Writing: An Interdisciplinary Journal, 4*, 257–280.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Blair, T. K., & Crump, W. (1984). Effects of discourse mode on syntactic complexity of learning disabled students' written expression. *Learning Disability Quarterly*, *7* (1), 19-29. doi:10. 2307/1510257

Bollen, K.A. (1989). *Structural equations with latent variables.* New York, NY: John Wiley.

Bryant, F.B., Yarnold, P. R., & Michelson, E.A. (1999). Statistical methodology: VIII. Using confirmatory factor analysis (CFA) in emergency medicine research. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine. 6: 54-66*.

Cain, K & Nash, H.M. (2011). The influence of connectives on young readers' processing and comprehension of test. *Journal of Educational Psychology,* 103(2), 429-441.

Chomsky, N. (1957). *Syntactic structures*. New York: Walter de Gruyter

Chung, K.W. K., & O'Neil, H. F. (1997). *Methodological approaches to online scoring of essays.* (ERIC reproduction service no ED 418 101).

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2013). Using automatic scoring models to detect changes in student writing in an intelligent tutoring system. In McCarthy, P. M., & Youngblood G. M., (Eds.). *Proceedings of the 26th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*. (pp. 208-213). Menlo Park, CA: The AAAI Press

Crossley, S.A., & McNamara, D.S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning*, *21* (2/3), pp.170–191.

Crismore, A., Markkanen, R., & Steffensen, M. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. Written Communication, 10, 39–71.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. Behavior

Research Methods, 1–11. http://doi.org/10.3758/s13428-015-0651-7

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

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

Crowhurst, M. (1980a). Syntactic complexity and teachers' quality ratings of narrations and arguments. *Research in the Teaching of English*, *14,* 223–231.

Crowhurst, M. (1980b). Syntactic complexity in narration and argument at three grade levels. *Canadian Journal of Education*, *5*(1), 6-13.

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

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. Assessing Writing, 18, 7-24

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). The Cambridge handbook of expertise and expert performance. New York, NY: Cambridge University Press

Evans, R. V. (1979). The relationship between the reading and writing of syntactic structures. *Research In The Teaching Of English*, *13*(2), 129 – 135.

Fan, X., B. Thompson, & Wang, L. (1999). Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6, 56-83.

Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional Science*, *26*, 473-488. doi:10.1023/A:1003202412203

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

Freedman, S. (1979). How characteristics of student essays influence teachers' evaluations. *Journal of Educational Psychology, 71,* 328–338.

Fuchs, D., & Fuchs, L.S. (2006). Introduction to response to intervention. What, why, and how valid is it? *Reading Research Quarterly. 41.* 93 – 99.

Geiser, S., & Studley, R. (2001). *UC and the SAT: Predictive Validity and Differential Impact of the SAT I and SAT II at the University of California*, Office of the President, University of California, October 29.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long, (Eds.). *Testing structural equation models*. (pp. 40 – 65). Newbury Park, CA: SAGE Publications.

Givon, T. (2009). *The genesis of syntactic complexity.* Amsterdam: J. Benjamins.

Graesser, A.C., & McNamara, D.S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371-398.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of

text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36* (2), 193–202. http://doi.org/10.3758/BF03195564

Graesser, A.C., Gernsbacher, M.A., & Goldman, S.R. (Eds.). (2003). *Handbook of discourse processes.* Mahwah, NJ: Lawrence Erlbaum.

Graesser, A. C. & McNamara, D. S. (2011). Computational Analyses of Multilevel Discourse Comprehension. Topics in Cognitive Science, 3: 371–398. doi:10.1111/j.1756-8765.2010.01081.x

Graesser, A.C., Millis, K.K., & Zwaan, R.A. (1997). Discourse comprehension. *Annual Review of Psychology,* 48, 163-189.

Graesser, A.C. & McNamara, D.S. (2010). Computational Analyses of Multilevel Discourse Comprehension. *Topics in Cognitive Science. 3*(2), 371-398 doi**:** 10.1111/j.1756-8765.2010.01081.x

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review,* 101, 371-395.

Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology.* Mahwah, NJ: Erlbaum.

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323–344). New York, NY: Guilford.

Graham, S., & Harris, K. R. (2005). *Writing better: Effective strategies for teaching students with learning difficulties.* Baltimore, MD: Brookes.

Graham, S., & Perry, M. (1993). Indexing transitional knowledge. *Developmental Psychology*, 2*9*, 779-788.

Graham, S., MacArthur, C. & Fitzgerald, J. (2013). *Best practices in writing instruction*, 2nd Ed. New York, NY: The Guilford Press.

Grobe, G. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English*, *15* (1), 75-85.

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

Harris, K., Graham, S., Mason, L., & Friedlander, B. (2008). *Powerful writing strategies for all students*. Baltimore, MD: Brookes.

Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication, 17*, 307-352. doi: 10.1177/0741088300017003001

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy, & S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences, and applications (pp. 1–27). Mahwah, NJ: Erlbaum.

Hempelmann, C. F., Rus, V., Graesser, A. C., & McNamara, D. S. (2006). Evaluating State-of-the-Art Treebank-style Parsers for Coh-Metrix and Other Learning Technology Environments. *Natural Language Engineering*, *12* (02), 131–144.

Houck, C., & Billingsley, B. (1989). Written expression of students with and without learning disabilities: Differences across the grades. *Journal of Learning Disabilities*, *22*, 561–565.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: Sage Publications.

Hu, L., & Bentler, P.M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6 1-55.

Hunt, K.W. (1965). Grammatical structures written and three grade levels (Research Report No. 3). Champaign, IL: National Council of Teachers of English.

Hunt, K.W. (1970). Syntactic maturity in school children and adults. Monographs of the Society for Research in Child Development, *35* (1), 1- 61.

Jagaiah, T. (2016). Syntactic complexity measures: Variation by genre, grade-level, and students writing abilities. Unpublished manuscript.

Jöreskog, K. & Long, J.S. (1993), Introduction. In K. A. Bollen & J. Scott Long (Eds.), Testing Structural Equation Models, Newbury Park, CA: Sage.

Jurafsky, D., & Martin, J., (2008). *Speech and language processing.* Englewood, NJ: Prentice Hall.

Kagan, D.M. (1980). Syntactic complexity and cognitive style. *Applied Psycholinguistics*, *1*(1), 111-122. doi:10. 1017/S014271640000076X

Kaufman, P., Bradbury, D. & Owings, J. (1992). *Characteristics of at-risk students*. Published in *NELS:88*. Washington, DC: National Center for Education Statistics.

Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. Memory and Cognition, 15 (3), 256-266

King, M., & Rentel, V. (1979). Toward a theory of early writing development. *Research in the Teaching of English*, 13, 243-255.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, UK: Cambridge University Press.

Kline, F. M., Schumaker, J. B., & Deshler, D. D. (1991). Development and validation of feedback routines for instructing students with learning disabilities. *Learning Disabilities Quarterly, 14*, 191–207

Landauer, T., McNamara, D.S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis.* Mahwah, NJ: Erlbaum.

Lane, S.E., & Lewandowski, L. J. (1994). Oral and written compositions of students with and without learning disabilities. *Journal of Psychoeducational Assessment*, *12*(2), 142-153. doi:10. 1177/073428299401200204

Longo, B. (1994). The role of metadiscourse in persuasion. *Technical Communication,* 41, 348-352.

Louwerse, M.M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics,*12, 291-315.

Markels, R. B. (1984). *A new perspective on cohesion in expository paragraphs.* Carbondale: Southern Illinois Press.

McCarthy, P.M., & Boonthum-Denecke, C. (Eds.). (2011). *Applied natural language processing and content analysis: Identification, investigation, and resolution.* Hershey, PA: IGI Global.

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language, 25,* 431– 444.

McCutchen, D. (1994). The magical number three, plus or minus two: Working memory in writing. In J.S. Carlson (series ed.) & E.C. Butterfield (vol ed.). *Advances in cognition and educational practice. Vol.2: Children's writing: Towards a process theory of the development of skilled writing* (pp. 1- 30). Greenwich, CT. JAI Press.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. Educational Psychology Review, 8, 299-324. doi: 10.1007/BF01464076

McCutchen, D., & Perfetti, C.A. (1982). Coherence and connectedness in the development of discourse production. *Text*, 2, 113-139

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

McNamara, D.S., & Graesser, A.C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P.M. McCarthy & C. Boonthum (Eds.). (2011). *Applied natural language processing and content analysis: Identification, investigation, and resolution.* Hershey, PA: IGI Global.

McNamara, D.S., Graesser, A.C., & Louwerse, M.M. (2012). Sources of text difficulty: Across genres and grades, In J. P. Sabatini, E. Albro, & Reilly, T. O. (Eds.), *Measuring up*: *Advances in how we assess reading ability* (pp. 89-116). Plymouth, UK: Rowman & Littlefield Education.

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

McNamara, D.S., & Kintsch, W. (1996). Learning from Text: Effects of prior knowledge and tect coherence. *Discourse Processes*, 22, 247-287.

McNamara, D. S.,Louwerse, M.M.. McCarthy, P. M., & Graesser, A. C., (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes,* 47, 292 -330.

Meyers, L.S., Gamst, G.C., & Guariona, A.J. (2006). *Applied multivariate research: Design and interpretation.* Thousand Oaks, CA: Sage.

Moran, M.R. (1981). Performance of learning disabled and low achieving secondary students on formal features of a paragraph-writing task. *Learning Disability Quarterly*, *4*(3), 271-80.

Morris, N.T., & Crump, W. (1982). Syntactic and vocabulary development in the written

language of learning disabled and non-learning disabled students at four age levels. *Learning Disability Quarterly*, *5*(2), 163-172. doi:10.2307/1510577

Myhill, D. & Jones, S. (2009). How talk becomes text: Investigating the concept of oral rehearsal in early years' classrooms. *British Journal of Educational Studies*, 57: 265–284. doi: 10.1111/j.1467-8527.2009.00438.x

Myklebust, H. R. (1973). *Development and disorders of written language (Vol. II): Studies of normal and exceptional children.* New York: Grune and Stratton.

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

National Commission on Writing (2003). *The neglected ''R.''* College Entrance Examination Board, New York.

Newcomer, P. L., & Barenbaum, E. M. (1991). The written composing ability of children with learning disabilities: A review of the literature from 1980 to 1990. *Journal of Learning Disabilities*, 24, 578-593.

No Child Left Behind Act of 2001, 20 U.S.C. § 6301, *et seq*. (2007).

Noeth, R. J., & Kobrin, J. L. (2007). *Writing changes in the nation's K-12 education system.* Research Notes-34. Iowa City, IA: The College Board.

Newcomer, P. L., & Barenbaum, E. M. (1991). The written composing ability of children with learning disabilities: A review of the literature from 1980 to 1990. Journal of Learning Disabilities, 24, 578-593.

Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan, 48*, 238-243.

Page, E. B. (1994). Computer grading of student prose, using modern concepts and software. *The Journal of Experimental Education, 62*(2), 127-142.

Page, E. B. (2003). Project Essay Grade: PEG. In M. D. Shermis, & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp.43-54). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Perron, J.D. (April, 1977). *Written syntactic complexity and modes of discourse*. A paper presented at the annual meeting of the American Educational Research Association, New York. (ERIC Document Reproduction Service No. ED 139 009)

Perry, J.L., Nicholls, A.R. Clough, P.J. Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in Physical Edcuation and Exercise Science. 19,* 12- 21.

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences,* 27, 169 – 226.

Prater, D. L., & Mayo, N. B. (1984). Cognitive developmental level and syntactic maturity. *Journal of Research and Development In Education*, *17* (3), 1-7.

Price, G.B., & Graves, R. L. (1980). Sex differences in syntax and usage in oral and written language. *Research in the Teaching of English*, *14* (2), 147-53.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language*, *30* (1), 3-26. doi:10. 1177/0142723709350531

Rodrigues, R. J. (1980). Bilingual and monolingual English syntax on the Isle of Lewis, Scotland. *Research in the Teaching of English*, *14* (2), 139-46.

Rousseau, M. K., Bottge, B. A., & Dy, E. B. (1993). Syntactic complexity in the writing of students with and without mental retardation. *American Journal on Mental Retardation*, *98* (1), 113-120.

Rubin, D. L. & Piche, G.L. (1979). Development in syntactic and strategic aspects of audience adaptation skills in written persuasive communication. *Research in the Teaching of English*, *13*, 293 – 316.

Saddler, B., Asaro, K., & Behforooz, B. (2008). The effects of peer-assisted sentence-combining on four young writers with learning disabilities. *Learning Disabilities: A Contemporary Journal, 6*, 17–31.

Saddler, B., Behforooz, B., & Asaro, K. (2008). The effects of sentence-combining instruction on the writing of fourth-grade students with writing difficulties. *Journal of Special Education, 42*, 79–90. doi:10.1177/0022466907310371

Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence-combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology*, *97*, 43–54.

San Jose, C.P.M. (1972). Grammatical structures in four modes of writing at the fourth grade level. Unpublished doctoral dissertation, Syracuse University.

Sanford, A.J., & Emmott, C. (2012). *Mind, brain and narrative.* Cambridge: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (1986). Written composition. In M. Wittrock (Ed.), Handbook on research on teaching (3rd ed., pp. 778–803). New York: Macmillan.

Schultz, J. M. (1994). Stylistic Reformulation: Theoretical Premises and Practical Applications. *The Modern Language Journal, 78*(2)169-178.

Secretary of the State of Connecticut. (2010). *Connecticut mastery test fourth generation*: *Interpretive guide.* CT: Connecticut State Board of Education.

Shermis, M. D., & Barrera, F. D. (2002). Automated Essay Scoring for Electronic Portfolios. *Assessment Update*, *14*(4), 1.

Shermis, M. D. & Burstein, J. (2003). *Automated Essay Scoring: A cross disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

Simon, H. A., & Chase, W. G. (1973). Skills in chess. American Scientist, 61, 394-403.

Singer, M., & Leon, J. (2007). Psychological studies of higher language processes: Behavioral and empirical approaches, In F. Schmalhofer & C. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp.9-25). Mahwah, NJ: Lawrence Erlbaum.

Smith, W. L. (1974). Syntactic recording of passages written at three levels of complexity. *Journal of Experimental Education*. *43* (2) 66 – 72.

Smith, W.L., & Swan, M. B. (1978). Adjusting syntactic structures to varied levels of audience. *Journal of Experimental Education*. *46* (4) 29 – 34.

Stewart, M. F. (1978). Syntactic maturity from high school to university: A first look. *Research in the Teaching of English*. *12*(1), 37 – 46.

Stewart, M. F., & Grobe, C. H. (1979). Syntactic maturity, mechanics of writing, and teachers' quality ratings. *Research in the Teaching of English*, *13* (3) 207-15.

Strong, W. (1986). Creative approaches to sentence combining. Urbana, OL: ERIC Clearinghouse on Reading and Communication Skill & National Council of Teachers of English.

van de Kopple, W.J. (1985). Some exploratory discourse in metadiscourse. *College Composition and Communication,* 36, 82-93.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Wagner, R. K., Puranik, C.S., Foorman, B., Foster, E., Tschinkel, E., & Kantor, P.T. (2011). Modeling the development of written language. *Reading and Writing: An Interdisciplinary Journal*, *24*, 203–220.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin,* 123, 162-185.