WRITING ASSESSMENT IN MIDDLE SCHOOL STUDENTS: ANALYZING SPELLING WITHIN A MULTIDIMENSIONAL LANGUAGE FRAMEWORK

By

Lake Eiseler Sweet

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

School Psychology- Doctor of Philosophy

2022

ABSTRACT

WRITING ASSESSMENT IN MIDDLE SCHOOL STUDENTS: ANALYZING SPELLING WITHIN A MULTIDIMENSIONAL LANGUAGE FRAMEWORK

By

Lake Eiseler Sweet

Although producing quality written expression is a vital skill, many students in the United States struggle to produce proficient written language. There are many academic and career outcomes related to the ability to produce written expression, yet many schools lack formalized writing assessment and instruction. As such, many questions remain related to individual differences in writing ability and best practices in assessment and instruction.

To answer these questions, it is necessary to establish a model of written expression and what specific variables exist within the model to be used to assess written language. Modern writing assessment theory uses levels of language as a framework with commonly assessed dimensions of accuracy, complexity and productivity. This framework has yet to be firmly established in the literature, and the variables included in each level are just beginning to be explored. One salient variable in writing research, assessment and instruction is spelling ability, and how this ability may influence the production of written language.

This study furthers the work by Wilson et al. (2017), Troia and colleagues (2019) and many others (e.g., Berninger et al., 2006; Flower & Hayes, 1981) with the ultimate goal of developing a model of written language to guide assessment and instruction in schools. Specifically, data were drawn from Truckenmiller and colleagues (2020) study piloting a writing assessment tool, Writing Architect, which sampled 526 students from third to eight grades; this study used sixth, seventh and eighth grades with a resulting sample size of 290 students.

Results indicated spelling was a significant predictor of writing quality, in that better spelling indicated better writing quality. The same was true for text. For the sentence-level variable, a higher score indicated worse writing quality in a significant way. The word variable did not significantly predict writing quality in the model. The significant interaction between spelling and text variables suggests that the effect of text on writing quality is even higher when spelling ability is also high.

Findings highlight the importance of writing and spelling instruction in school. The findings for this age group help identify how writing abilities may change over the trajectory of development and vary individually. Additionally, this analysis echoes the call for further research to establish variables for automated writing assessment.

ACKNOWLEDGEMENTS

I would like to start by thanking many individuals who have contributed to my personal and professional growth. First, Dr. Jodene Fine, who has provided support, guidance, and importantly, constructive criticism when it is necessary. Dr. Fine has been instrumental in shaping my confidence and skills as a researcher and clinician. Second, I would like to thank Dr. Adrea Truckenmiller for allowing me to collaborate with her on this research and for going above and beyond in her willingness to discuss and guide my learning on these topics. I would also like to express my gratitude to Dr. Martin Volker, for his support always, and for his willingness to serve on this committee. I would like to thank Dr. Troy Mariage for agreeing to serve on my committee and providing thought provoking feedback, especially related to practice. Finally, I would like to thank Nicole Jess for her support with statistics as I worked and learned through this process.

Personally, I would like to thank John, Grover and Laura, for continued support and constant encouragement. You helped me find a necessary balance between work and play. Thank you for your love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

Producing quality written expression is a vital skill for students in the 21st century. To succeed in elementary, middle, high school and careers, the ability to express thoughts through writing is vital. Communicating through written expression, critical in all stages of life, is important for social, cultural and academic pursuits (Graham, 2006). Writing is used to influence others, communicate ideas and learn (Graham, Gillespie & McKeown, 2013). Writing is crucial in the development of other academic skills, especially reading (Graham & Hebert, 2011).

Outcomes related to written expression include grade retention, standardized assessment success and graduation (Jenkins et al., 2004). However, according to the National Assessment of Educational Progress, about 25% of students score at the proficient level in writing assessment (2011). Beyond this, many schools lack formalized writing assessments and interventions to address this established need. With the increased focus on supports, interventions and individualized instruction in the current educational climate, the question remains: why are so many students still failing to write proficiently? And how can we accurately assess the areas in which they struggle for targeted instruction and intervention?

Understanding why and how students differ in their writing abilities is an important first step to addressing these issues. Researchers need to find variables and measures that differentiate good and poor writers to begin to understand what is constraining the writing ability of some students. To do so, there first must be an understanding of the different components of written language and how they interact to influence writing quality. Beyond research to guide understanding of why students struggle to write, evidence-based interventions in writing also depend on detailed analysis of written work.

Assessment of writing is problematic on many fronts. Even if there were a universally used writing assessment, teachers struggle with finding the time to provide qualitative or quantitative assessment of written samples for students. Innovations in technology have allowed for new types and levels of analysis of written language. With new advances in computational software, samples of written language can be analyzed for hundreds of variables at word, sentence and text-level with ease. However, understanding which of these variables are meaningful for models of writing influencing instruction and intervention is not yet established.

Assessment in writing today is often based on the writing theory, which generally includes the relations between transcription, cognitive skills, text generation and working memory (e.g., Berninger et al., 1996). However, more work is needed to understand the interaction of the variables within the model and their influence on writing quality. Further, this model has not been studied at all developmental levels with consideration for variation within individuals.

A recent article by Wilson and colleagues (2017) analyzed how different variables from online text analysis software fit into a framework incorporating different levels of language, and how this model predicted writing outcomes for students. Results of this study were promising in terms of providing fast, meaningful analysis to guide instruction and intervention in student writing. However, more work analyzing these variables and their fit into a model of writing to guide assessment is necessary. Building on the work by Wilson et al., (2017) Troia and colleagues (2019) analyzed the same levels of language, adding commonly assessed dimensions of language (accuracy, complexity, productivity) to make a 3x3 matrix of areas of written expression for analysis. The Troia et al. (2019) work to establish a model and the interaction of

levels and dimensions of language has the potential to transform how writing is assessed and instructed in schools.

This study aims to further the work by Wilson et al. (2017) , Troia and colleagues (2019) and many others (e.g., Berninger et al., 2006; Flower & Hayes, 1981) with the ultimate goal of developing a model of written language to guide assessment and instruction in schools. An evidence-based model of the levels of language is important to establish to understand why students struggle with writing, in what specific ways they struggle, and to monitor their growth in writing.

Specifically, this study will look at the influence of spelling on writing quality overall and at specific levels of language. This study aims to identify to what extent spelling blocks or assists the development of components of writing. By understanding the role of spelling in language development, we can understand how to intervene in spelling to improve written language abilities in students. The ultimate goal of this study is to further the model of the components of written language and how they interact with each other to create effective, efficient assessment and intervention of writing.

The outcomes of interest in this study are overall writing quality, with a focus on the complexity of word-, sentence- and text-level variables of complexity. A major goal of the study is a specific analysis of the role of spelling ability and its interaction with all levels of the model. Covariates included in the model include special education status, gender, grade level and race/ethnicity.

CHAPTER TWO: LITERATURE REVIEW

Model of Written Language: Theory

To lay a foundation for empirical research on writing theory, it is important to understand how modern writing theory came to be. In this aim, it is also important to understand how certain variables within the theory have shifted and changed over time. An exploration of major theories of written expression leads to a more complete picture of the components of modern writing theory that are to be tested in this study: specifically, the different levels of language within theories, the performance dimensions of written language commonly assessed and specifically, spelling. Establishing these components and the reasons they are included in theories of written language will lead to a clearer understanding of why language is assessed the way it is, and why it is important to understand the relationships between these variables with the ultimate goal of a better way to assess written language.

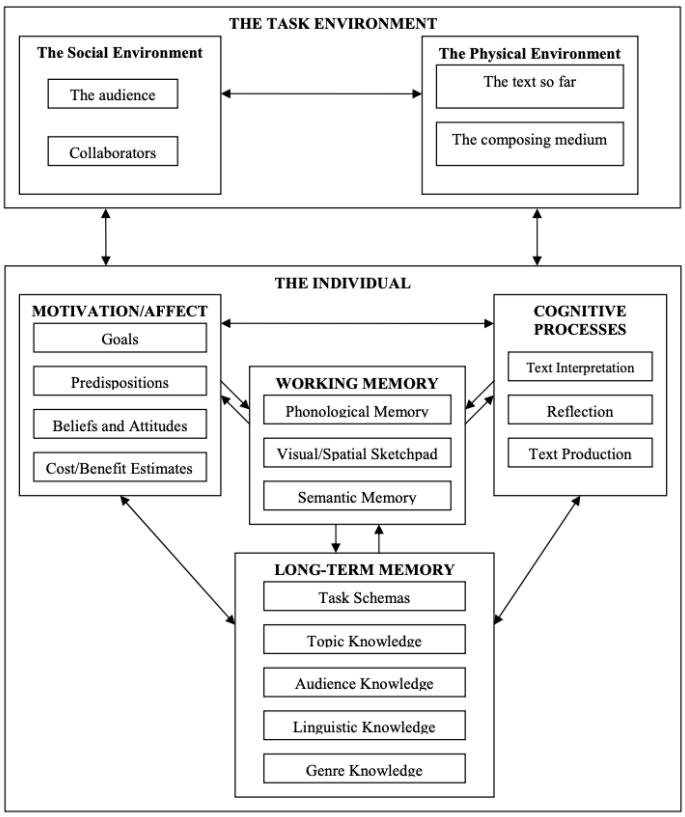
There are many theories explaining the writing process and the individual components and skills necessary to write. Pre-1980, these theories mainly focused on the sequential stages of writing such as planning, drafting and revising (e.g., Ashbaugh, 1921, Littwin, 1935, Crowley, 1977). In 1981, Flower and Hayes published a capacity theory of writing, focusing on the cognitive processes involved in writing versus the stages of the process. In 1982, Berninger and Garvey published an article that introduced the different levels of language in writing, guiding an understanding of both the development of written ability and how to assess written language. These theories together inform our understanding of written expression and writing assessment today.

Flower and Hayes (1981) published one of the first widely accepted theories of writing that included cognitive components and focused on the process of writing instead of the finished

writing product. This theory focused on planning, translating and reviewing as the three main components in the cognitive process of writing. Hayes and Flower (1981) posited that writing is made up of distinct cognitive processes, and propose a model explaining this instead of the existing stage model of writing (see Figure 1.). Text generation (planning), according to Flower and Hayes (1981) involves coming up with ideas and translating these ideas into language. Transcription (translating) involves translating language into writing (typing or handwriting). They also put forth the idea that these components are organized in a hierarchical manner, and that the activity is goal-directed.

Figure 1.

Flower and Hayes (1996) Model of Writing



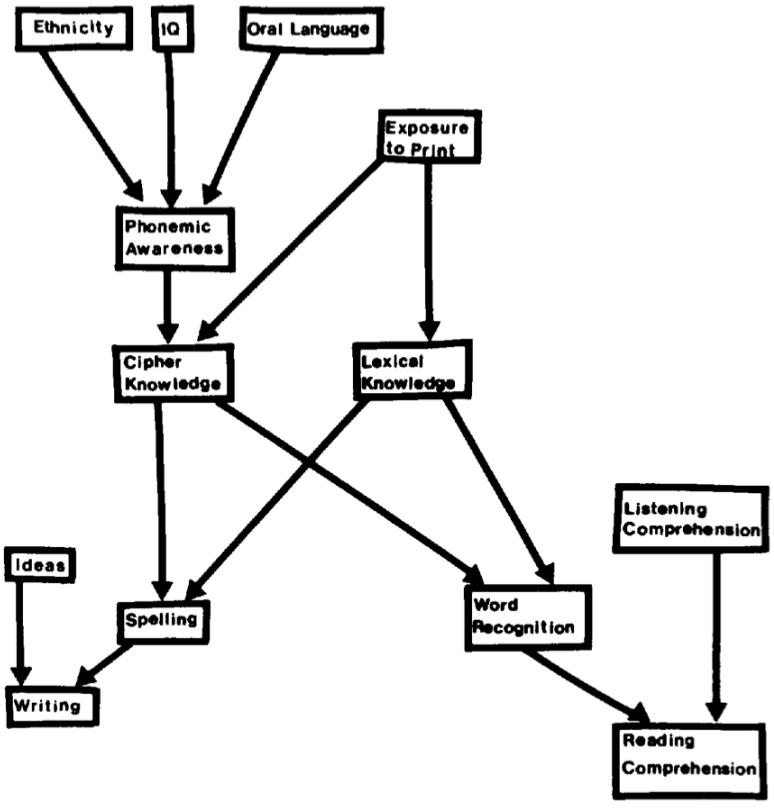
Frameworks exploring levels of language go back as far as the early 1980s as well, with work by Berninger and Garvey in 1982. Their work began to quantitatively analyze oral

communication between children. This in-depth examination of language structure, grammar and syntax was the beginning of much work by Berninger and colleagues on the understanding of the structure of different levels of language instead of just language as a whole.

Juel, Griffith, and Gough (1986) tested a simple view of reading and writing theory (see Figure 2., below). In their longitudinal analysis, the authors examined many different levels of language and components of writing, including phonemic awareness, spelling, lexical knowledge, ideation and story writing. Their theory importantly considered the lower level skills necessary to spell, and how spelling, in turn, influences writing quality. These authors laid a foundation for understanding the role of spelling in writing, and how this skill may facilitate or hinder the development of other writing skills.

Figure 2.

Juel et al. (1986) Model of Literacy Acquisition



In later work, Juel (1988) introduced the most simplified model of writing, on which most later models of writing are based. This model includes the components of spelling and ideation (generating/developing and communicating ideas). All later models feature spelling and ideation prominently, with or without other processes that are associated with these skills. However, again, more research is needed to quantify the relationship between these important variables.

Many other researchers have added to the components and skills involved in the writing process. Whitaker, Berninger, Johnston and Swanson (1994) expanded on earlier theories such as Juel's 1988 simple view of writing and added levels of language in the ideation component of the theory. These levels of ideation include word, sentence and text (paragraph). The levels of language framework also had roots in work by Durst and Newell (1989) and van Dijk and Kiptsch (1983) and others, who introduced and conceptualized the different levels of language, and began the work of empirical research on these processes.

The levels of language are critical in writing assessment, allowing for an understanding of how different skills develop and interact in individuals in prediction of written ability. However, work is needed beyond establishing the existence of different levels of language as discrete variables. How do these levels vary in individual ability, and how can this be measured? How do these levels interact in their prediction of writing quality? And what other variables influence the development and interaction of the levels of language?

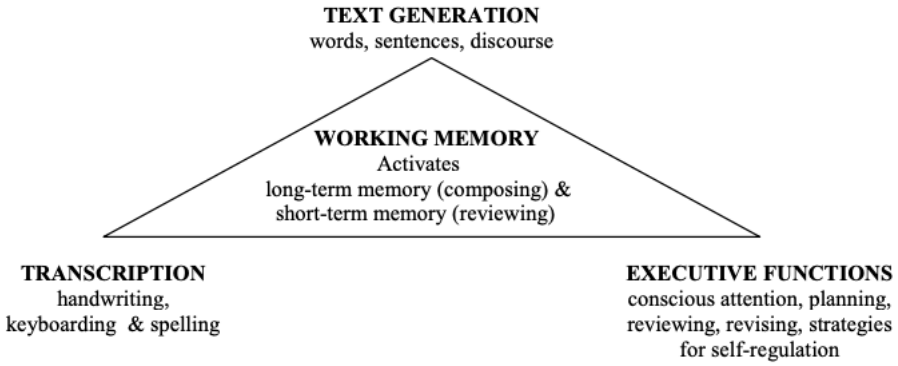
McCutchen (1996) introduced theory that focused on the variable of working memory in the writing process, especially as it had been applied to reading research. McCutchen also applied established reading theory to writing and the writing process. Processing and storage capacity were of specific importance to the theory, which she examined in relation to Flower and

Hayes components of writing. McCutchen argued for a capacity theory of writing, where different components of writing use cognitive resources, and where lower and higher-level processes compete in complex ways for these resources, influencing writing quality. McCutchen used research on working memory and reading, and current theories of writing, and attempts to combine the two.

McCutchen also introduced the idea of the competition between different components of writing and begins to think about the ways in which these components interact according to a capacity theory to influence writing quality. Understanding why different levels of language may exert influence over other levels has foundations in capacity theory. This theory was influential in understanding why children were not able to cognitively operate all of the skills necessary for writing at once. The cognitive capacity was a limiting factor in writing ability. However, how exactly this capacity functioned and how the skills competed for the cognitive resources was not established in the original theory and has yet to be firmly established.

Figure 3.

Berninger & Amtmann (2002) Simple View of Writing



In later work, Berninger and colleagues (2002) investigated a simple view of writing in third grade students (see Figure 3., above). The simple view of writing in its simplest form, is

broken down into the components of transcription and ideation. Transcription includes the mechanics and conventions of written language (handwriting, grammar, etc.), whereas ideation includes planning to write, drafts, word level complexity, and structure. Specifically, their experimental study focused on struggling writers and investigated instructional practices. Their study was based on theories that spelling (a transcription skill) may cause struggles in writing or alternately, planning, text generating, reviewing, and revising may cause struggles. The sample included 96 third grade students who met certain inclusion criteria (including verbal IQ within the average range) were randomly assigned to one of the four groups. The three treatment groups had instructional components related to the theory described above. Assessments administered following treatment included handwriting automaticity, normed measure of spelling, spelling inventory, compositional fluency and compositional quality.

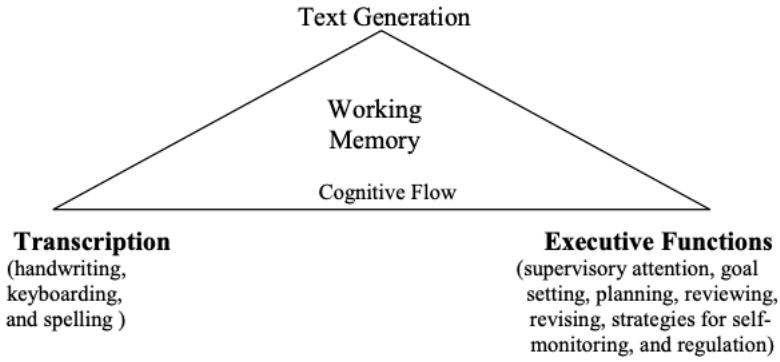
The authors found that for children in the treatment conditions for spelling and quality of writing, explicit instruction increased the children's skills in the areas in which they were instructed, and that composition training increased writing quality. This study was important for understanding how separate skills involved in the writing process may develop in different individuals. The paper by Berninger and colleagues was based on Berninger's simple view of composing (Berninger, 2000), which the author states is based on cognitive, developmental, neuropsychological and educational research. It posits that writing is based on a triangle of transcription, self-regulation of executive function, and text generation (at the top of the triangle, see Figure 3.).

If transcription skills are automatic, more working memory is available for higher level skills (text generation). This leads directly to Kim and colleagues (2018) investigation of text generation (fluency). Berninger also mentions in the article that in the theory, executive function

is involved in processes like goal setting, planning, reviewing and revising, which is traced directly to the McCutchen (1996) paper introducing capacity theory.

Figure 4.

Berninger et al. (2006) Not so Simple View of Writing



In further work by Hayes and Berninger (2014), the authors offer an integrated, cognitive theory based on empirical research (see Figure 4.). The theory attempts to take cognitive theories and apply them to writing specifically to inform a cognitive theory of writing. The theory includes a developmental perspective of the development of writing abilities in children. The authors put forth the integration of three levels of writing production: resource, process and control (Hayes & Berninger, 2014). At the resource level, cognitive processes such as attention, long-term and working memory and reading ability are included. At the process level, the task environment interacts with components of the writing process. At the control level, planning and schemas and task initiation operate. These levels are integrated in skilled writers (usually typically developing, older children), and prior to integration, writing is constrained by whichever level takes the most processing (Hayes & Berninger, 2014). Text generation, including the components of proposing, transcription, translation and evaluation constrains

writing ability/quality until 5th grade, according to the model (Hayes & Berninger, 2014). In the intermediate grades, the processes mature and become automated (requiring fewer resources).

Kim and colleagues (2018) operationalize the components of writing included in their theory: transcription, text generation and writing fluency. Transcription was defined as encoding sounds into print and includes handwriting and spelling. Text generation was operationalized as generating ideas and encoding ideas at word, sentence and text level as oral language skills.

Previous models had incorporated text generation and transcription, but, according to Kim and colleagues (2018), the previous models had failed to consider the influence of writing fluency. Writing fluency was defined as automaticity in writing and occurs at the word, sentence and text level. Kim et al. (2018) tested this model that incorporated writing fluency along with text generation and transcription in prediction of overall writing quality. Writing fluency was measured using the percentage of correct writing sequences in written text within a certain period of time (e.g., five minutes). The authors hypothesized fluency would mediate the effects of transcription and text generation in their prediction of writing quality for second and third grade students.

Finally, the authors hypothesized that the relationships between these components would change with the individual's development of the individual skills. Younger children would be constrained by transcription skills, which would result in a weaker relationship between text generation and fluency. The authors also theorized that fluency would be more strongly related to writing quality as skills develop.

Data analysis was completed using confirmatory factor analysis and structural equation modeling. Results from a multi-group structural equation model indicated handwriting fluency was related to text writing fluency and spelling was related to text writing fluency for second

graders; oral language was not related to fluency, nor was the outcome of writing quality. For third grade students, and writing fluency and spelling were related to fluency, as was oral language. Oral language, handwriting fluency, and text writing fluency were related to writing quality.

This study importantly uses empirical research to attempt to understand the way many skills interact in their prediction of writing quality. It is an excellent example of the current exploration of how different components of written language interact in their prediction of writing quality, with consideration of individual differences. The study considers different components of written language, such as fluency and spelling ability and examines how they change in relation to each other, other abilities involved in written expression and overall writing quality. However, the study focuses specifically on one age group, and research needs to further this understanding by examining all ages or conducting longitudinal studies of written language development.

The study by Kim and colleagues (2017) is an example of empirical research investigating a modern theory of writing. An understanding of writing theory allows for ideas of how different skills and components must be developed to produce written language. While overall quality of written expression is important, understanding how these skills and processes interact with each other and with overall quality is crucial, and this study is a step in the direction of quantifying variables in modern writing theory. This modern theory, explained below, is also the basis for the current study, which aims to expand the quantitative understanding of the components of writing theory and their interactions with each other. First, a summary of modern writing theory is necessary.

Modern Writing Theory

The modern theory of writing used in this study continues to be based heavily on the seminal work of Juel in 1988. This simplified model of writing that includes spelling and ideation heavily influenced all later writing models, whether they included additional skills and components or not. Models following Juel's simple view of writing (1988) often broke ideation into different levels of language (word, sentence and text). The levels of language are a common component of many of the theories of written language throughout history Abbott, Berninger, & Fayol, 2010; Hayes & Berninger, 2014. This hierarchical framework is useful for evaluating writing performance but has limited support in terms of empirical research (Troia et al., 2019).

Each of these levels of language is also multidimensional and includes productivity, accuracy and complexity of language (e.g., Troia et al., 2019). These performance dimensions are commonly used to guide the assessment of written language (Troia et al., 2019). Productivity involves quantitative measures such as number of words written, words written in time constraints and diversity of words written. Accuracy includes measures of spelling and grammar. Complexity of written language involves word choice, sentence structure and length, and semantic and content analysis of paragraphs. These variables, productivity, accuracy and complexity, are woven throughout the history of the theories of written language. These performance dimensions are common in assessing writing as well as foci of interventions.

Accuracy of Written Language

Accuracy of written language is a dimension of performance that measures skills such as spelling of words, capitalization, grammar, and punctuation. These conventions are important to convey ideas clearly and are also related to overall quality of writing in children (Olinghouse, 2008). Previous research indicates children who struggle with mastery of language because of

disability struggle with accuracy (Gillam & Jonston, 1992). The ability to produce accurate written language also interacts with the other dimensions in its prediction of writing quality.

Productivity of Written Language

Measures of productivity in the levels of language framework include things like correct word sequences, word count, fluency (words written in a time constraint), and number of T-units (independent clauses). Research on productivity of written language has found individual differences in these measures related to grade level and writing ability. This was summarized by Troia and colleagues (2019) as older and more skilled writers produce more written language.

Complexity of Written Language

Word-level complexity involves vocabulary and word choice in writing (National Assessment Governing Board, 2010). A higher word-level complexity in a writing sample depends on many variables, including motor skills, phonological skills, vocabulary knowledge (Dockrell et al., 2009). Sentence-level complexity involves syntactical, mechanical and grammatical choices (National Assessment Governing Board, 2010). Text-level complexity of written language includes measures of organization, cohesion and development of ideas. As with other performance dimensions, complexity measures have been correlated with better writing skill and age (e.g., Beers & Nagy, 2009).

Summarized Modern Theory

Historically, many theories have built upon each other to bring us to the way we view writing today: a modern writing model that includes cognitive processes and resources, interacting with each other in complex ways to produce written language. Importantly all theories since Juel's simple view of writing in 1988 have included the processes of spelling and ideation. Ideation (or generating, developing and communicating ideas) can be viewed has

including different levels: the word, sentence and text levels, which are useful for assessment and intervention. Beyond these levels, written language can also be quantified in different ways in terms of productivity, accuracy and complexity: the performance dimensions of written language.

The majority of writing research focuses on the writing process, and not on the levels of language involved in the ability to produce writing (Abbot, Berninger and Fayol, 2010). This framework of the levels of language is helpful for understanding writing assessment, instruction and intervention. Although grounded in writing theory, research on this framework is still developing (Kim et al., 2019, Troia et al., 2019; Wilson et al. 2017).

Troia and colleagues published a 3x3 matrix in 2019 that combined the levels of language frameworks with the commonly assessed components of written language: productivity, accuracy and complexity. Using this framework, Troia et al. (2019) analyzed student written work. This built on work by Wilson and colleagues (2017) that analyzed complexity at the word, sentence and text level in written expression. Troia and colleagues (2019) found differences in performance across grade and writing ability for the variables at different levels of language.

Their results (see Figure 5.) indicated certain variables predicted writing quality in their sample, and also revealed which variables differed by grade level and by ability level. Troia and colleagues (2019) found that overall, type-token ratio (total number of unique words divided by total number of words) for content words, total words written, narrativity (the degree of word familiarity, similarity, simplicity and cohesion represent narrative structure; Troia et al., 2019), percent word accuracy (total number of errors in capitalization and spelling divided by total words), percent grammatical sentences, mean punctuation errors per sentence and handwriting style (evaluated using four categories- manuscript only, mostly manuscript, mostly cursive or

cursive only) were all significant predictors of narrative quality. Further analysis revealed type-token ratio of content words, total words written, mean textual lexical diversity (average length of word strings that maintain a certain level of lexical diversity, a measure of complexity of text; Troia et al., 2019), mean syllables per word, mean content word frequency, narrativity, percent word accuracy, mean punctuation errors per sentence, handwriting style and process use (evaluation of student work for evidence of using a writing process) all had grade level changes in variables. Finally, variables that differentiated between poor and good writers included type-token ratio of content words, total words written, mean textual lexical diversity, mean content word frequency, mean words per sentence, percent word accuracy, percent grammatical sentences, handwriting style and process use.

The results from Troia et al.'s (2019) study indicate many factors vary at individual, grade and ability levels, and variables at different levels of language and dimensions of discourse influence writing quality. However, the ways these variables interact in their prediction of writing quality needs to be further fleshed out to develop a solid model. Additionally, some results are confounding, and need to be assessed in more samples to apply the results to a population.

Table 1.

Significant Predictors of Narrative Quality

	Word	Sentence	Text
Productivity	Type-token ratio for content words		Total words written
Accuracy	Percent word accuracy	Grammatical sentences, mean punctuation errors	
Complexity			Incidence of connectives narrativity

Troia and colleagues' 2019 work is important in establishing some quantitative levels for how these variables interact in a model of written language. The study used a sample of 362 fourth through sixth grade students from schools in the Midwest. The majority of participants (63%) were white (8% African American, 6% Latino/a, 5% Native American, 1.5% Asian American, 15.5% other/multiethnic students; Troia et al., 2019). This work can be expanded upon by examining other grade levels, inclusion of students from diverse backgrounds and examination of ability levels.

As seen through the review of literature, much work has been done recently in the detailed analysis of the components of language, moving toward more comprehensive and informative assessment. However, much work is still needed to understand how the identified components of writing interact in individuals. A modern approach to assessing writing, based on automated, evidence-based tools is the next frontier in writing research and instruction.

Also, in an effort to move forward the literature on this model of writing, this study aims to add a specific analysis of a writing skill, spelling. Analyzing spelling specifically and how it constrains other skills in the model in their prediction of writing quality is meaningful for understanding the model overall. Spelling is a skill taught in classrooms and has been found to influence writing in students, but the specific way in which it operates needs further examination.

Spelling

Specifically, this study aims to examine a word-level, performance dimension of accuracy: spelling. The study is specifically examining how spelling constrains the complexity of other levels of language in their prediction of writing quality. Understanding the role of spelling in this model has potential for improving the assessment of written language and for designing targeted writing intervention.

Understanding how spelling ability influences the levels of language and their prediction of writing quality is meaningful for instruction and intervention. Previous research on spelling ability has identified some patterns in children's spelling and its relationship to writing ability. Students with greater spelling skills take more changes; on the opposite end of the spectrum, students who struggle to spell get stuck on spelling words correctly and don't engage in higher order cognitive processes (i.e., those involved in sentence and text levels of writing) (Beers & Nagy, 2009). Understanding this relation between spelling and the hierarchical word, sentence and text levels of language illustrates how spelling ability can constrain what children choose to write.

Further, research has identified spelling as variable that constrains writing quality in many ways. Spelling ability constrains text production (speed) and the quality of text produced when children were asked to write, but not when they were asked to dictate a text (Sumner, Connelly, & Barnett, 2014). In studies of text production, children were found to be slower in producing text at the word and sentence level when they had lower spelling abilities (Sumner, Connelly, & Barnett, 2013). Other research has expanded on this correlation, finding kids who often know word is spelled incorrectly often fail to move past trying to spell the word. Because of this, they do not engage in higher order cognitive processes necessary to produce written language (Beers & Nagy, 2009).

In one of the only studies on the levels of language framework in writing, Abbott, Berninger and Fayol (2010) found that individual differences in spelling accounted for unique variance in both word-level spelling and text-level composition in all grades included in the study (first through seventh). The authors interpreted this significant finding as individuals with better spelling skills were more likely to translate the ideas into words and the words into written

text. Alternately, the authors suggested an explanation related to working memory; word-level working memory is related to all writing skills and spelling at these grades. The authors went on to note that the relation between text-level language and spelling indicated these skills develop concurrently, especially during middle school. The relation found between spelling and text-level writing also led the authors to hypothesize a possible top-down interrelation between the variables.

Abbot, Berninger and Fayol (2011) also found that spelling was the most stable skill across grade levels of the writing-related skills they examined. The authors used this finding to argue for the significance of spelling in children's writing education, as their results also indicated spelling was significantly related to other writing skills. However, in their literature review, they identified a lack of strategic, consistent spelling instruction in classrooms across the country (Abbott et al., 2010).

Clearly, spelling influences the production of written language. However, some studies indicate it influences the word-level of language, some indicate the sentence-level, some indicate the text-level, and some indicate multiple levels. There is no consensus on how spelling influences written language, which levels it influences most, or how it specifically influences written language for different individuals. This relation is clearly complex and has potential to inform writing instruction and assessment.

Writing Assessment

This empirical research aims to forward the literature by providing a quantitative analysis of the ways components of writing interact. Examination of writing theory leads to a clearer picture of what variables should be assessed in writing assessment and why they are important. Understanding how variables in writing theory interact with each other in practice can

revolutionize the way educators assess and intervene in writing and spelling. The next important question is, how are these variables assessed in modern writing assessment? And what are the needs that still exist in assessing writing? The first step in doing so is understanding the current status of writing assessment.

Writing is difficult to assess for many reasons. First, assessing writing in a standardized way is problematic because of a lack of standardized variables to be measured, and a lack of agreement on variables for overall measures of quality. Second, there is a lack of standardized assessments that measure both overall writing quality and the minute variables of the different levels of language that would be informative to guiding instruction and intervention.

The approaches to writing assessment have changed much in recent years as writing research has advanced. Trait rubrics have been commonly used in writing assessment, asking evaluators to score writing samples on a certain number of traits of writing such as 6 + 1 Traits of Writing (Culham, 2003). However, research has indicated these trait rubrics may not be the most accurate or meaningful measure for assessing writing. For example, the traits share variance (Troia et al., 2015), indicating the scores for each trait would not provide information to guide intervention or instruction. That the traits in the rubrics share variance indicates they may not be measuring separate constructs, and therefore may not be providing meaningful information on distinct skills or variables in the student's writing. Further, according to Troia and colleagues (2019), scoring writing with trait rubrics can be intensive in terms of time and money.

One common method of writing assessment, using formative feedback, seems to be beneficial to developing writing skills (Graham et al., 2007). Formative assessment is providing qualitative and quantitative feedback that is designed to improve student writing. Commonly referred to in practice as 'assessment FOR learning', formative assessments include collecting

data on student learning, interpreting and using it to guide instruction (William, 2006). Types of formative assessments for writing include feedback from teachers, computer software, scoring rubrics and self-assessments (Graham et al., 2007).

The evidence for the effectiveness of these types of assessments varies widely (Graham et al., 2015). The benefit of formative assessment varies by the type used, and the downside is the time necessary to collect the data necessary to provide meaningful feedback. Some types of formative assessment are more labor intensive then others, but also more effective at promoting skills in students. Other types of formative assessment are not valid/reliable and are too dependent on individual judgement. For example, a subjective rubric, teacher feedback or self-assessment may not be a reliable or valid source of feedback.

A real need for educators and students is less labor-intensive way for teachers to assess written expression and provide specific feedback. This feedback, in turn, can lead to specific instruction and intervention. Automated tools for assessing writing would allow this, but researchers first need to establish which components of writing are meaningful for identifying struggling writers and for intervention.

Coh-Metrix

Advanced in technology and models of writing assessment have led to the development of software designed to analyze different components of written language. Coh-Metrix is an online software used to analyze written (typed) text at the word-, sentence- and text levels using 200 measures of cohesion, language and readability (Graesser et al., 2004). Coh-Metrix was developed to evaluate texts for use in schools. Texts used in all levels of education are commonly assessed using readability formulas, which use word and sentence length to evaluate text. However, these readability formulas are thought to be too simplistic in their analysis of text

and are often misused and misunderstood both by text manufacturers and consumers (Graesser et al., 2004). These identified issues in analyzing text led the developers of Coh-Metrix to develop a more sophisticated, hyper-detailed way to analyze text.

Another benefit of systems such as Coh-Metrix is the potential to address many needs in writing assessment. The ease and speed of use allow fast, detailed analysis of any type of written expression, including that produced by students, thanks to its easy to input web interface. However, the software generates a mass of data that is not in and of itself meaningful to instruction, including 50 types of cohesion relations and more than 200 types of language, text and readability variables (Graesser et al., 2004). The multitude of data generated by Coh-Metrix does not translate easily into data teachers can use to understand student proficiency in writing compared to state standards or student progress in specific writing skills found in most writing theory.

Writing Architect

Additionally, tools like the Writing Architect (Truckenmiller et al., 2020) have been developed, a software interface that includes the assessment and analysis of written language. In the software interface, students read a prompt and then respond to the prompt. The web interface captures the students' responses automatically at three, seven and 15 minutes. The software uses trained researchers and assistants to score written samples from students. Scorers are instructed to evaluate written language using correct minus incorrect writing sequences (CIWS), and essay quality using a measure was based on the Smarter Balanced Assessment Consortium writing rubric (2014) and revised by an expert writing panel. The software captures measures of paragraph typing fluency (characters typed correctly in 90 seconds), a task outside of the prompt response.

Research on Writing Architect has only recently been published as the software itself is newly developed, but the results of the research have been meaningful (Truckenmiller et al., 2020). A study published in 2020 indicated Writing Architect scores of writing fluency predicted 70% to 95% of the variance in writing achievement in middle school students and 31% of the variance in third grade students (Truckenmiller et al., 2020). This study is important in establishing Writing Architect as a tool for assessment and progress monitoring in schools, but the measure needs further validation and correlation to other established measures.

This study aims to use these two assessment tools, Coh-Metrix and Writing Architect, along with other established measures to further the understanding of the quantifiable relations between variables commonly used in writing assessment. Specifically, Writing Architect and Coh-Metrix variables will be used together to analyze a model of interaction with levels of language and the dimensions of performance of writing. Along with analyzing the variables from these tools, many other factors influencing the development of writing skills in individuals will be considered.

Factors Influencing Writing Development (Variables Controlled for in Current Study)

Why do students differ in their ability to write? Along with understanding the components of writing and how they interact to influence writing quality for a typical student, a model of writing needs to incorporate how individual difference influence these interactions. Gender, race, socioeconomic status, age, and special education eligibility are individual level variables that may influence the way the levels of language interact in predicting writing quality. Most importantly, identifying the malleable variables suitable for intervention is an important goal of this work. But first, individual differences within the student are also meaningful for establishing a model for evaluating writing, and for designing instruction and intervention.

Grade Level/Age

Previous sections have highlighted the importance of word- and sentence- level skills like vocabulary and syntax use in overall writing ability. The importance of these skills has been established through theory (e.g., Flower & Hayes, 1981, Scardamalia & Bereiter, 1987) and empirical research. Such empirical research has established that these skills do change over the course of a student's academic career (e.g., Olinghouse & Leaird, 2009, Troia et al., 2019 However, to design meaningful writing instruction with these skills in mind, a developmental trajectory aligned with grade levels must be established.

As stated previously, research in writing has found older students used more diverse vocabulary, with use of less frequent and longer words (Olinghouse & Leaird, 2009). In a study of narrative text writing in fourth, fifth and sixth grade students, word accuracy (measured using errors in capitalization and spelling) and word productivity (measured using the ratio of number of different words to total number of words) were significant predictors of narrative quality, differentiated good and poor writers, and had observable grade level changes between fourth, fifth and sixth grade students (Troia et al., 2019).

Although the development of many components of writing can follow a developmental trajectory, Troia and colleagues (2019) noted that younger students had fewer errors in capitalization and spelling, but more errors in punctuation. The authors interpreted these findings as younger students struggling with punctuation and choosing to use fewer challenging words in terms of spelling and capitalization. The authors found further confirmation for this interpretation with the results of shorter word length and less lexical complexity in younger students.

In terms of syntax, in the study investigating levels of language in written assessment by Troia and colleagues (2019), the authors identified sentence-level predictors of writing quality in narrative text for fourth, fifth and sixth grade students. In this study, some measures of syntax did not change in a significant way between grade levels, while others had linear changes, and others followed a non-linear trajectory. Such findings illustrate the need for a better developmental understanding of syntax across grade-levels.

Ability Level

Understanding how a writer's overall ability interacts with indices of word- and sentence-level language ability is also important for designing instruction. Available theories (e.g., Flower & Hayes, 1981) seem to suggest poor and good writers would differ in their word- and sentence-level language abilities. Indeed, research has illustrated that writers without learning disabilities (who scored better in measures of overall writing ability) had higher scores than writers with learning disabilities in productivity, sentence complexity and lexical diversity (Koutsoftas & Gray, 2012). In a similar study, children without learning disabilities scored better in cohesion (a measure of lexical continuity) than children with learning disabilities (Koutsoftas & Petersen, 2016). Although such studies are informative for understanding disability and intervention, a deeper understanding of ability level and its interaction with components of written expression is necessary. Considering disability adds dimensions to the analysis but understanding differences between good and poor writers without that dimension would also be informative.

Gender

Another important variable in understanding written expression and individual performance is gender. Overall, girls perform better in writing than boys (e.g., National Center for Education Statistics, 2012). However, the reasons behind this gender difference are yet to be

established. Some research has identified specific components of writing difference between the genders; for example, transcription differences in boys and girls, with girls outperforming boys in written orthographic fluency (Berninger & Fuller, 1992). In other studies, this gender gap was maintained even when language, reading, attention, spelling, handwriting automaticity and rapid automatized naming were controlled for (Kim et al., 2015). Some research suggests this difference in written ability between the genders is due to attitude (e.g., Knudson, 1995), whereas other research has found the relations between gender, attitude and written ability murky (Graham et al., 2007).

General findings on gender and written expression suggest a hypothesis for the current study of girls outperforming boys in measures of word- and sentence-levels of language and in overall measures of written expression. However, the reasons behind this gender gap are unclear. Identifying specific components of written expression, and where and how boys and girls differ, would be beneficial in designing instruction to close this gender gap in written expression.

Race/Ethnicity

Research on race and ethnicity and writing achievement and assessment is severely limited. Results from standardized writing assessments, such as those published through the National Assessment of Educational Progress (2011) indicate that in eighth grade, writing scores were higher for Asian students than for other racial/ethnic groups, followed by white students, students of two or more races, American Indian/Alaskan native students, native Hawaiian/other Pacific Islander students, Hispanic students and black students. Further, the same report found that in the 2007 assessment, the achievement gap between white and black eight grade students decreased, while the gap between white and Hispanic eight grade students did not significantly change (these were the only reported achievement gaps).

Using three creative writing samples, researchers analyzed for performance differences between Caucasian, African American, Latino/a, and Asian eight-grade students; they found no significant differences between the Caucasian and African American students, but did find significant differences only on the poetry task between Latino/a and Caucasian students and between Latino/a and Asian students.

Factors Influencing Writing Development (Variables Not Controlled For)

Cognitive Factors

Writing is an activity that involves the integration of multiple processes and skills. Writing also relies on multiple cognitive processes and the ability to write develops according to different trajectories. These processes interfere with each other during development, where advances or lack of development in one area may hinder abilities in another area (O'Rourke et al., 2018). The relationships are complex, vary by individuals, and are not fully understood.

According to Kim and colleagues (2018), writing requires higher-order cognitive processes which demand cognitive capacity. When lower-level skills such as transcription are efficient and automatic, they require less cognitive capacity, and there is more cognitive capacity available for higher level skills. Transcription skills are not considered attention demanding for skilled writers; for less skilled writers, transcription skills take cognitive capacity and less capacity is available for higher level skills. This theory of cognitive capacity for writing influences many models and predictions about how the skills involved in writing will interact with each other.

Swanson and Berninger (1995) investigated working memory and short-term memory specific to phonological skills in relation to measures of writing. Their results indicated working memory related significantly to writing measures, especially in the level of text generation. The

authors conclusions supported a capacity theory of writing with individual differences in children. Further, in an experimental study, Kellogg (2001) asked participants to use metacognition (cued by auditory reminders) to observe their writing process. The results of the study indicated different writing processes were in competition for working memory resources to produce writing.

Attention

In typical developing children, the development of attentional development has been studied extensively (Gupta & Kar, 2009), as the skill is an integral component of learning, influencing many academic areas as well as behavior. The ability to pay attention has a role in language, literacy and mathematics learning (Kruschke, 2003). Inattention in the classroom causes barriers to learning for children.

Studies of attention and writing ability often focus on children who have been diagnosed with attention-deficit/hyperactivity disorder (ADHD). Children with ADHD are more likely to have a learning disability (including written expression; Mayes and Calhoun, 2007) and often struggle with dysgraphia (Pitcher et al., 2003). Results from studies analyzing attention and writing abilities seem to indicate a probable correlation between the two abilities. In studies using structural equation modeling, models including components of attention were better fits than those just including components of language and literacy (Kent et al., 2014).

Instruction

The development of the ability to write is not fully understood but is thought to be dependent on both the individual and the context (Graham, 2006), as well as on external factors of instruction, curriculum and pedagogy (Schultz & Fetzo, 2000). The development of the ability to write depends on learning the fundamental skills associated with the writing process such as

semantics, grammar and spelling (Hayes, 1996). However, variability is widespread in writing instruction between teachers in terms of curriculum and instructional practices in both the processes and skills associated with written expression (Cutler & Graham, 2008). The writing practices employed in a classroom influences the development of writing skills of students, and the development of certain skills depends on explicit instruction (Graham, 2006). As such, curriculum, instruction and specific variations within instruction are independent variables that influence writing ability in individual students.

Purpose of Present Study

This study is based on models first developed by Flower and Hayes (1980) and refined by Berninger and colleagues (2003). The model of the simple view of writing (Berninger et al., 2003) posits of word-, sentence- and text-level variables interact in their prediction of overall writing quality. Troia and colleagues (2018) outline how components of written expression align with these levels of language to provide clear constructs for analysis. Previous studies have examined some variables in the model (e.g., Kim et al., 2018, Wilson et al., 2017), but the way these variables interact specifically for the typical and struggling writer has yet to be firmly established.

Truckenmiller and colleagues (2020) furthered the research literature surrounding this model using a newly developed assessment of writing, the Writing Architect (WA). This study uses the Writing Architect, which has promise for assessment and progress monitoring in the classroom. The Writing Architect has the ability to further assess the components of writing and their interaction in prediction of writing quality, both alone and in conjunction with analysis tools such as Coh-Metrix.

The outcomes of interest in this study are overall writing quality, with a focus complexity of word-, sentence- and text-level variables and a specific analysis of the role of spelling ability and its interaction with all levels of the model. Covariates included in the model include special education status, gender, grade level and race/ethnicity. Finally, in an additional analysis, a latent variable to dichotomize students into two groups, poor and good writers, will be constructed.

To analyze the main model, structural equation modeling and path analysis will be used. Path tracing will be used to examine the relationships between variables in the model and the fit of the model for predicting writing quality.

Table 2.

Levels of Language and Components of Written Expression

		Performance Dimensions of Written Language		
		Word	Sentence	Text
Levels of Language	Productivity			
	Accuracy	Spelling		
	Complexity	Word choice: meaningfulness, specificity, concreteness	Sentence structure: mean length, syntactical similarity, words before main clause	Text structure: (LSA) semantic overlap between paragraphs, given-newness of sentences, LSA cosines for adjacent sentences

Research Questions and Hypotheses

How does spelling constrain the other levels of language complexity in their prediction of writing quality?

Hypothesis. Spelling will constrain all levels of language in their prediction of writing quality. Spelling will constrain most at word- and text-levels of language complexity.

Rationale. Students who know word is spelled incorrectly often fail to move past trying to spell the word. Because of this, they do not engage in higher order cognitive processes necessary to produce written language (Beers & Nagy, 2009). Spelling has been found to account for unique variance in both word-level spelling and text-level composition in first through seventh (Abbot et al., 2010).

Question 1a:

How does spelling constrain the word level of language complexity (vocabulary) in its prediction of writing quality?

Hypothesis 1a. Greater spelling ability will be associated with greater word-level writing ability and higher writing quality.

Rationale. In studies of text production, children were found to be slower in producing text at the word and sentence level when they had lower spelling abilities (Sumner, Connelly, & Barnett, 2013). Students who know word is spelled incorrectly often fail to move past trying to spell the word. Because of this, they do not engage in higher order cognitive processes necessary to produce written language (Beers & Nagy, 2009). Previous studies have found spelling important at the word- and text- level of language (Abbot et al., 2010).

Question 1b:

How does spelling constrain the sentence level of language complexity (sentence construction) in its prediction of writing quality?

Hypothesis 1b. Greater spelling ability will be associated with greater sentence-level writing ability and higher writing quality.

Rationale. In studies of text production, children were found to be slower in producing text at the word and sentence level when they had lower spelling abilities (Sumner,

Connelly, & Barnett, 2013). Students who know word is spelled incorrectly often fail to move past trying to spell the word. Because of this, they do not engage in higher order cognitive processes necessary to produce written language (Beers & Nagy, 2009).

Question 1c:

How does spelling constrain the text level of language complexity (text complexity) in its prediction of writing quality?

Hypothesis 1c. Greater spelling ability will be associated with greater text-level writing ability and higher writing quality.

Rationale. Spelling ability constrains text production (speed) and the quality of text produced when children were asked to write, but not when they were asked to dictate a text (Sumner et al., 2014). Previous studies have found spelling important at the word- and text-level of language (Abbot et al., 2010). At the text level there are many more components involved in the prediction of quality (Berninger & Amtmann, 2002, Berninger et al., 2006).

Table 3.

Summary of Study Hypotheses and Analyses

Hypothesis	Analysis	Variables	Method
1. Spelling constrains most at word-, less at sentence- and the least at the text-level	Effect of spelling on the predictive ability of all levels of language on writing quality	Word-level writing abilities (independent) Sentence-level writing abilities (independent) Text-level writing abilities (independent) Spelling ability (independent) Writing quality (dependent) Age (control) Gender (control)	SEM/path analysis
1a. Greater spelling ability will be associated with greater word-level writing ability and higher writing quality	Effect of spelling ability on word-level writing abilities and writing quality	Word-level writing abilities (independent) Spelling ability (independent) Writing quality (dependent) Age (control) Gender (control)	SEM/path analysis
1b. Greater spelling ability will be associated with greater sentence-level writing ability and higher writing quality	Effect of spelling ability on sentence-level writing abilities and writing quality	Sentence-level writing abilities (independent) Spelling ability (independent) Writing quality (dependent) Age (control) Gender (control)	SEM/path analysis
1c. Greater spelling ability will be associated with greater text-level writing ability and higher writing quality	Effect of spelling ability on text-level writing abilities and writing quality	Text-level writing abilities (independent) Spelling ability (independent) Writing quality (dependent) Age (control) Gender (control)	SEM/path analysis

CHAPTER THREE: METHOD

Study Model

The model of the simple view of writing (Juel, 1988) includes the components of spelling and ideation. Future theories of writing included additional levels of ideation at the word-, sentence- and text-level variables interact in their prediction of overall writing quality (Berninger et al., 2003). Troia and colleagues (2018) outline how the productivity, accuracy and complexity dimensions written expression align with these levels of language to provide clear constructs for analysis.

Some initial studies have examined some variables in this model (e.g., Kim et al., 2018, Wilson et al., 2017), but the way these variables interact specifically for the typical and struggling writer has yet to be firmly established. The outcome of interest in this study is students' overall writing quality with informational text. Specifically, the role of spelling will be analyzed in relation to students' complexity of word-, sentence- and text-level variables (see Table 4, below).

Table 4.

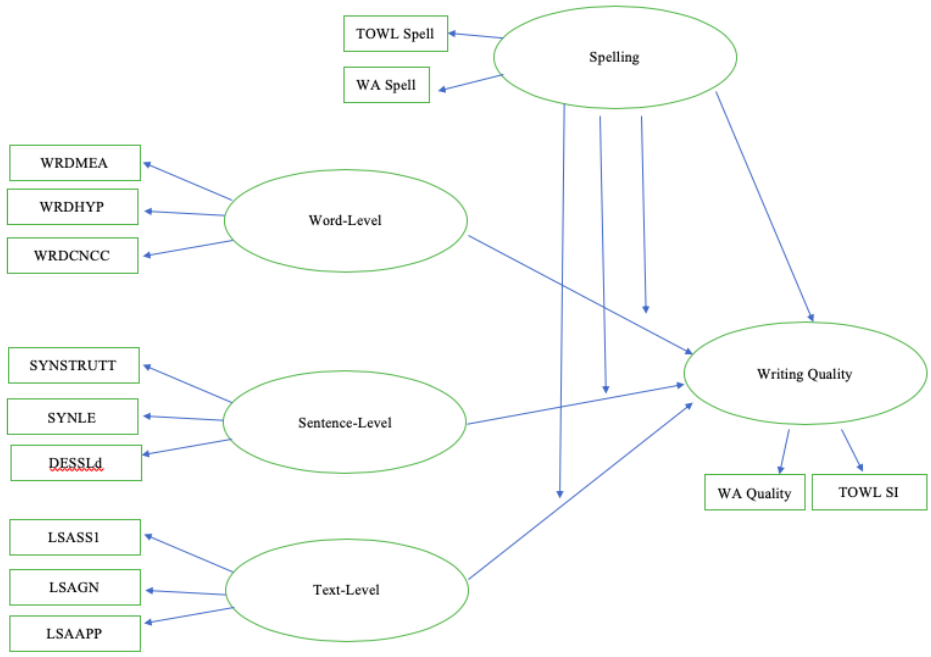
Levels of Language and Components of Written Expression

	Word	Sentence	Text
Productivity			
Accuracy	Spelling		
Complexity	Coh-metrix measures of word-level complexity (WRDMEA, WRDHYP, WRDCNCC)	Coh-metrix measures of sentence-level complexity (SYNSTRUCT, SYNLE, DESSLD)	Coh-metrix measures of text-level complexity (LSASS, LSAGN, LSAPP)

To analyze the model, structural equation modeling will be used to examine the relationships between variables in the model and the fit of the model for predicting writing quality. Figure 5. below is a visual representation of the theoretical study model showing expected interactions among latent variables and their observed measures. Covariates included in the model include gender, and grade.

Figure 5.

Hypothesized Model for Analysis



Study Design

This study used an existing data set composed of data from a study by Truckenmiller and colleagues (2020) that piloted Writing Architect 1.0, an online writing assessment tool. The study is a secondary data analysis of a measure validation study. There were 3 forms administered in a counterbalanced design. The sample was recruited as a convenience sample.

Data Collection Procedures

The WA tool was designed for use with grades 3 to 8, to assess written input produced in response to short informational text. The students read or click to listen to the informational text and are then prompted to compose a response through a computer interface (with a paper copy available). A time limit of three minutes for planning and fifteen minutes for composition is imposed. A sample of prompts and directions is included in *Appendix A*. The tool suppressed spelling and grammar checking by the internet browser, so students received no support or feedback in those areas during the task. After completion of the composing task, students complete a paragraph copying task to measure typing fluency.

For the original study in the winter of 2017, authors recruited seventeen general education teachers from five school districts, resulting in a total of 28 classrooms participating. These classrooms were in five rural and suburban school districts and included grades 3, 5, 6, 7 and 8 (no fourth-grade classrooms agreed to participate). Students were exempted from the study for lack of consent (n=2) or disability status (n=5). The sample resulted in a total of 526 students.

Data was collected using Writing Architect, an online writing assessment interface developed by Truckenmiller and colleagues (2020) in group administration in computer labs. Administration occurred during winter and spring of 2017. Writing Architect includes instructions in the interface, but these were also administered aloud by trained assistants. Before administration began, students received a copy of the passage, a blank page to plan and headphones. Administration time was within one class period including an allotted 3 minutes for planning, and 15 minutes for writing. Finally, writing samples were entered into an online scoring interface, Coh-Metrix (Graesser et al., 2003), and scored for three measures in each category of word-, sentence- and text-level complexity to represent corresponding levels of the

model. In addition, students were administered components of the Test of Written Language, Fourth Edition (TOWL-4) as part of the study.

The WA prompts were scored by trained assistants using a codebook (see *Appendix B*). Measures from the WA include hand-scoring of spelling and writing quality, as well as automated scoring using Coh-Metrix. Essay quality was a multi component measure made up of seven dimensions, each scored on a five-point rubric by trained assistants; the seven components were purpose, logical coherence, concluding sentence of section, cohesion, supporting details from source materials, language and vocabulary choice and grammar/usage/mechanics. Because each student was administered three different forms in a counterbalanced order, the mean of the three scores from each student was calculated and used in the analyses.

Research assistants were trained using a scoring and coding manual (see *Appendix B, Appendix C).* Training required the assistants to demonstrate higher than 95% agreement on three samples before they scored prompts (Truckenmiller et al., 2020). Interrater reliability was calculated for 28% of the samples using a two-way random model at ICC = .81 (Truckenmiller et al., 2020).

Sample

These students were recruited from rural and urban school districts in Florida and Michigan. The sample for this study included grades 6, 7 and 8 to allow for deeper analysis of variables at a specific group of grades, with a resulting sample size of 290 students. This sample included demographics described in Table 5.

Table 5.

Demographic Characteristics of Final Sample

Characteristic	Sample n
Child Gender	
Male	130
Female	160
Child Race/Ethnicity	
White/Non-Hispanic	225
African American	25
Hispanic or Latino	13
Native Hawaiian/Other Pacific Islander	13
More Than One Race	12
Grade	
6th Grade	123
7th Grade	59
8th Grade	108

Variables and Measures

Table 6 identifies the constructs in the present study as well as the proposed variables involved in constructing a latent variable of the construct and the data sources. Each level of language construct is hypothesized to be made up of three variables from Coh-Metrix. Both the writing quality and spelling constructs are hypothesized to be made up of a TOWL-4 and Writing Architect variable.

Table 6.

Construct, Variable and Data Source

Construct	Variables	Data Source
1. Spelling	Spelling standard score, spell correct	TOWL-4, WA
2. Word-level language	WRDMEA, WRDHYP, WRDCNCC	WA, Coh-Metrix
3. Sentence-level language	DESSLD, SYNSTRUTT, SYNLE	WA, Coh-Metrix
4. Text-level language	LSAPP, LSAGN, LSASS	WA, Coh-Metrix
5. Writing quality	WA rubrics, TOWL Spontaneous Index	WA, TOWL-4

Spelling

Spelling ability was assessed using the Test of Written Language, Fourth Edition (TOWL-4). The TOWL-4 Spelling subtest requires students to spell dictated sentences, scoring students on punctuation and spelling ability. The resulting score is scaled, with a range of 1 to 20 with 8-12 representing an average score. The grade-based coefficient alpha for the Spelling subtest is .91, .89 and .91 for 6^{th}, 7^{th} and 8^{th} grade. The TOWL-4 reports reliabilities ($\alpha > .80$) and validity coefficients with other assessments ($r = .54$).

The Writing Architect spelling variable was calculated by dividing word-level spelling errors by total words in the student's written response, reported as the percentage of words spelled correctly by the student.

Coh-Metrix Variables

The online computational tool Coh-Metrix analyzes text to produce 108 indices of linguistic and discourse representation (Graesser, McNamara, & Louwerse, 2003). These indices provide measures of written text and are grouped in categories according to what they are designed to measure. Categories include descriptive, text easability principal component scores, referential cohesion, LSA, lexical diversity, connectives, situation model, syntactic complexity, syntactic pattern density, word information, and readability. Although there are no performance benchmarks for Coh-Metrix to distinguish good from poor writers, quality of writing or response to instruction, studies have begun to identify specific Coh-Metrix indices that are associated with higher writing quality (e.g., Troia et al., 2019; Wilson et al., 2017).

Coh-metrix provides norms for the text base, separated by grade level and text genre (language arts, social studies and science). The norms were created by analyzing text created by the Touchstone Applied Science Associates which has a database of 119,627 paragraphs taken

from 37,651 writing samples (McNamara et al., 2014) from textbooks. As such, the norms are not relevant for analyzing student writing.

The Coh-Metrix variables used in this study to examine word-, sentence- and text- level complexity of language were based on work by Wilson and colleagues (2017). In their study, they attempted to establish a model of writing using these Coh-Metrix indices on middle school writing samples. They established a model for analyzing middle school writing samples using the 9 measures of complexity from Coh-Metrix described below.

Three measures of word-level complexity were gathered using the online system for computing coherence metrics, Coh-Metrix, to analyze the writing samples gathered through WA. The three measures of word-level complexity included in the data set were WRDMEA, WRDHYP, and WRDCNCC. The variable WRDMEA is vocabulary measure from Coh-Metrix, measuring meaningfulness ratings for words. Meaningfulness ratings for 2,627 words were developed by Toglia and Battig (1978); examples include the rating for 'people' (612) and 'abbess' (218'). Higher meaningfulness rated words are highly associated with other words; lower meaningfulness ratings may indicate more abstract words, perhaps used by more skilled writers.

The variable WRDHYP measures hypernymy (word specificity) where each word (noun or verb) is measured on hierarchical scale for specificity. Using WordNet (Fellbaum, 1998), the hierarchical scale, where a lower score indicates less-specific words and a higher value indicates more-specific words; more specific words may be expected in a text from a more skilled writer. Finally, the variable WRDCNCC100 is an index of how concrete or non-abstract a word is; words that are more concrete are those things you can hear, taste or touch. Text that contains

abstract words is more difficult for readers to comprehend (McNamara & Graesser, 2011); hypothetically, more abstract words would be expected from more skilled writers.

Three measures of sentence-level complexity were gathered using Coh-Metrix (DESSLD, SYNSTRUTT, SYNLE). The variable DESSLD is a sentence construction measure from Coh-Metrix, this is the standard deviation of the measure for the mean length of sentences within the text; in this variable, a large standard deviation indicates that the text has large variation in terms of lengths of sentences. In general, longer sentences are expected to be more syntactically complex and would hypothetically indicate better writing quality (this is an established indicator of difficult for reading; McNamara & Graesser, 2011).

The variable SYNSTRUTT is a measure of sentence-to-sentence syntax similarity, specifically the proportion of intersection tree nodes between all sentences and across paragraphs. This variable is measured by calculating the consistency of syntactic construction using a tree displaying nodes. The more uniform the syntactic components of a text are, the easier the readability of the text (Crossley, Greenfield, & McNamara, 2008); hypothetically, a higher score here would be expected for better writers.

Finally, for sentence-level complexity, the variable SYNLE is the mean number of words before the main verb of the main clause in sentences, this is a good index of working memory load. This is also a measure of sentence level difficulty, with more difficult sentences having more words before the main verb of the main clause. A study by McNamara, Crossley, and McCarthy (2010) identified this index as one of the three most predictive indices of essay quality.

As with the other levels of language, three measures of text-level complexity were gathered using Coh-Metrix analysis of the writing samples. The measures of text-level language

complexity involve latent semantic analysis (LSA), which is a measure of semantic overlap (Landauer et al., 2007). LSA analyzes meaning overlap between words (explicit words and words related in meaning) by using singular value decomposition. This statistical technique computes the similarity between parts of the text (word, sentence, text-level) and reports a geometric cosine (McNamara & Graesser, 2011). Semantic overlap is an important way of measuring the cohesion of text, which has implications for readability; more cohesive texts are easier for readers to understand (McNamara & Graesser, 2011); hypothetically more cohesive texts would be expected from more skilled writers. When there is semantic overlap between words, sentences and paragraphs in the text, the content of the writing is linked between these areas. Coh-Metrix measures the cohesion between different parts of text, from sentences next to each other, or adjacent sentences (a more localized measure) to paragraphs (a more global measure).

These measures, LSAPP, LSAGN, LSASS, were used to construct a latent variable of text-level complexity for writing. The variable LSAPP from Coh-Metrix is an index of latent semantic analysis (LSA) that is computed as the mean of the LSA cosines between adjacent paragraphs. The variable LSAGN is the average given-new of each sentence. This measure, unique to Coh-Metrix, classifies text into three categories: given, partially given or not given based on how much given versus new information is in each sentence compared with the content of prior text. These variables have a range of 0 to 1, with 0 indicating less given information (lower cohesion of text) and 1 indicating less new information (higher cohesion of text). Finally, LSASS is computed as the mean of LSA cosines for adjacent sentence-to-sentence units, and measures how consistent adjacent sentences are overlapped semantically.

Writing Achievement

The writing achievement latent variable was first hypothesized to be made up of the TOWL-4 Story Composition score and the writing quality score from Writing Architect (minus the mechanical score to not confound spelling in the analysis). These measures combined a standardized measure of writing quality with a rubric developed in research. It was important to capture an accurate representation of the student's written ability in this score, hence the multiple measures.

The TOWL-4 Spontaneous Index estimate's the test-taker's writing ability using subtests that evaluate spontaneously composed essays. Two subtests are used to calculate this composite score, Contextual Conventions and Story Composition. The Contextual Conventions subtest requires the student to write a story in response to a stimulus, earning points for satisfying orthographic and grammatic conventions. The grade-based coefficient alpha for the Contextual Conventions Subtest is 74, .82 and .80 for 6th, 7th and 8th grade. Composition subtest requires students to write in response to a prompt. The written response is scored on organization, theme, plot, character development, prose and vocabulary use. Specifically, the scorer is asked to respond to 11 items in rating the writing. The grade-based coefficient alpha for the Story Composition subtest is .75, .72 and .65 for 6th, 7th and 8th grade. The grade-based coefficient alpha for the Spontaneous Index is .84 for all three grade levels. The TOWL-4 reports reliabilities ($\alpha > .80$) and validity coefficients with other assessments ($r = .54$).

The Writing Architect essay quality score was calculated using rubrics in the WA Codebook (see *Appendix B*) by trained research assistants. The final written response at 15 minutes was scored using a rubric developed by researchers (Troia, 2018). This rubric measured the written response on (a) purpose, (b) logical coherence, (c) concluding sentence or section, (d)

cohesion, (e) supporting details from source materials, (f) language and vocabulary choice, and (g) grammar/usage/mechanics. Each of the seven dimensions was scored on a scale of 0 to 5 with 0 indicating 'no evidence of dimensional quality; severely flawed/incomprehensible' and 5 indicating 'excellent evidence of dimensional quality, virtually no flaws/fully comprehensible' (Truckenmiller et al., 2020). This measure had reliability of r = .70. The variable used in this study included all scoring dimensions except mechanical (which would include spelling).

Statistical Analysis

The model was adapted from Wilson and colleague's model of assessing text complexity at the word-, sentence and text-level (2017) combined with the model used by Troia and colleagues in 2019. The intent was to use a similar structure with the three-indicator latent variables for each level of language, combined specifically with an analysis of the interaction and effects of spelling, a transcription measure.

Analysis began with three-indicator latent variables for each of the levels of language: word, sentence and text. Many different iterations of this model were attempted, but the fit was never appropriate. Issues with the latent variables resulted in the model fit being poor for the sample. Specifically, the sentence and word variables caused issues.

Eventually, different combinations of the indicators for sentence and word variables were tested in the model. Every combination of two-indicator latent variable and one indicator were tested for the word, sentence and text levels. Finally, the model fit with one indicator for the sentence and word variables, and a two-indicator latent variable for the text variable.

The measurement model with the best fit included then two-indicator latent variables for text, spelling and writing quality, and a one-indicator variable for word and sentence levels. This

measurement part of the model had very high factor loadings and a very good model fit. These methods and analyses are reported in detail below.

Preliminary/Descriptive Analysis

Data analysis was completed using Mplus (Muthén & Muthén, 2017) and SPSS software. Before testing the model fit, preliminary descriptive analyses and tests of normality were conducted. These tests and statistics were carried out to better understand the data set and constructs included. Identified outliers were included in the study as the study aims to understand how the model fits for all children, including gifted and those with special needs. Means and standard deviations of all continuous variables were calculated as well as frequencies and percentages of categorical variables. Tests of normality indicated normal distributions for most variables, although some transformations were performed because of the differences in variable range and extremely small scales and ranges for some variables. For example, the original SYNLE variable had a range of 0 to .391, mean of 0.172, and standard deviation of 0.070 whereas WRDCNC had a range of 3.8278 to 4.9522, a mean of 4.9522, and standard deviation of 0.159.

Table 7.

Original Means and SDs of Sample

Variable	Mean	SD
SpellSS	10.93	2.756
WRDMEA	6.407	0.684
WRDHYP	7.805	2.610
WRDCNCC	425.898	15.910
DESSLD	7.330	3.429
SYNSTRUTT	247.029	38.909
SYNLE	0.618	0.130
LSAPP	0.271	0.089
LSAGN	0.758	0.090
LSASS	0.253	0.121
Spontaneous Index	104.56	18.769
Essay Quality	12.474	5.358

Data Transformations

Coh-metrix variables were initially reported in different ways (scales, count, proportions). These variables were transformed to more z-scores because of vast differences in range and standard deviation of the variables. The z-score variables were used for analysis.

Measurement Model

Latent Variables

To test if the hypothesized measurement model fit the data, full information robust maximum likelihood model was used. Mixed modeling was used to allow for random slopes and intercepts, which allows for the creation of interaction terms.

The latent factor structure was tested using MPlus (Muthén & Muthén, 2017). Each latent variable is described above along with the indicators included. Ultimately, every possible combination of the latent variables was tested in the model. For the attempted latent variable for sentence complexity, the final model used just one measure. The same was true for the word level complexity variable. Again, balances between the statistical fit and theoretical value were

considered when making these decisions. Ultimately, the model fit was best with these decisions. In the end, writing quality, spelling and text were latent variables, and the sentence and word constructs were not. Every variation of the model was attempted before selecting the final model with the best fit. Table 8, below displays the measures of fit for the full model (with all latent variables), then for each model as variables were removed.

Table 8.

Latent Factor Analysis, Fit of Different Models

	Chi Sq	**RMSEA (90 Percent C.I)**	**CFI/TLI**	**SRMR**
Full Model	610.917, p=0.000	0.163 [0.151, 0.175]	0.695, 0.661	0.175
Remove Word 1	502.383, p=0.000	0.163 [0.150, 0.176]	0.737, 0.700	0.142
Remove Word 1 & 3	302.964, p=0.000	0.151 [0.135, 0.167]	0.818, 0.795	0.192
Removed Sent 1, Word 1 & 3	312.225, p=0.000	0.177 [0.159, 0.195]	0.795, 0.762	0.148
Remove Sent 3, Sent 1, Word 1 & 3	139.195, p=0.000	0.152 [0.129, 0.176]	0.898, 0.881	0.094
Remove Text 2, Sent 1 & 3, Word 1 & 3	28.828, p<0.0042	0.0700 [0.037,0.102]	0.981, 0.977	0.041

Many analyses of fit were used for the data including chi-square statistic, comparative fit index (CFI; Bentler, 1990), Tucker-Lewis index (TLI; Tucker and Lewis, 1973), root mean square of error approximation (RMSEA; Steiger, 1990), Chi square and standardized root mean square residual (SRMR). RMSEA is an index of absolute fit, comparing the measurement model to a perfect model, and the left confidence interval should be less than .05 to indicate good fit (Browne and Cudeck, 1993. CFI and TLI are incremental fit indices where the measurement model is compared with a baseline model, and the index should be greater than .90 to indicate

good fit (Bentler and Bonnet, 1990). Chi square was also used to analyze the fit of the measurement model and should be significant to indicate good model fit. The standardized root mean square residual is another fit index that is calculated as the square root of the difference between residuals in a covariance matrix and hypothesized model; this value should be below 0.08 to indicate a good fit.

In the final measurement model, all factor loadings where high and the model fit was very good. Model fit for measurement model: $\chi 2 = 28.828$, df= 12, $p < 0.0042$; RMSEA =0.070 [0.037, 0.102]; CFI = 0.981; TLI= 0.977; SRMR =0.041.

Parameter Estimates of Measurement Model

After establishing the latent variables, the standardized parameter estimates of the model were calculated. In the measurement model, all factor loadings are significant ($p<0.05$), indicating good fit within the model (see Table 9, below). The magnitude of all factor loadings, a measure of the variance explained by each variable for the factor, are all above 0.5, which according to Hair and colleagues (1990) is practically significant. Correlations between the different latent variables indicate that the factors are related, which is to be expected for variables measuring components of the same construct.

Table 9.

Standardized Parameter Estimates of the Three Levels of Language Model

	Estimate	SE	Estimate/SE	p-value
Writing Quality: Factor Loadings				
Qual Total	0.813	0.032	25.088	0.000
Spontaneous Index	0.660	0.404	16.513	0.000
Spelling: Factor Loadings				
Spell Correct	0.694	0.043	16.106	0.000
Spelling TOWL	0.902	0.040	22.337	0.000
Text: Factor Loadings				
LSAPP	0.926	0.018	50.658	0.000
LSASS	-0.894	0.020	-45.617	0.000
Correlations				
Spell with Writing Quality	0.770	0.055	13.995	0.000
Text with Writing Quality	0.889	0.033	26.638	0.000
Text with Spelling	0.470	0.059	7.990	0.000
Residual Variances				
WQ1	0.339	0.053	6.442	0.000
WQ2	0.565	0.053	10.724	0.000
Spel 1	0.518	0.060	8.663	0.000
Spel 2	0.186	0.073	2.546	0.011
Text1	0.143	0.034	4.212	0.000
Text 2	0.200	0.035	5.706	0.000

CHAPTER FOUR: RESULTS

After the measurement model was established, the model was tested for main effects and interactions of variables. Each interaction was added, one at a time, to test for the effect on the model. The interaction of spelling with each level of language (word, sentence and text) was analyzed. The model was then compared using the chi-square test of loglikelihood in addition to the Akaike information criterion (AIC) and Bayesian information criterion (BIC) results.

Main Effects Model

First, a model testing only the main effects of the variables on writing quality was analyzed. Analysis of the main effects model indicated some variables were significant predictors of Writing Quality for the 6-8th grade students. The model indicated WQ was significantly predicted by Spelling (β=0.431, p =0.000), Text (β=0.676, p=0.000), and Sentence 2 (β=-0.114, p=0.004).

Interactions

Loglikelihood and information/comparative fit criteria were examined to compare model fit of different models. Changes in the model were balanced between helping the model statistically and preserving the theoretical substance of the model. MPlus first models the outcome on all main effects, then interactions are created. To assess interactions in the model, each interaction and combinations of interactions were added to the model and first analyzed using AIC and BIC fit criteria (see Table 10, below). When the interactions were added to the main effects model, the AIC and BIC were compared; in general, the lowest number for each of the fit criteria signifies the best model fit. Further, examining how the numbers changed with the addition of the interaction can signify better fit (lower number in AIC/BIC) or worse fit (higher

number in AIC/BIC). The lowest AIC and BIC are bolded in the table below, signifying the interaction or interaction combination with the best fit.

Table 10.

Analysis of Interactions

	AIC	**BIC**	**SS Adj BIC**
Main Effects Only	3868.907	3942.236	3878.813
Add Sent x Spell	3870.555	3947.550	3880.956
Add Word x Spell	3870.781	3947.776	3881.182
Add Text x Spell Interaction	**3857.897**	**3934.892**	**3868.298**
Add Sent x Spell, and Text x Spell	3859.869	3940.530	3870.765
Add Word x Spell and Text x Spell	3872.477	3953.138	3883.373

The interactions in the model were further assessed using a chi square test of loglikelihood which can be used to test between nested models. In this analysis, each interaction was added to the main model and analyzed using this test, with a significant result indicating better fit when the interaction is added to the model. In this analysis, a scaling correction was used so the calculated differences would be chi-square distributed (Satora & Bentler, 2010). The following formula was used for the scaling correction:

1. Compute the difference test scaling correction where p0 is the number of parameters in the nested model and p1 is the number of parameters in the comparison model.
2. cd = (p0 * c0 - p1*c1)/(p0 - p1)
3. = (39*1.450 - 47*1.546)/(39 - 47) = 2.014
4. Compute the chi-square difference test (TRd) as follows:
5. TRd = -2*(L0 - L1)/cd

   = -2*(-2606 + 2583)/2.014 = 22.840

When analyzed, the test indicated that only the interaction of text by spelling should be included in the model (β=0.143, p=0.000). The sentence and spelling interaction, and the word and spelling interactions were not included in the final model. These interactions were added to the model, and the ratio test was conducted and found the interactions did not contribute significantly to the model (see Table 11, below). These results indicate spelling does not constrain the sentence level of language or word level of language in their prediction of writing quality.

Table 11.

Loglikelihood Analysis of Interactions

Model	**Log likelihood (Scaling Correction Factor)**	**Number of free parameters**	**Difference test**	**P-value**
Main Effects Only	-1914.454 (1.1085)	20	-	-
Add Sent x Spell	-1914.278 (1.1134)	21	0.2906	0.5899
Add Word x Spell	-1914.391 (1.0671)	21	0.5270	0.4679
Add Text x Spell	**-1907.949 (1.0999)**	**21**	**14.0209**	**>0.001**

Final Structural Equation Model

Spelling was a significant predictor of writing quality, in that better spelling indicated better writing quality (β=0.431, p =0.000). The same was true for text (β=0.676, p=0.000). For the sentence variable, a higher score indicated worse writing quality in a significant way (β=-0.114, p=0.004). The word variable did not significantly predict Writing Quality in the model (β=0.064, p=0.060). The significant interaction between Spelling and Text suggests that the effect of text on writing quality is even higher when spelling is also high (spelling moderates the

effect of text on writing quality). Grade, gender and ethnicity variables were included as important control variables in the model and were all significant predictors of writing quality (see Table 12, below).

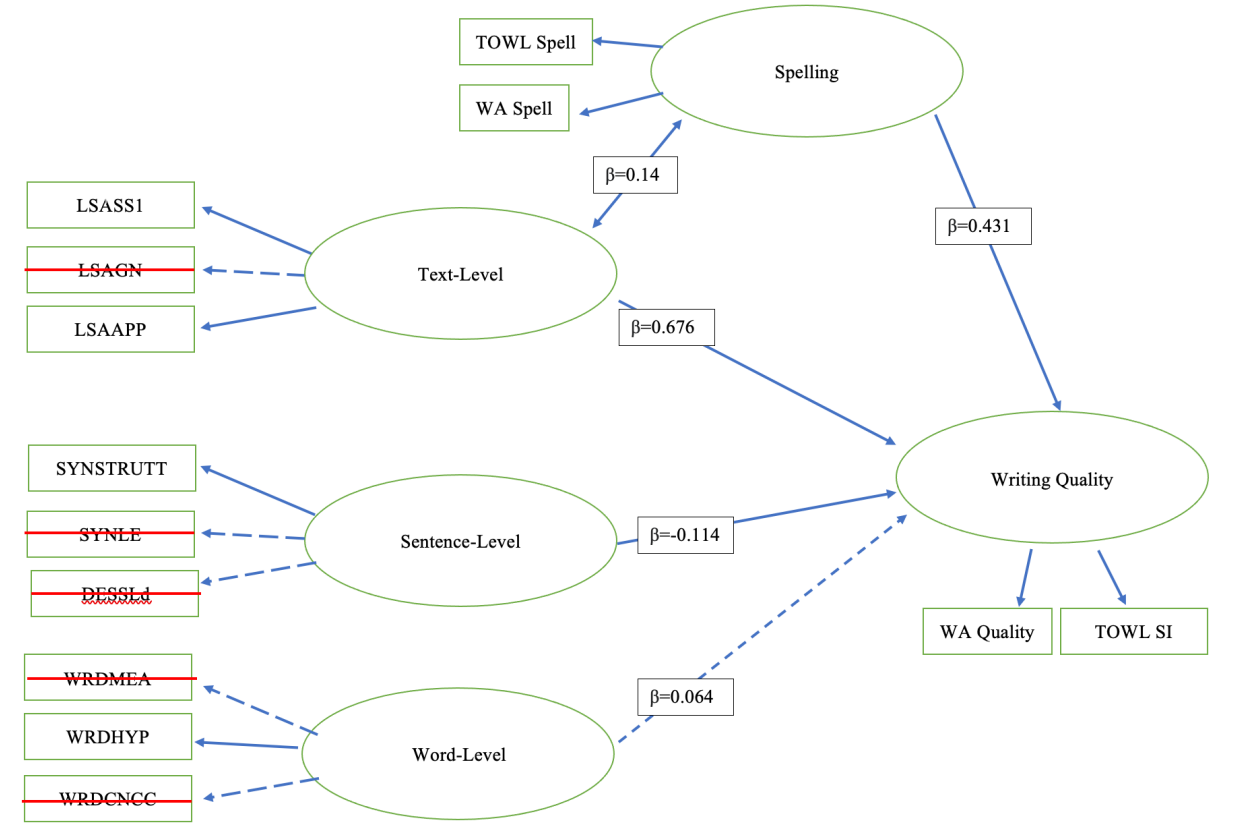
Table 12.

Structural Equation Model with Covariates

	Estimate	SE	Estimate/SE	p-value
Writing Quality: Factor Loadings				
Quality Total	0.844	0.030	27.736	0.000
Spontaneous Index	0.677	0.041	16.402	0.000
Spelling: Factor Loadings				
Spell Correct	0.725	0.046	15.873	0.000
Spelling TOWL	0.863	0.040	21.771	0.000
Text: Factor Loadings				
LSAPP	0.940	0.018	50.967	0.000
LSASS	-0.882	0.021	-42.638	0.000
Regressions				
Writing Quality on Spell	0.431	0.056	7.636	0.000
Writing Quality on Text	0.676	0.053	12.830	0.000
Writing Quality on Text x Spell	0.143	0.031	4.594	0.000
Writing Quality on Sent 2	-0.114	0.039	-2.891	0.004
Writing Quality on Word 2	0.064	0.034	1.880	0.060
Writing Quality on Grade	-0.063	0.012	-5.380	0.000
Writing Quality on Gender	0.157	0.042	3.770	0.000
Writing Quality on Ethnicity	0.118	0.036	3.243	0.001
Covariances				
Text With Spell	0.464	0.068	6.772	0.000
Residual Variances				
Writing Quality 1	0.288	0.051	5.613	0.000
Writing Quality 2	0.542	0.056	9.713	0.000
Spell 1	0.526	0.066	7.936	0.000
Spell 2	0.744	0.068	10.886	0.000
Text1	0.883	0.035	25.484	0.000
Text 2	0.777	0.036	21.319	0.000

Figure 6.

Final Measurement Model



CHAPTER FIVE: DISCUSSION

In the sample included in this study, spelling significantly predicted writing quality, in that better spelling indicated better writing quality. The same was true for text measures of complexity, where the interaction was also significant. Finally, a significant interaction was found wherein if the spelling scores and text scores were better, the results for writing quality were even greater. Below, research questions are reviewed with their hypotheses, the relevant results and discussion related to each question.

Research Questions

Spelling Moderates Text-Level Complexity in Predicting Writing Quality

Research question one examined if spelling constrained the other levels of language complexity in their prediction of writing quality. The hypothesis for this research question was that spelling would constrain all levels of language in their prediction of writing quality but would do so most at word- and text-levels of language complexity. This hypothesis was based on previous research indicating that performing higher order cognitive processes are more difficult when students have limited performance in lower order processes (Beers & Nagy, 2009). Additionally, spelling has been found to account for unique variance in both text-level composition and future development of word-level spelling in first through seventh grades (Abbot et al., 2010).

As anticipated, text level cohesion variables best predicted writing quality and spelling significantly constrained or facilitated a student's ability to effectively express text level cohesion. Spelling moderated the text level of language in the prediction of writing quality, but not the word and sentence levels of language. There are various reasons that spelling moderated only the text level of language in its prediction of writing quality. Previous studies have found

some differences in the way levels of language behave in models of written expression, explored below with consideration of current findings.

First, according to the capacity theory of writing put forth first by McCutchen (1996), the skills involved in producing these different levels of language may be competing for cognitive resources. This competition influences writing quality. Perhaps, in this sample of older students, the lower levels of language are developed and using fewer resources, but the higher level of language (text) is still challenging for middle school students and is further constrained by spelling ability in its production of quality text. Hayes and Berninger (2014) suggest that text generation, including the components of proposing, transcription, translation and evaluation constrains writing ability/quality until 5th grade. In the intermediate grades the processes mature and become automated and so require fewer resources. In this sample, the text processes may not be mature yet and the students still require cognitive resources.

These results also fit with findings of Summner, Connelly and Barnett (2013) that children were found to be slower in producing text at the word and sentence level when they had lower spelling abilities. Additionally, previous research on this subject found students failed to engage in higher order cognitive processes when stuck on a lower order process such as spelling (Beers & Nagy, 2009). Finally, previous studies have found spelling important at the word- and text- level of language (Abbot et al., 2010).

Additionally, Abbott, Berninger and Fayol (2010) suggested an explanation related to working memory; word-level working memory is related to all writing skills and spelling. The authors went on to note that the relation between text-level language and spelling indicated these skills develop concurrently, especially during middle school. These results support the research by Abbot and colleagues (2010) in finding a significant relationship between text-level language

and spelling, and further the research by finding the two interact in their prediction of writing quality for middle school students.

It may also be due to word and sentence level complexity predicting only a small amount of unique variance in writing quality, which is similar to findings from others (Troia et al., 2019; Truckenmiller & Petscher, 2020). Research by Troia et al. (2019) also only found text level complexity (incidences of connectives and narrativity) a significant prediction of narrative quality. The same study found no word or sentence levels of complexity were significant predictors of writing quality in grades 4 through 6. The current study supports this finding in suggesting that text is a predictor of narrative quality and goes further by analyzing how spelling fits into this relationship. The finding that spelling has a significant interaction with text in predicting writing quality adds to the understanding of how a model of written language may work in students.

In terms of syntax, in the study investigating levels of language in written assessment by Troia and colleagues (2019), the authors identified sentence-level predictors of writing quality in narrative text for fourth, fifth and sixth grade students. In this study, some measures of syntax did not change in a significant way between grade levels, while others had linear changes, and still others followed a non-linear trajectory. Such findings illustrate the need for a better developmental understanding of syntax across grade-levels.

Troia et al. (2019) suggested that sentence-level variables, which were found to be stable in that study, may have been in a "quiet period of development" in their 5-8th grade sample. This could be true in this sample of slightly older students, and also be true for the word-level variables included in this analysis. Perhaps these skills have reached a plateau in development, where only the text-level variable is predicting variability in writing quality.

Finally, differences in samples, differences in methods and differences in analysis could explain these findings. In terms of samples, this sample used middle school students, which is a somewhat novel sample for analyzing this model. Understanding how the model may work differently across different age levels is important for future research and practice, and these results add to the existing data from mostly elementary students. That the model worked differently in middle school students may be an indication of differences in how language works as skills develop.

However, the sample was very similar in age to that used by Wilson and colleagues in 2017 to test a very similar model. This sample differed demographically and did not include students in special education and did include 7th grade students (in addition to 6th and 8th in the Wilson et al., 2017 sample). It is not possible to say why exactly the results differed from previous research, but it is important to continue to test this model in different samples to establish individual and developmental differences in how it operates.

In terms of methods, this analysis was based on writing samples that were informational in nature. Informational text is important for many academic areas but may differ in language structure from narrative or persuasive text. The most recent study to test a similar model by Wilson and colleagues (2017) used an argumentative writing prompt. Understanding how the components of written language differ in different types of writing has implications for instruction and assessment.

The analysis used here was similar to previous analyses of models of language (e.g., Wilson et al., 2017) but also differed in some ways. For example, Wilson and colleagues (2017) made some adjustments to the text and sentence latent variables in the model (e.g., Heywood cases and shared variances). Wilson and colleagues (2017) also found differences in the different

grade levels when they were modeled alone, likely indicating different relations between syntax and cohesion across grades. Finally, Wilson and colleague's 2017 results on cohesion (text level) were different from previous research on high school students that did not find cohesion to be a predictor of quality. As stated previously, this research adds to the existing evidence on how this model may differ individually and developmentally and across different samples and increases the need for further exploration of the model's functioning.

Spelling Did Not Influence Word-Level Complexity

The results of analysis of the current sample indicated spelling did not moderate word-level writing ability in its prediction of writing quality. When the spelling and word-level interaction was added to the model, the loglikelihood test indicated the interaction should not be included in the model. Rather, spelling had a rather large and direct relation to writing quality above and beyond its interaction with student's vocabulary word choice. The implication is that spelling has an influence on writing quality throughout middle school and that influence is independent of its relation with word choice.

Similar to Troia's 2019 finding that complexity of words did not significantly predict narrative writing quality, we found that word-level complexity, measured as hypernymy in this study, did not significantly predict informational writing quality. Troia found that complexity variables did differentiate between higher and lower performing students. This suggests that word complexity may be a target for instruction, depending on how much a writing quality outcome values use of complex words.

Beyond these theoretical explanations for the findings regarding word-level complexity, the findings could also be due to an error in the methods or measurement. For example, perhaps the Coh-Metrix measures used did not measure word level language abilities. Also, in the current

analysis, the word-level variable was not a latent variable, and less information is stored in a single observed variable than in a latent variable by nature. Future research should examine the reliability and validity, and especially content validity of the word choice/ word complexity/vocabulary construct for use in models.

Spelling Does Not Constrain Sentence-Level Complexity

Question 1b examined how spelling constrained the sentence level of language complexity in its prediction of writing quality. The results of analysis of the current sample indicated spelling did not moderate sentence-level writing ability in its prediction of writing quality. When the spelling and sentence-level interaction was added to the model, the loglikelihood test indicated the interaction should not be included in the model. Rather, spelling had a rather large and direct relation to writing quality above and beyond its interaction with student's sentence complexity. The implication is that spelling has an influence on writing quality throughout middle school and that influence is independent of its relationship with measures of sentence syntax in student writing.

This finding is inconsistent with a similar model analysis by Wilson and colleagues (2017) and analysis in other samples (e.g., Beers & Nagy, 2009, Sumner, Connelly, & Barnett, 2013), but is consistent with other research (e.g., Troia et al., 2019). There are many possibilities as to why the model worked in this way for this sample of students. In terms of syntax, in the study investigating levels of language in written assessment by Troia and colleagues (2019), the authors identified sentence-level predictors of writing quality in narrative text for fourth, fifth and sixth grade students. In this study, some measures of syntax did not change in a significant way between grade levels, while others had linear changes, and others followed a non-linear

trajectory. Such findings illustrate the need for a better developmental understanding of syntax across grade-levels.

As with the word level of language, methods and sample variations should be considered. The Coh-Metrix variables again may not measure what was intended in the study, as they are relatively new automated variables used in writing research. As with the word-level of language, students may have developed skills by the middle school age to compensate for their lower spelling abilities, and we may not see the constraint on this level in prediction of written ability. Finally, also in line with the word-level findings, the sentence-level variable was not a latent variable, and less information is stored in a single, observed variable than in a latent variable by nature.

Spelling and Text Have Significant Interaction

Question 1c examined the relationship between spelling and the text level of language complexity (text complexity) in its prediction of writing quality. The hypothesis for this question was that greater spelling ability would be associated with greater text-level writing ability and higher writing quality. This hypothesis was based on previous research in which spelling ability constrained constrains text production (speed) and the quality of text produced when children were asked to write, but not when they were asked to dictate a text (Sumner et al., 2014) and research that indicated at the text level of language, there are many more components involved in the prediction of quality.

Results of the current analyses found spelling and text to have a significant interaction, indicating spelling moderates the relationship between text level language in this sample. This finding is consistent with previous research that found spelling ability to constrain text production quantity and quality (Sumner et al., 2014). Also, in accordance with previous

research, cognitive processing/capacity theories of language may posit that at the text level there are so many lower order cognitive processes involved that the student may not produce text/produce quality text when focusing on these lower order processes (Berninger & Amtmann, 2002, Berninger et al., 2006).

Limitations

As outlined in the capacity theory in relation to the simple view of writing, there are many variables that influence written expression. Research is not conducted in a perfect world and measuring and controlling for every variable is not plausible. As such, there are limitations to the present study. There are many additional variables that could have been collected and analyzed that may have influenced the model and results. Some skills related to the ability to write in direct and indirect ways would inform the development of a complete, exhaustive model of written expression, such as memory, motor skills, self-regulation, and pre-spelling literacy skills, to name a few. In an ideal situation, a full cognitive assessment of participants would inform a model of written expression. In Kim and Schatschneider's 2017 model, some components of cognitive processing and development such as theory of mind, inferencing ability and working memory were included. Verbal IQ and overall IQ could also influence writing quality and the model. These were not included in the current study.

Variables outside of the child such as explicit spelling instruction, literacy instruction, and curriculum would also influence spelling and writing ability. None of these were included in the study model. Other external to child variables such as educational policy at local, state and national levels could certainly be hypothesized to influence education, but such an analysis would add multiple levels of complexity.

In terms of methods, there are also limitations to the current study. The outcome measure used in this study was a latent variable using the Writing Architect essay quality score and the TOWL-4 Spontaneous Index, which is measure of narrative essay writing ability. Although this latent variable improves upon models with only one measure, the use of a narrative task from the TOWL-4 may make this a different construct than informational writing quality. Additional research is needed to improve the availability and understanding of writing assessment (IES, 2017). The methods for this study focused on informational writing prompts, which may not generalize to other types of writing. Finally, using a first-draft only assessment differs from authentic writing situations but it is the most common way automated scoring has been analyzed (Wilson et al., 2017).

The use of Coh-Metrix variables, while promising, may also be problematic. These variables are just beginning to be understood in their measurement of student's written expression and how they may fit into developmental models of written expression. They are not established in terms of reliability or validity for use in analyzing student writing samples. In many cases, it is still unclear if they in fact measure what the purport to measure, or even what they purport to measure. Although these variables have vast potential for automated writing assessment, further research is needed.

The sample and analyses included in this study may also be a limitation in terms of establishing a model of written expression. The development of skills necessary for written expression may not follow a linear trajectory and analyzing multiple grade levels together may not be the best way to understand a model of written expression. In fact, previous research has identified that certain components of written language do not seem to follow a linear developmental trajectory (Troia et al., 2019) and attempting to use such a trajectory may not lead

to a clear model of written expression. Wilson and colleagues (2019) also found that correlation directionality was different when each grade level was analyzed separately, indicating different relations between variables at different grade levels. Future research should analyze grade levels separately as well as together to understand how the model functions differently in these different samples and in terms of development.

Implications for Practice

This research has many implications for practice. A model of written expression, developed with individual and developmental considerations and established through research can guide assessment, intervention and instruction of writing in schools. This model and its implications for school practice would also influence the training of teachers, development of writing assessments and development of curriculum.

In the general education classroom, an accurate and efficient writing assessment would allow for the identification of which students needed targeted instruction in writing, as well as what skills to target in instruction. Further, such an assessment would allow for progress monitoring of the different skills involved in written expression for students included in interventions. These students also would benefit from feedback on their writing targeted to their needs and specialized instruction.

This study adds to existing research attempting to establish a model, and uniquely, begins to analyze how another literacy skill, spelling, should fit into considerations of intervention and assessment of written expression. Results in this sample strengthen the existing literature suggesting that spelling plays a major role in writing quality and can constrain the text level of complexity. These results should guide practitioners to consider explicit spelling instruction as a

possible way to improve writing quality for students and that its likely to be at least as important, if not more important than diversity of vocabulary and sentences.

This study moves toward establishing how automated variables may fit into a model of writing and assess student growth and proficiency. The development of assessments that could be semi-automated in nature has many implications for the way teachers assess and respond to assessments with instruction and intervention in the classroom. The Coh-Metrix variables used in this study have great potential for use in assessment. An assessment of writing that is partially or fully automated would be beneficial to teachers, psychologists and ultimately, students. This study analyzed how nine specific Coh-Metrix variables fit into a model of written expression and found some did not in the current sample. However, the variables that did fit into the model of written expression could be used in schools to analyze specific levels of language in student writing, as well as overall writing quality.

A meta-analysis of spelling instruction found strong support for teaching spelling explicitly to develop spelling skills and additionally influences reading development (Graham and Santangelo, 2014). This meta-analysis and the results of the present study and others should be a strong argument for teaching spelling explicitly for spelling and writing ability.

Abbot, Berninger and Fayol (2011) also found that spelling was the most stable skill across grade levels of the writing-related skills they examined. The authors used this finding to argue for the significance of spelling in children's writing education, as their results also indicated spelling was significantly related to other writing skills. However, in their literature review, they identified a lack of strategic, consistent spelling instruction in classrooms across the country (Abbott et al., 2010).

Teacher preparation for all grades should include knowledge of the components of written language and how written expression is assessed. A deep understanding of these topics would allow teachers to better intervene when students struggle. This research specifically helps teachers to understand how spelling may influence different levels of language and writing quality. With a more accurate and thorough model like this study puts forth, teachers can analyze student writing for specific areas of need and foci for intervention.

A real need for educators and students is less labor-intensive way for teachers to assess written expression and provide specific feedback. This feedback, in turn, can lead to specific instruction and intervention. Automated tools for assessing writing would allow this, but researchers first need to establish which components of writing are meaningful for identifying struggling writers and for intervention.

Implications for Future Research

Most assessment and authentic writing tasks have technology supports available (e.g., spell-check, and thesaurus). For example, the National Assessment of Educational Progress also reports interesting statistics on the ways in which middle schoolers use the computer-based writing assessment and what tools they may use in this computer interface (e.g., thesaurus spell-check). The influence of these readily accessible tools may change the way students and adults produce written expression and may change the model of written expression entirely.

The ultimate goal of both writing instruction and intervention is to improve writing quality, and the strategies to do this greatly depend on the purpose and form of writing being done (Troia et al., 2019). More research is needed on how genres of writing may differ in models of written expression. For example, does spelling constrain the levels of language differently in

informational vs. narrative text? Many further questions exist as to how language differs in genres of written expression and how language and skills may vary across these genres.

Wilson and colleagues (2017) argue the use of on-demand, first draft only written products in assessment differs from authentic writing situations. Future research should aim to complete a study with multiple drafts of the same writing and analyze the first vs. final draft for differences in how a model of written expression fits different drafts. Such a study would more closely resemble the way individuals write in schools and in the real world.

The Coh-Metrix variables used in this study, as well as the myriad of other variables in the Coh-Metrix system, have tremendous potential for semi-automated assessment of written expression. However, these variables are in their infancy in terms of being understood as they fit into a model of written expression to be used in assessment. Further research is needed to establish how these Coh-Metrix variables measure written language, predict outcomes for students and can be of best use for progress monitoring and assessing written expression.

Conclusion

Overarching questions that guided this study included: why are so many students failing to reach proficiency in written expression? And: how can we accurately assess the areas in which these students struggle for targeted instruction and intervention? These questions are meaningful for student outcomes in school and life, as written expression is a vital skill for students and has outcomes related to grade retention, standardized assessment success and graduation (Jenkins et al., 2004).

This study aimed to further the development of a model of written language to guide assessment and instruction in schools with a specific focus on the role of spelling in the model of written language. A comprehensive model of written language would guide understanding of

what skills make up the ability to produce written expression and how these skills and their interactions change individually and developmentally. Establishing a model of written expression, the cognitive processes involved, and how these components and skills interact would allow an understanding of what needs to be assessed in written expression and importantly, why certain students may struggle.

This study tested one empirically established model of written expression to analyze how spelling may influence the levels of language in their prediction of writing ability. The hypothesized model, based on work by Wilson and colleagues (2017) used three levels of language (word, sentence, text) and including spelling, with specific aims of analyzing how spelling constrained these levels in their prediction of writing quality. This study found spelling constrained only the text-level of language in its prediction of writing quality, and that an interaction existed in which if the spelling scores and text scores were better, the results for writing quality were even greater.

These results suggest that spelling should not be relegated to the accuracy dimension of word-level skills within a levels of language model for written expression. Overall, the study has many implications for future research and practice. Mainly, this study adds to desperately needed analysis of samples to understand how the model of written expression behaves differently across individuals and developmental levels. This study also directs future research in the same goals, by confirming and questioning various prior studies, and adding an understanding of how spelling may fit into the model. Ultimately, the need for more research on written expression is needed to establish a comprehensive model of written expression for assessment and instruction to understand why students struggle in writing, in what specific ways they struggle, and to monitor their growth in written expression.

APPENDICES

APPENDIX A

The Writing Architect Prompts

Informative (passage) prompts

Remember, a well written informative paper (1) has a clear main idea and stays on topic, (2) includes a good introduction and conclusion, (3) uses information from the article stated in your own words plus your own ideas, and (4) follows the rules of writing.

Frigid Northern China hosts snow and ice sculpture festival

How would you like to spend Thanksgiving in space?

This is how bats can land upside down

13 Year Old World War II Veteran

Here's a food wrapper you can eat

Plastic bottle village

Swat up: Six reasons to love flies

Can an Elevated Bus Solve China’s Traffic Woes

What sort of spider can capture its prey without a web?

Scientists find that dogs understand what you're saying

Why do Alaskan volcanoes erupt so often?

Furniture of the Future

Thorny Devil

Kingdom of Ghana

Visits to National Parks Sets Record

Trapped Ants

Study Reveals Surprising Facts About Our Choice of Emojis

Example Informational Passage

Why do Alaskan volcanoes erupt so often?

Source: Smithsonian Tween Tribune

A remote volcano in Alaska's Aleutian Islands has erupted 10 times in less than a month. Experts say more eruptions are possible.

Bogoslof volcano has sent up ash clouds that have reached as high as 35,000 feet.

The Alaska Volcano Observatory is a joint program of the U.S. Geological Survey and the University of Alaska Fairbanks. It says 90 volcanoes have been active within the last 10,000 years. And they could erupt again. More than 50 have been active since about 1760. That is when record-keeping began.

Like Bogoslof, most are on the 1,550-mile-long Aleutian Arc. The area forms the northern portion of the Pacific "Ring of Fire." The ring is a horseshoe-shape zone around the Pacific Ocean of frequent earthquakes and volcanic eruptions. These are triggered by the subduction of an oceanic plate beneath continental plates.

Volcanoes in Alaska erupt regularly. Pavlof Volcano sent up ash clouds in 2013. Cleveland volcano blew in December 2011. Redoubt volcano 100 miles southwest of Anchorage blew in March 2009, dropping ash during the medals ceremony for the U.S. alpine ski championships at Alyeska Resort in Girdwood. Some volcanoes erupt and spit out additional ash intermittently for weeks, as Bogoslof seems to be doing.

The Alaska Volcano Observatory was formed in response to the 1986 eruption of Mount Augustine. The observatory has tools to predict eruptions. As magma moves beneath a volcano before an eruption, it often generates earthquakes. They swell the surface of a mountain and increase the gases emitted. The observatory samples the gases. It also measures earthquake activity and watches for landscape deformities.

The observatory uses mathematical models to forecast how fast ash particles will be transported in the atmosphere. And to determine where ash could fall. The observatory runs the models when it detects that a volcano might erupt. It also updates them when they blow.

What makes Alaska volcanoes so dangerous? Volcanoes in Hawaii ooze lava. But volcanoes in Alaska tend to explode.

Instead of a red river of lava, Alaska volcanoes typically shoot ash up to 50,000 feet. That is more than nine miles. It reaches the jet stream.

That ash is not the kind you left after a campfire. Instead, it's an abrasive kind of rock fragment. The particulate has jagged edges. It has been used as an industrial abrasive to polish metals.

Particulate can injure skin, eyes and breathing passages. The young, the elderly and people with respiratory problems are especially susceptible. Ash under a windshield wiper can scratch glass. However, most volcanoes are far from communities. So ash fall that requires breathing masks or new air filters on a car is infrequent.

USGS geophysicist John Power once likened flying through an ash cloud to flying into a sandblaster.

Ash can scrape the moving parts of jet engines such as turbine blades. However, ash on hot parts of a jet engine is potentially more dangerous, according to the observatory. The engines operate near the melting temperature of volcanic ash.

"Ingestion of ash can clog fuel nozzles, combustor, and turbine parts causing surging, flame out, immediate loss of engine thrust, and engine failure," according to the observatory.

Using information provided by the Federal Aviation Administration, the observatory estimates that more than 80,000 large aircraft per year, and 30,000 people per day, fly on routes downwind of Aleutian volcanoes. These are along great-circle routes between Europe, North America and Asia.

Airlines get excited when an ash cloud rises above 20,000 feet.

The jet stream can carry ash for hundreds of miles. Ash from Kasatochi Volcano in August 2008 blew all the way to Montana.

Redoubt volcano blew on Dec. 15, 1989. It sent ash 150 miles away into the path of a KLM jet carrying 231 passengers. Its four engines flamed out.

As the crew tried to restart the engines, "smoke" and a strong odor of sulfur filled the cockpit and cabin. The jet dropped more than 2 miles, from 27,900 feet to 13,300 feet. The crew finally was able to restart all engines. The plane landed safely at Anchorage.

So what are the chances for a major, catastrophic eruption?

"That's always a possibility but big eruptions have precursor signals," said USGS research geophysicist Chris Waythomas. "That just doesn't happen in 20 minutes."

Months of below-ground unrest can precede a major eruption. The Alaska Volcano Observatory, Waythomas said, likely would be tipped off by movement of the huge volume of magma involved.

"It has to break a lot of rock to get to the surface," he said.

Prompt: There are no communities near the Alaskan volcanoes. Write an informative paper describing why people would not want to live near these volcanoes. Remember, a well written informative paper (1) has a clear main idea and stays on topic, (2) includes a good introduction and conclusion, (3) uses information from the article stated in your own words plus your own ideas, and (4) follows the rules of writing.

APPENDIX B

Code Book: The Writing Architect

Figure 7.

Writing Architect Codebook Page One

WRITING ARCHITECT CODEBOOK

CONTENTS

VARIABLES IN WRITING ARCHITECT

COMPUTER CALCULATED

- Time spent writing
- Typing fluency-number of words
- total words (at 3 minutes, 5, 7, 10, & 15)
- number of sentences
- number of paragraphs

HUMAN-SCORED, COMPUTER TALLIED

All variables are available for the 15 minute (final) sample. Some can also be marked in the 3, 5, 7, and 10-minute sample.

- correct sequences
- incorrect sequences
    - word-level errors (calculated as number of misspelled words divided by total number of words):
        - spelling

1

Figure 8.

Writing Architect Codebook Page Two

        - capitalization
    - sentence-level errors (calculated as number of incorrect sequences divided by total number of sequences)
        - punctuation
        - word meaning doesn't fit (semantic)
        - missing word
        - transposed or extra word
        - subject/verb disagreement, noun/modifier disagreement, tense error
- percent correct sequences
- percent incorrect sequences
- correct minus incorrect sequences
- writing quality total (0 to 35)
    - purpose (0-5)
    - discourse level (0-5)
    - conclusion (0-5)
    - sentence-level cohesion (0-5)
    - supporting details (0-5)
    - language choice (0-5)
    - mechanics (0-5)
- coh-metrix – all variables
- planning

STUDENT FACING MATERIALS

To take the assessment like a student. Go to
https://msu.co1.qualtrics.com/jfe/form/SV_9nVzmrwdhZJSznD?Q_JFE=qdg

Use the student ID number 9999.

*****If there are any notes from the administrator in the packet, transfer them to the notes section in Filemaker at the final essay section. **********

PLANNING SHEETS

Find the student and prompt in filemaker that matches the physical copy.

In the 15-minute tab, click 'final score'. Enter the planning score here.

Figure 9.

Writing Architect Codebook Page Three

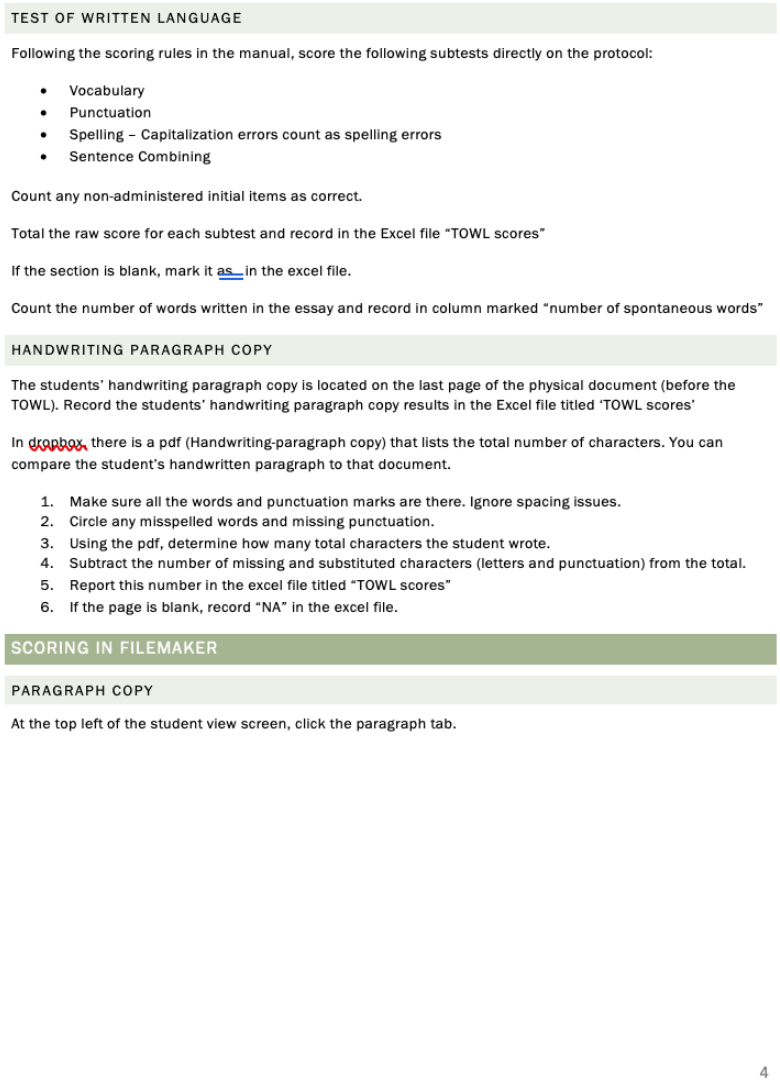
SCORING RULES

0 = minimal planning	• less than or equal to 5 words
1 = drafting	• More than 5 words • Block of text similar to the text in the typed sample
2 = basic outline	• List of 2 or more thoughts • List may or may not use bullets
3 = organizer	• Uses terms like intro, conclusion, main idea, points, warrants, claims • Draws a venn diagram • Separates space into 2 columns

3

Figure 10.

Writing Architect Codebook Page Four



TEST OF WRITTEN LANGUAGE

Following the scoring rules in the manual, score the following subtests directly on the protocol:

- Vocabulary
- Punctuation
- Spelling – Capitalization errors count as spelling errors
- Sentence Combining

Count any non-administered initial items as correct.

Total the raw score for each subtest and record in the Excel file "TOWL scores"

If the section is blank, mark it as in the excel file.

Count the number of words written in the essay and record in column marked "number of spontaneous words"

HANDWRITING PARAGRAPH COPY

The students' handwriting paragraph copy is located on the last page of the physical document (before the TOWL). Record the students' handwriting paragraph copy results in the Excel file titled 'TOWL scores'

In dropbox there is a pdf (Handwriting-paragraph copy) that lists the total number of characters. You can compare the student's handwritten paragraph to that document.

1. Make sure all the words and punctuation marks are there. Ignore spacing issues.
2. Circle any misspelled words and missing punctuation.
3. Using the pdf, determine how many total characters the student wrote.
4. Subtract the number of missing and substituted characters (letters and punctuation) from the total.
5. Report this number in the excel file titled "TOWL scores"
6. If the page is blank, record "NA" in the excel file.

SCORING IN FILEMAKER

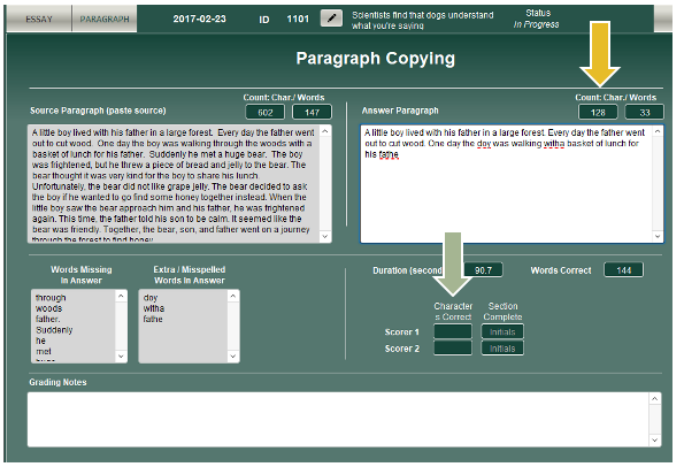
PARAGRAPH COPY

At the top left of the student view screen, click the paragraph tab.

4

Figure 11.

Writing Architect Codebook Page Five



If the text box is blank, leave the 'characters correct' box blank and fill in your initials.

The numbers at the top right are the number of characters that the student typed and the number words they typed. We will subtract the number of letters incorrectly typed from the character count.

Incorrectly typed letters includes substitutions and additions only. Spacing issues, and missing letters or missing punctuation are ignored.

The above example has 128 total characters (see yellow arrow in the figure above). The example has only 1 letter substitution (doy for boy). Therefore the Characters Correct count to be entered into the box (green arrow) is 127.

Fill in the characters correct box and the Section Complete box.

ESSAY SCORING

As you complete each section of scoring, type your initials in the initials box.

If you encounter a blank essay section, list the student ID number and the prompt title in the excel file in dropbox ('remove from filemaker.xlsx').

If there are 2 essays for 1 prompt, score both.

QUALITY RUBRIC

Scoring Dimensions:

- Orients the reader to the purpose of the text effectively and creatively
- Groups related ideas to enhance text coherence logically and insightfully
- Provides a concluding sentence or section that follows smoothly from prior ideas

5

Figure 12.

Writing Architect Codebook Page Six

- Links ideas using words or phrases precisely and effectively for strong cohesion
- Develops ideas using facts, examples, experiences, descriptive details, dialogue/quotes (from source materials as appropriate) that are relevant and impactful
- Uses language and vocabulary that is precise, varied, and apt for the type of text
- Is free of errors in grammar, usage, and mechanics (spelling, capitalization, and punctuation)

Rating Scale for Each Dimension:

0. No evidence of dimensional quality; severely flawed/incomprehensible
1. Minimal evidence of dimensional quality; substantially flawed/difficult to read
2. Some evidence of dimensional quality; notably flawed but readable
3. Adequate evidence of dimensional quality; a few consistent flaws but readable
4. Strong evidence of dimensional quality; some inconsistent flaws/easy to read
5. Excellent evidence of dimensional quality; virtually no flaws/fully comprehensible

Scoring Procedures:

a) Separate group of papers (20 – 25 papers of the same prompt from the same class) to score into three sets representing low, medium, and high overall quality based on initial impression.
b) Score each of the seven dimensions separately using the six-point scale above, working in reverse order from last to first dimension.
c) Sum the total points awarded across dimensions to obtain the paper's score (range from 0 to 35)
d) In the notes section:
    a. If the student seems to be copying the passage, write "duplication" and do not score for quality.
    b. If the essay is completely non-responsive to the prompt, write "nonresponsive" and do not score for quality.

Enter scores in Filemaker

- Search for the student number
- Go to minute 15, then click the 'final score' box.
- Enter scores in the Writing Quality section

Figure 13.

Writing Architect Codebook Page Seven

COH-METRIX

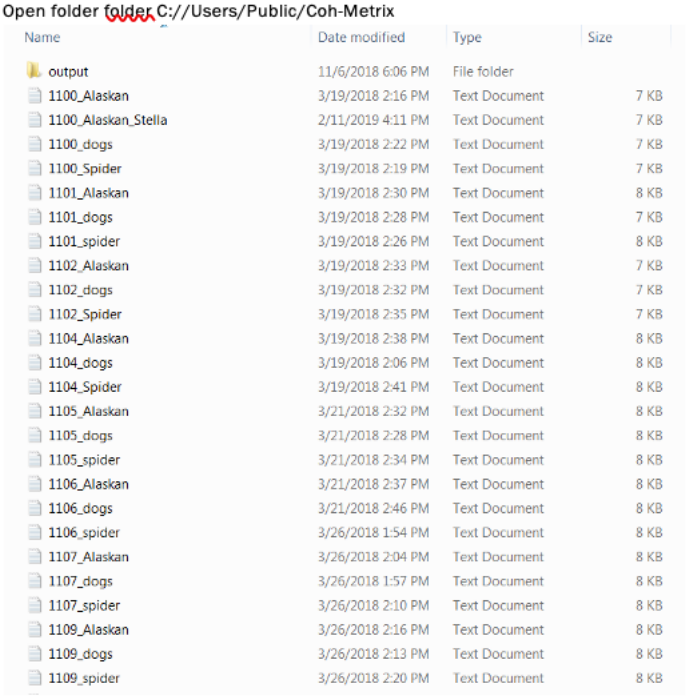
A video demonstrating how to score Coh-Metrix with an example can be found here: https://mediaspace.msu.edu/media/coh-metrix+scoring/1_braqpw9t

1. Copy and paste the text from the final score page (the 'source data' box) to the Coh-Metrix tab; then correct the text for:
- spelling
- punctuation
- capitalization
- missing words
- egregious syntax errors
2. After each one, click save. Name the text file with ID number and prompt (5100_vet) and save in folder C://Users/Public/Coh-Metrix.
3. A dialog box will ask you if you'd like to open the file, click open.
4. Copy and paste the text that you edited at the bottom of the text file.
5. Indicate how many texts you completed for that student ID in the excel file titled "Coh-Metrix Progress"

If there are any duplicates of the same passage and ID, do NOT run it through Coh-Metrix. Highlight the ID number in the excel document "Coh-Metrix progress".

Second Scorer

1. Open folder folder C://Users/Public/Coh-Metrix



Name	Date modified	Type	Size
output	11/6/2018 6:06 PM	File folder	
1100_Alaskan	3/19/2018 2:16 PM	Text Document	7 KB
1100_Alaskan_Stella	2/11/2019 4:11 PM	Text Document	7 KB
1100_dogs	3/19/2018 2:22 PM	Text Document	7 KB
1100_Spider	3/19/2018 2:19 PM	Text Document	7 KB
1101_Alaskan	3/19/2018 2:30 PM	Text Document	8 KB
1101_dogs	3/19/2018 2:28 PM	Text Document	7 KB
1101_spider	3/19/2018 2:26 PM	Text Document	8 KB
1102_Alaskan	3/19/2018 2:33 PM	Text Document	7 KB
1102_dogs	3/19/2018 2:32 PM	Text Document	7 KB
1102_Spider	3/19/2018 2:35 PM	Text Document	7 KB
1104_Alaskan	3/19/2018 2:38 PM	Text Document	8 KB
1104_dogs	3/19/2018 2:06 PM	Text Document	8 KB
1104_Spider	3/19/2018 2:41 PM	Text Document	8 KB
1105_Alaskan	3/21/2018 2:32 PM	Text Document	8 KB
1105_dogs	3/21/2018 2:28 PM	Text Document	8 KB
1105_spider	3/21/2018 2:34 PM	Text Document	8 KB
1106_Alaskan	3/21/2018 2:37 PM	Text Document	8 KB
1106_dogs	3/21/2018 2:46 PM	Text Document	8 KB
1106_spider	3/26/2018 1:54 PM	Text Document	8 KB
1107_Alaskan	3/26/2018 2:04 PM	Text Document	8 KB
1107_dogs	3/26/2018 1:57 PM	Text Document	8 KB
1107_spider	3/26/2018 2:10 PM	Text Document	8 KB
1109_Alaskan	3/26/2018 2:16 PM	Text Document	8 KB
1109_dogs	3/26/2018 2:13 PM	Text Document	8 KB
1109_spider	3/26/2018 2:20 PM	Text Document	8 KB

7

Figure 14.

Writing Architect Codebook Page Eight

2. Open each file, scroll to the bottom. Check the writing part for:
    - spelling
    - punctuation
    - capitalization
    - missing words
    - egregious syntax errors
    - if the last sentence is incomplete, just leave it as is
3. If it's all correct, add "-D" to the end of the file name.
4. If not, make the corrections. Save it. Close it. Add "redo" to the filename.

WRITING SEQUENCE SCORING

CORRECT WRITING SEQUENCES

A sequence is defined as the connection from one written UNIT to the next (i.e., adjacent writing units). Adjacent writing units could be two words or it could be a word and a punctuation mark.

Marking a correct writing sequence indicates that BOTH adjacent writing units correctly utilize spelling, punctuation, grammar, syntax, and capitalization. This manual uses a carat (^) to designate correct writing sequences. Carats have been added to the two example sentences below to demonstrate the scoring of correct writing sequences.

> **^I^ think^ the^ author^ cares^ about^ what^ his^ family^ thinks^ about^ him^. ^He^ wants^ to^ make ^his ^family^ happy^.**

In the above example, note that correct sequences are marked between every word and marked before and after each punctuation mark. In this example each sequence is considered to be correct because each word is correctly spelled, the syntax and grammar features are correct, words are appropriately capitalized, and end punctuation is used appropriately.

When errors in spelling, punctuation, grammar, syntax, or capitalization are made, the sequence is marked as incorrect.

INCORRECT WRITING SEQUENCES

Every writing sequence (i.e., adjacent writing units) should receive a marking. The marking indicates if that particular sequence is correct or incorrect. This manual uses the letter (X) to designate incorrect writing sequences. Incorrect writing sequences are marked if the unit on EITHER side of a sequence violates a rule of spelling, syntax, grammar, punctuation, or capitalization. The specific rules for each of these areas is detailed below.

SPELLING

An incorrect sequence is marked on both sides of a word that is spelled incorrectly.

- The word is spelled incorrectly for the context of the sentence.

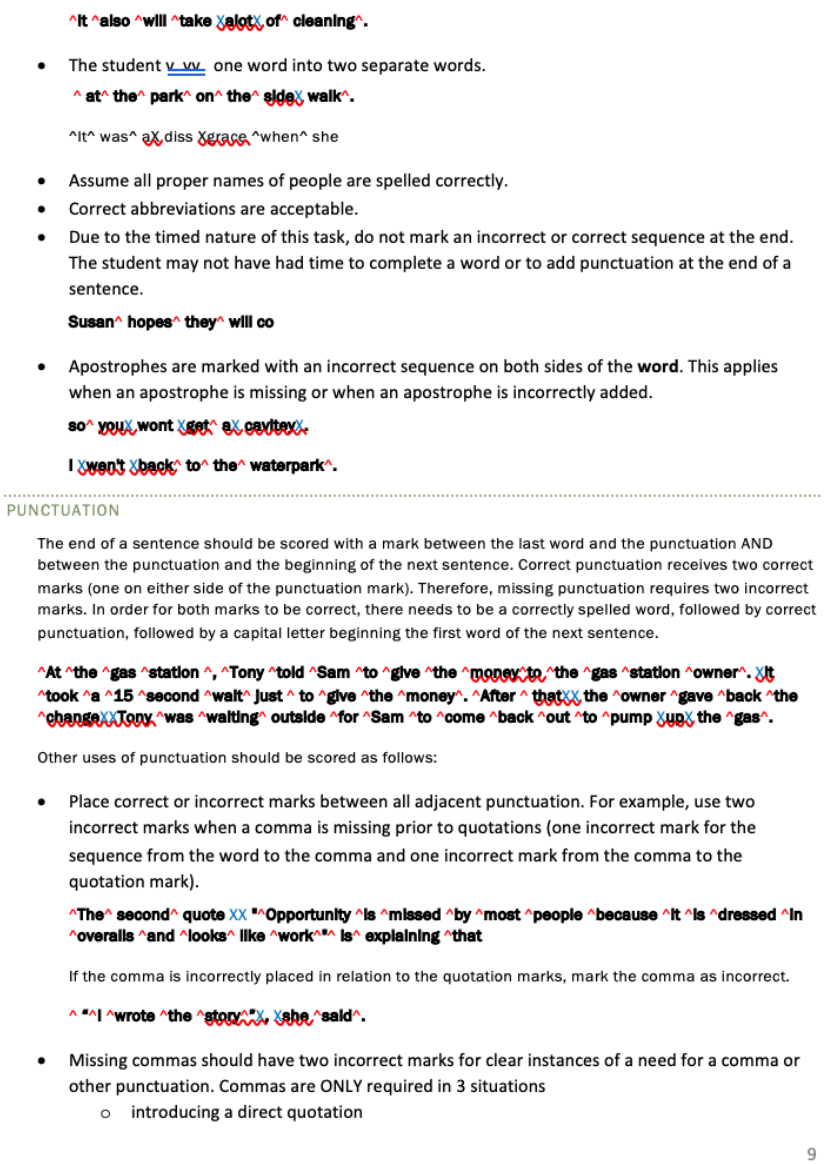
  ^Mom ^and^ the^ boy^ will^ ride X there X bikes^.

- The student combines two words into one.

8

Figure 15.

Writing Architect Codebook Page Nine



^It ^also ^will ^take Xalot of^ cleaning^.

- The student v vv one word into two separate words.

^ at^ the^ park^ on^ the^ side walk^.

^It^ was^ a diss Xgrace ^when^ she

- Assume all proper names of people are spelled correctly.
- Correct abbreviations are acceptable.
- Due to the timed nature of this task, do not mark an incorrect or correct sequence at the end. The student may not have had time to complete a word or to add punctuation at the end of a sentence.

Susan^ hopes^ they^ will co

- Apostrophes are marked with an incorrect sequence on both sides of the **word**. This applies when an apostrophe is missing or when an apostrophe is incorrectly added.

so^ you wont Xget^ a cavitey.

I Xwen't Xback^ to^ the^ waterpark^.

PUNCTUATION

The end of a sentence should be scored with a mark between the last word and the punctuation AND between the punctuation and the beginning of the next sentence. Correct punctuation receives two correct marks (one on either side of the punctuation mark). Therefore, missing punctuation requires two incorrect marks. In order for both marks to be correct, there needs to be a correctly spelled word, followed by correct punctuation, followed by a capital letter beginning the first word of the next sentence.

^At ^the ^gas ^station ^, ^Tony ^told ^Sam ^to ^give ^the ^money^to ^the ^gas ^station ^owner^. Xit ^took ^a ^15 ^second ^wait^ just ^ to ^give ^the ^money^. ^After ^ thatXX the ^owner ^gave ^back ^the ^changeXXTony ^was ^waiting^ outside ^for ^Sam ^to ^come ^back ^out ^to ^pump Xup the ^gas^.

Other uses of punctuation should be scored as follows:

- Place correct or incorrect marks between all adjacent punctuation. For example, use two incorrect marks when a comma is missing prior to quotations (one incorrect mark for the sequence from the word to the comma and one incorrect mark from the comma to the quotation mark).

^The^ second^ quote XX "^Opportunity ^is ^missed ^by ^most ^people ^because ^it ^is ^dressed ^in ^overalls ^and ^looks^ like ^work^"^ is^ explaining ^that

If the comma is incorrectly placed in relation to the quotation marks, mark the comma as incorrect.

^ "^I ^wrote ^the ^story^". Xshe ^said^.

- Missing commas should have two incorrect marks for clear instances of a need for a comma or other punctuation. Commas are ONLY required in 3 situations
    - introducing a direct quotation

9

Figure 16.

Writing Architect Codebook Page Ten

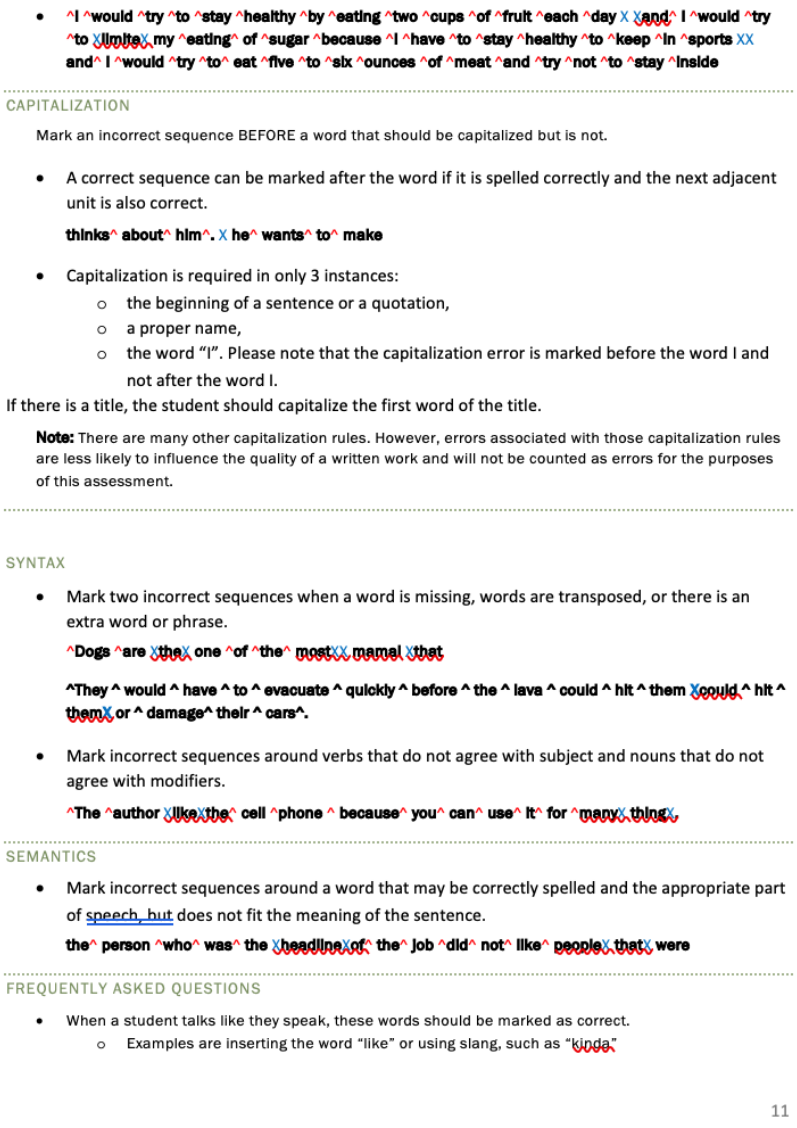
- in a series
  - The CCSS has not specifically stated whether Oxford commas are necessary, so both of the following examples are correct uses of commas in a series
    Example: ^It ^will ^make ^people ^sick^, ^cough^ or^ itch^.
    Example: ^It ^will ^make ^people ^sick^, ^cough^ ^ or^ itch^.
- after an introductory phrase or clause
- The comma rule does NOT apply to transition words.
  Example: ^Also ^we^ can ^ride
- If commas or any other extraneous mark of punctuation are inserted incorrectly, mark incorrect writing sequences around the comma
  example: ^The ^article ^tells ^you ^what ^plants ^are ^good ^to ^have ^in ^FloridaX Xand ^what^ plants ^are ^bad ^to ^have ^in ^Florida^.
- When numbers are used, mark correct sequences when commas are used correctly or not used at all. Only mark incorrect sequences if the commas are misused.
  example: ^Maldives ^has ^a ^population ^of ^340000^.
  example: ^Maldives ^has ^a ^population ^of X 3400,00 X.

- If quotation marks, colons, semicolons, ellipses, parentheses, or dashes are used correctly, mark correct sequences on both sides of the punctuation. Only mark incorrect sequences if these punctuation marks are clearly misused or if their absence creates a sentence fragment.
  - Example: just ^three ^natural ^ingredients X ; Xmycelium ^(^the ^vegetative ^part ^of ^mushrooms^)^, ^hemp^, ^and ^salt^.
- Hyphens – ignore them. If a missing hyphen creates a word that can't stand alone, score it as incorrectly split.
  - ^A ^seven-^year-^old ^boy ^joined ^the ^Navy^.
  - ^A ^seven ^year-^old ^boy ^joined ^the ^Navy^.
  - ^A ^7-^year-^old ^boy ^joined ^the^ Navy^.
  - ^Using ^the ^post-test^ data
  - despite^ the^ almost^ non livable^ conditions^.
- Incorrect spacing around punctuation is ok and correct sequences should be used. In other words, spacing does not affect scoring
- For any errors where the incorrect punctuation is used, use the 'missing punctuation' button.
- For run-on sentences determine where a logical break should be located and mark TWO incorrect sequences in that location.
- Some will have multiple run-ons. After identifying the first run on, the next determination of a run-on should be based on assuming the previous mistake was corrected.
  - For example: ^This ^story^ is ^about ^a ^girl xname xariellexx that^ she ^was ^going ^fishing ^with ^her^ dad ^and ^her^ brotherx ieovx and xarielle^ did^ not^ want^ to^ go ^fishing xxshe^ was^ going ^to ^the ^mall ^but^ her^ momx coughtx a^ cold xx now^ she ^has ^to ^gox fichingx ^Then ^she ^got^ her ^stuff

10

Figure 17.

Writing Architect Codebook Page Eleven



- ^I ^would ^try ^to ^stay ^healthy ^by ^eating ^two ^cups ^of ^fruit ^each ^day X Xand^ I ^would ^try ^to Xlimite^ my ^eating^ of ^sugar ^because ^I ^have ^to ^stay ^healthy ^to ^keep ^in ^sports XX and^ I ^would ^try ^to^ eat ^five ^to ^six ^ounces ^of ^meat ^and ^try ^not ^to ^stay ^inside

CAPITALIZATION

Mark an incorrect sequence BEFORE a word that should be capitalized but is not.

- A correct sequence can be marked after the word if it is spelled correctly and the next adjacent unit is also correct.

  thinks^ about^ him^. X he^ wants^ to^ make

- Capitalization is required in only 3 instances:
  - the beginning of a sentence or a quotation,
  - a proper name,
  - the word "I". Please note that the capitalization error is marked before the word I and not after the word I.

If there is a title, the student should capitalize the first word of the title.

Note: There are many other capitalization rules. However, errors associated with those capitalization rules are less likely to influence the quality of a written work and will not be counted as errors for the purposes of this assessment.

SYNTAX

- Mark two incorrect sequences when a word is missing, words are transposed, or there is an extra word or phrase.

  ^Dogs ^are Xthe^ one ^of ^the^ mostXX mamal Xthat

  ^They ^ would ^ have ^ to ^ evacuate ^ quickly ^ before ^ the ^ lava ^ could ^ hit ^ them Xcould ^ hit ^ themX or ^ damage^ their ^ cars^.

- Mark incorrect sequences around verbs that do not agree with subject and nouns that do not agree with modifiers.

  ^The ^author Xlike^the^ cell ^phone ^ because^ you^ can^ use^ it^ for ^manyX thingX.

SEMANTICS

- Mark incorrect sequences around a word that may be correctly spelled and the appropriate part of speech, but does not fit the meaning of the sentence.

  the^ person ^who^ was^ the Xheadline^of^ the^ job ^did^ not^ like^ peopleX thatX were

FREQUENTLY ASKED QUESTIONS

- When a student talks like they speak, these words should be marked as correct.
  - Examples are inserting the word "like" or using slang, such as "kinda."

11

Figure 18.

Writing Architect Codebook Page Twelve

- A sequence should be marked before the very first word of a paragraph.
  - Example at the beginning: ^The ^increase ^in ^tourism...
  - Example in the middle:

    ...after ^attracting ^them ^over^.

    ^A ^spider ^that ^is ^alike...

- If the student uses a title

  ^Ants ^and^ Their^ Amazing^ Talents

  ^In^ Poland^ there^ are^ many^ ants...

- You will likely never have more than 2 incorrect sequences next to each other. If there is ever a case to be made for 3 in a row, bring it up to the group for discussion.
- How do you score the last word the student wrote? Never mark a sequence (correct or incorrect) after the last thing that the student wrote. If the student was in the middle of typing a word, ignore that last word. If they completed the word or have punctuation marked, score the sequence before the word/punctuation, but not after.

  Example: ^I ^would ^eat ^all ^of ^the ^healthy ^things
  ^that^ could ^keep^ me he
  ^Broiled ^meat ^or ^grilled ^can ^give^ you n
  ^I ^would ^eat ^all ^of ^the ^healthy ^things^.
- Agreement - mark incorrect sequences for subject verb disagreement and incongruencies with number for adjective NOTE: we are NOT looking for consistent verb tense within a sentence nor are we looking for inconsistent number when the words are not adjacent

  Example: ^There XisX a^ lot^ of^ different^ ways
  Example: ^We ^had ^3 XcarX
  Nonexample:

  If you find subject/verb disagreement that is separated by more than 2-3 words, make a note in the notes section for it to be reviewed.

- Numerals are ok and marked as correct as long as they are used correctly.
- Case - In the following example, the word should be myself rather than me. However, we are not marking incorrect sequences for problems with case.
  ^Laws^ help ^me^ and^ other^ people^ keep^ safe^.

- Symbols and abbreviations - if the symbol or abbreviation is used correctly it should be marked as correct.
  Examples:
  ^&^
  ^T.V.^ OR ^TV^
  I^like ^dogs^etc^.

- If punctuation is missing, mark the missing punctuation, but do NOT also penalize for missing capital letter.

  Example: ...^stay ^healthy ^to ^keep ^in ^sports XX and^ he^would ^try ^to^ eat ^five ^to ^six ^ounces ^of ^meat

12

Figure 19.

Writing Architect Codebook Page Thirteen

- If the student including 'planning' notes in their essay; score it like connected words.

  Examples: ^(^introduction^) @consequences ^of ^increased ^tourism ^at ^national ^parks^.

  ^(^paragraph^)@ increased ^tourism ^has^ caused ^a ^lot ^of ^chaos^.

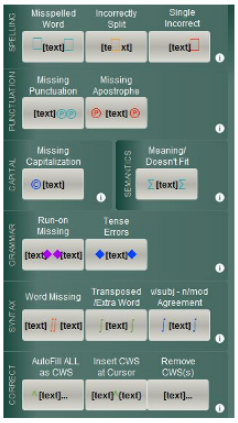
USING FILEMAKER SYMBOLS

Each button in filemaker has an information box indicating the corresponding rule.

IGNORE THE APOSTROPHE BUTTON (use spelling instead)

IGNORE THE RUN-ON BUTTON (use punctuation instead)

IGNORE THE TENSE ERRORS BUTTON (use syntax agreement instead)



- Samples may have multiple errors in rapid succession and it may be difficult to determine which error to mark first. The general rule is to mark the error that comes first when reading the sentence.
  - In the following sentence, the missing word between 'most' and 'mamal' would be marked first. The mark for the misspelling of 'mamal' would only come after the word 'mamal' and not added before the word, because the sequence is already marked incorrect for having a missing word.

    The ^Dogs ^are ∑the∑ one ^of ^the^ most mamal that

- If there are 2 spelling errors next to each other, use the regular misspelled word for the first one and the 'single incorrect' button for the second error. This is important for the 'spelling error' calculation to provide a correct calculation.

13

Figure 20.

Writing Architect Codebook Page Fourteen

An Alska volcanoe is

- Always use the when the spelling error is the 2nd error in a row
  ^Some ^ spiders ſusesſ thier webs ^ to ^ catch ^ prey^.

LIMITATIONS

A writing sequence only considers the correctness between adjacent written units; it does not always reflect grammatical issues at the paragraph level and sometimes misses grammatical issues at the sentence level. For example, it will not reflect shifts in verb tense or voice from sentence to sentence. At the sentence level, it is also sometimes very difficult to determine where a long run-on sentence should end, or if there are multiple places to mark insertion of end punctuation. Furthermore, it does not capture errors regarding parallel construction at the sentence or paragraph level.

At first, scoring of CIWS is time-consuming due to the many rules. After some practice with the scoring, the average time to score CIWS is less than 5 minutes. There may be some situations that do not fit the rules listed above, and it may be difficult to determine whether correct or incorrect sequences should be marked. Since writing has an infinite number of possible responses, difficult-to-score situations are highly likely. In these situations

1 – Score the writing using your best judgment

2 – Copy the sentence or the entire writing sample in the FAQs document on Dropbox.

3 – Include the student ID number & prompt name

ADDITIONAL EXAMPLES

Several practice examples are provided here. Each example below was written by a Florida student in response to the writing prompts in the open response diagnostic assessment of the FAIR-FS. The scoring is demonstrated below each example. The calculation of CIWS for each individual example is also listed.

I rember like it was yesterday. It was me, my sister , my other sister and my mom. My sisters and I and my mom were walking to the water ride and I looked at a sigh and I didn't see my family then the first thing I did was I went back to the waterpark and I asked if they had seen a tall wuman and two kid. They said that they were at the waterslide. I ran up to my mom . And that is the time I was lost.

^I X rember X like^ It^ was ^yesterdayXX It ^was^ me^ ,^my^ sister^ ,^my^ other^ sister ^and^ my^ mom^.^My^ sisters ^and^ I ^and ^my^ mom^ were ^walking^ to ^the^ water^ ride^ and^ I ^looked^at ^a Xsigh X and^ I ^didn't^ see^ my^ familyXX then ^the^ first^thing^ I^did ^was ^I Xwent Xback^ to^ the^ waterpark ^ and^ I ^asked^ If Xthey Xhad^ seen ^a Xtall Xwuman Xand ^two X kidX.XThayX said ^that Xthey X were^ at ^the ^watersideXX I ^ran^ up^ to^ my^ mom^.^And ^that ^is ^the ^time ^I^was^ lost^.

76 *(CWS)*

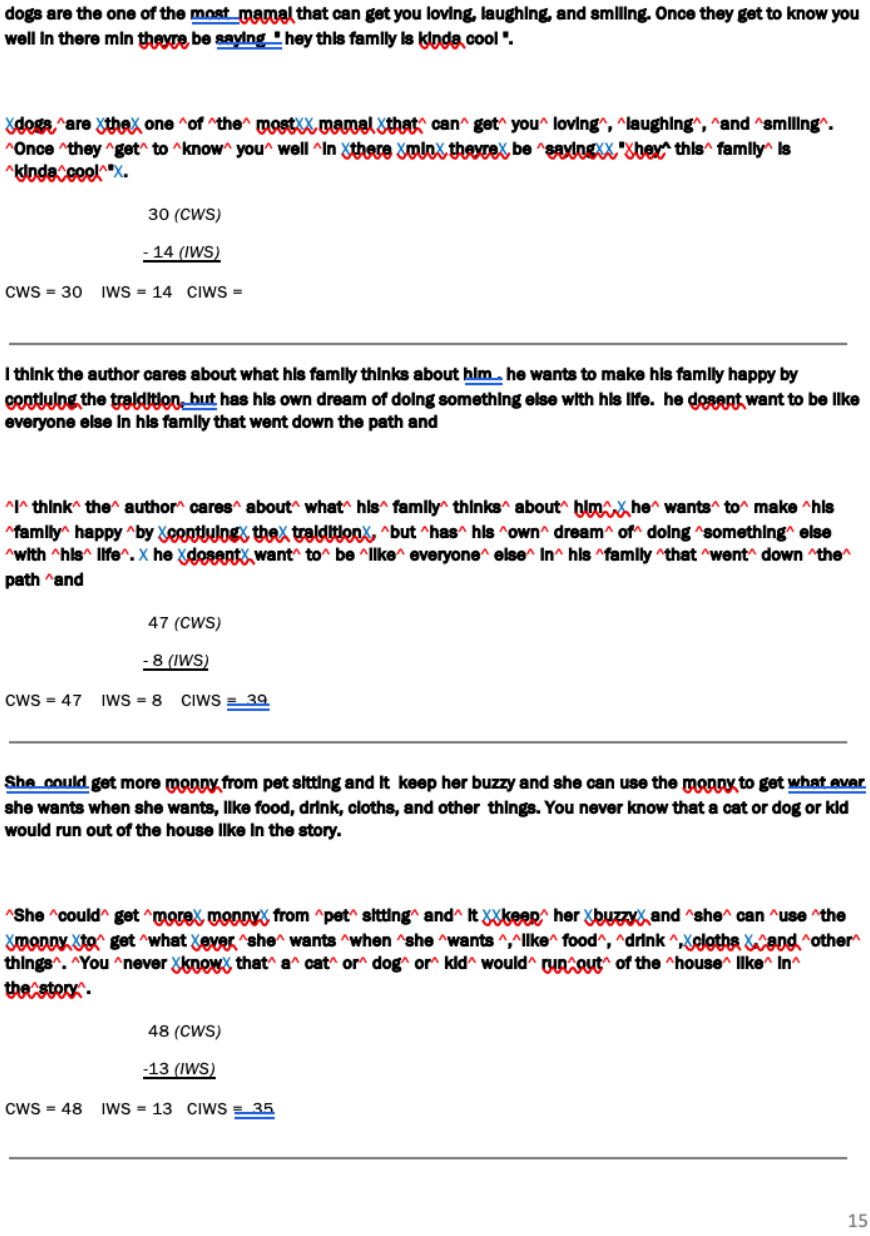
-23 *(IWS)*

CWS = 76 IWS = 23 CIWS = 53

14

Figure 21.

Writing Architect Codebook Page Fifteen



dogs are the one of the most mamal that can get you loving, laughing, and smiling. Once they get to know you well in there min theyre be saying " hey this family is kinda cool ".

Xdogs ^are XtheX one ^of ^the^ mostXX mamal Xthat^ can^ get^ you^ loving^, ^laughing^, ^and ^smiling^. ^Once ^they ^get^ to ^know^ you^ well ^in Xthere XminX theyreX be ^sayingXX "Xhey^ this^ family^ is ^kinda^cool^"X.

30 *(CWS)*

- 14 *(IWS)*

CWS = 30 IWS = 14 CIWS =

I think the author cares about what his family thinks about him , he wants to make his family happy by contiuing the traidition, but has his own dream of doing something else with his life. he dosent want to be like everyone else in his family that went down the path and

^I^ think^ the^ author^ cares^ about^ what^ his^ family^ thinks^ about^ him^,X he^ wants^ to^ make ^his ^family^ happy ^by Xcontiuing X the X traidition X, ^but ^has^ his ^own^ dream^ of^ doing ^something^ else ^with ^his^ life^. X he Xdosent X want^ to^ be ^like^ everyone^ else^ in^ his ^family ^that ^went^ down ^the^ path ^and

47 *(CWS)*

- 8 *(IWS)*

CWS = 47 IWS = 8 CIWS = 39

She could get more monny from pet sitting and it keep her buzzy and she can use the monny to get what ever she wants when she wants, like food, drink, cloths, and other things. You never know that a cat or dog or kid would run out of the house like in the story.

^She ^could^ get ^more X monny X from ^pet^ sitting^ and^ it XX keep^ her Xbuzzy X and ^she^ can ^use ^the Xmonny X to^ get ^what Xever ^she^ wants ^when ^she ^wants ^,^like^ food^, ^drink ^,Xcloths X ,^and ^other^ things^. ^You ^never Xknow X that^ a^ cat^ or^ dog^ or^ kid^ would^ run^out^ of the ^house^ like^ in^ the^story^.

48 *(CWS)*

-13 *(IWS)*

CWS = 48 IWS = 13 CIWS = 35

15

Figure 22.

Writing Architect Codebook Page Sixteen

Well_e and her family ever since she was a little girl she and her family would always go out to a restraunt and they would get caught up on the news and talk about there weekends. she had said that she usaully gets the pancakes , but this time her dad said why don't you try something new like the french toast. and she said okay and when the surver brought it out. when she went and took a bute she said she liked it. So for now on she gets pancakes one time and the french toast the ne t when her and her family get together. but now that she is older and goes to college she doesn't do it ever sunday.

^Well X e X and ^her ^family X ever ^since ^she ^was^ a ^little ^girl ^she ^and ^her ^family ^would ^always ^go ^out ^to ^a X restraunt X and ^they ^would ^get ^caught ^up ^on ^the^ news ^and ^talk ^about X X there X weekends^. X she ^had ^said ^that ^she X usaully X gets ^the ^pancakes^. X but ^this ^time ^her ^dad ^said XX why^^don't ^you ^ try ^something ^new ^like ^the ^french ^toast^. X and ^she ^said ^okay X and X when ^the X surver X brought^ it^ out^. X when ^she ^went ^and ^took ^a X bute X she ^said ^she^ liked ^it^. ^So X for X now ^on ^she ^gets ^pancakes ^one ^time ^and ^the ^french ^toast ^the ^next X when^ her ^and ^her ^family ^get ^together^. X but ^now ^that ^she ^is ^older ^and ^goes ^to ^college XX she ^doesn't ^do ^it X ever X sunday^.

103 *(CWS)*

- 32 *(IWS)*

CWS = 103 IWS = 32 CIWS = 71

The relationship between lamar and his father is a person who tell his dad anything that deals with work or something private. Even though his father have good intention, but sometimes he will get a little to rough about the situation. Lamar father was a person is like if you cant ace the test you will not be able to drive.

^The^ relationship ^between X lamar ^and^ his^ father ^ is XX a^ person^ who X tell X his^ dad^ anything ^that^ deals ^with ^work ^or^something^private^. ^Even^ though ^his^ father X have X good ^intention^, X but X sometimes^ he ^will ^get^ a^ little X to X rough^ about^ the ^situation^. X Lamar X father was^ a X person X is ^like ^if^ you X cant X ace^the^ test ^you ^will ^not^ be ^able ^to ^drive^.

48 *(CWS)*

-17 *(IWS)*

CWS = 48 IWS = 17 CIWS = 31

APPENDIX C

Writing Architect: Quality Scoring Rubric

Table 13.

Writing Architect Quality Scoring Rubric

Scoring Dimension	No evidence of dimensional quality; severely flawed/difficult to read	Minimal evidence of dimensional quality; substantially flawed/difficult to read	Some evidence of dimensional quality; notably flawed but readable	Adequate evidence of dimensional quality; a few consistent flaws but readable	Strong evidence of dimensional quality; some inconsistent flaws/easy to read	Excellent evidence of dimensional quality; virtually no flaws/fully comprehensible
Orients the reader to the purpose of the text effectively and creatively	0	1	2	3	4	5
Groups related ideas to enhance text coherence logically and insightfully	0	1	2	3	4	5
Provides a concluding sentence or section that follows smoothly from prior ideas	0	1	2	3	4	5

Table 13 *(cont'd)*

Links ideas using words or phrases precisely and effectively for strong cohesion	0	1	2	3	4	5
Develops ideas using facts examples, experiences, descriptive details, dialogue/quotes (from source materials as appropriate) that are relevant and influenceful	0	1	2	3	4	5
Uses language and vocabulary that is precise, varied, and apt for the type of text	0	1	2	3	4	5
Is free of errors in grammar, usage, and mechanics (spelling, capitalization, and punctuation)	0	1	2	3	4	5

REFERENCES

REFERENCES

Abbot, R.D., Berninger, V.W. & Fayol, M. (2010) Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102 (22),* 281-298. https://doi.org/10.1037/a0019318

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of Response to Intervention Practices for Elementary School Reading* (NCEE 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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

Benson, B. J., & Campbell, H. M. (2009). Assessment of student writing with curriculum-based measurement. In G.A. Troia (Ed.) *Instruction and Assessment for Struggling Writers* (pp. 337-357). Guilford.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246. https://doi.org/10.1037/ 0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Berninger, V. W., Garcia, N. P., & Abbott, R. D. (2009). Multiple processes that matter in writing instruction and assessment. In G.A. Troia (Ed.) *Instruction and Assessment for Struggling Writers* (pp. 15-50). Guilford.

Berninger, V.W., Fuller, F., & Whitaker, D. A process model of writing development across the lifespan. *Educational Psychology, 8,* 193-218. https://doi.org/10.1007/BF01464073

Berninger, V.W., Stage, S.A., Smith, D.R. & Hildebrand, D. (2001). Assessment for reading and writing intervention: A three-tier model for prevention and remediation. In J.J.W. Andrews, D.H. Saklofske, H.L. Janzen (Eds.), *Handbook of psychoeducational assessment: Ability, achievement, and behavior in children* (195-223). Academic Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136–162). Sage.

Calfee, R. C., & Miller, R. G. (2007). Best practices in writing assessment. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction*. Guilford.

Campbell, H., Espin, C. A., & McMaster, K. (2013). The technical adequacy of curriculum-based writing measures with English learners. *Reading and Writing, 26*, 431-452. doi: 10.1007/s11145-012-9375-6

Christ, T. J., & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology, 47*(1), 55–75. doi:10.1016/j.jsp.2008.09.004

Cullham. R. (2003). *6+1 traits of writing.* Scholastic Inc.

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. Journal of Educational Psychology, 100(4), 907–919. https://doi.org/10.1037/a0012656

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*, 184-192. doi:10.1177/00224669030370030801

Deno, S. L., Marston, D., & Mirkin, P. K. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48*, 368-371. doi:10.1177/001440298204800417

Dockrell, J.E., Lindsay, G, Connelly, V. (2009). The influence of specific language impairment on adolescents' written text. *Exceptional Children, 75*, pp. 427-446

Durst, R. K., & Newell, G. E. (1989). The uses of function: James Britton's category system and research on writing. *Review of Educational Research, 59*(4), 375–394. https://doi.org/10.2307/1170204

Espin, C. A., De La Paz, S., Scierka, B. J., & Roelofs, L. (2005). The relationship between curriculum-based measures in written expression and quality and completeness of expository writing for middle school students. *The Journal of Special Education, 38*(4), 208–217. doi:10.1177/00224669050380040201

Espin, C., Shin, J., Deno, S. L., Skare, S., Robinson, S., & Benner, B. (2000). Identifying indicators of written expression proficiency for middle school students. *The Journal of Special Education, 34*(3), 140–153. doi:10.1177/002246690003400303

Espin, C., Wallace, T., Campbell, H., Lembke, E. S., Long, J. D., & Ticha, R. (2008). Curriculum-based measurement in writing: Predicting the success of high-school students on state standards tests. *Exceptional Children*, *74*(2), 174–193 doi:10.1177/001440290807400203

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*(2), 188–193.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology, 46*, 315–342. doi:10.1016/j.jsp.2007.06.003.

Furey, W. M., Marcotte, A. M., Hintze, J. M., & Shackett, C. M. (2016). Concurrent validity and classification accuracy of curriculum-based measurement for written expression. *School Psychology Quarterly, 31*(3), 369–382. doi:10.1037/spq0000138

Gillam, R., & Jonston, J. (1992). Spoken and written language relationships in language/learning-impaired and normally achieving school-age children. *Journal of Speech and Hearing Research, 35*, 1303–1315. doi/10.1044/jshr.3506.1303

Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 457–478). Mahwah, NJ: Erlbaum.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*, 170–182. doi:10.1037//0022-0663.89.1.170

Graham, S., Harris, K.R., MacArthur, C.A & Schwartz, S. (1991). Writing and writing instruction for students with learning disabilities: Review of a research program. *Learning Disability Quarterly, 14,* 89-114.

Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the influence of writing and writing instruction on reading. *Harvard Educational Review, 81*, 710–744.

Graham, S., Gillespie, A. & McKeown, D. Writing: importance, development, and instruction. *Read Writ* **26,** 1–15 (2013). https://doi.org/10.1007/s11145-012-9395-2

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99,* 445.

Graham, S., Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing* 27, 1703–1743. https://doi.org/10.1007/s11145-014-9517-0

Gupta, R., & Kar, B. R. (2009). Development of Attentional Processes in ADHD and Normal Children. *Progress in Brain Research*, *176*, 259–276. Retrieved from https://pdfs.semanticscholar.org/cfa1/992fef8b07c175ccad3ad167bc44d310f7d2.pdf

Hair, J., Anderson, R.E., Tatham, R.L. & Black, W.C. (1995). *Multivariate data analysis, 4th edition.* Prentice Hall Inc. doi/pdf/10.2307/1510517

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

Hammill, D. D., & Larsen, S. C. (2009). *Test of Written Language, Fourth Edition*. Pearson.

Hallquist, M., & Wiley, J. (2011). MplusAutomation: Automating Mplus model estimation and interpretation. R package version 0.6. Available online at: http://cran. r-project. org/web/packages/MplusAutomation/index. html.

Jenkins, J.R., Johnson, E., & Hileman, J. (2004). When reading is also writing: Sources of individual differences on the new reading performance assessments. *Scientific Studies of Reading, 8 (2),* 125-151. Htpps://doi.org/10.1207/s1532799xssr0802_2

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437–447. https://doi.org/10.1037/0022-0663.80.4.437

Juel, C., Griffith, P.L., Gough, P.B. (1996). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243-255.

Kaufman, J.C., Baer, J. & Gentile, C.A. (2011) Differences in gender and ethnicity as measured by rating three writing tasks. *Journal of Creative Behavior*, *38*, 56-69. https://doi.org/10.1002/j.2162-6057.2004.tb01231.x

Kellogg, R. T. (2001). Competition for working memory among writing processes. *The American Journal of Psychology, 114*(2), 175-91. doi:http://dx.doi.org/10.2307/1423513

Kent, S., Wanzek, J., Petscher, Y., Al Otaiba, S. & Kim, Y-S. (2014). Writing fluency and quality in kindergarten and first grade: the role of attention, reading, transcription, and oral language. *Reading and Writing, 27***,** 1163–1188. https://doi.org/10.1007/s11145-013-9480-1

Kim, Y-S. G., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2018). Theorization and an empirical investigation of the component-based and developmental text writing fluency construct. *Journal of Learning Disabilities, 51*(4), p.320-335. doi: 10.1177/0022219417712016.

Kruschke, J. K. (2003). Attention in Learning. *Current Directions in Psychological Science*, *12*(5), 171–175. https://doi.org/10.1111/1467-8721.01254

Lembke, E., Hampton, D., & Hendricker, E. (2013). Data-based decision-making in academics using curriculum-based measurement. In J. W. Lloyd, T. J. Landrum, B. Cook, & M. Tankersley (Eds.), *Research-based Approaches for Assessment* (pp. 18–31). Pearson.

Leifeld, P. (2013). texreg: Conversion of Statistical Model Output in R to LATEX and HTML Tables. *Journal of Statistical Software, 55*(8), 1–24. doi:10.18637/jss.v055.i08

Malecki, C. K., & Jewell, J. (2003). Developmental, sex, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools, 40*, 379–390. doi:10.1002/pits.10096

Marston, D., Deno, S., & Tindal, G. (1983). *A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress*. (Vol. IRLD-RR-126, pp. 29). Institute for Research on Learning Disabilities. doi:10.1177/073724778601100203

Mayes, S.D. & Calhoun, S.L. (2007) Learning, attention, writing, and processing speed in typical children and children with ADHD, autism, anxiety, depression, and oppositional-defiant Ddsorder. *Child Neuropsychology, 13*, 469-493, DOI: 10.1080/09297040601112773

McMaster, K. L. & Campbell, H. (2008). New and existing curriculum-based writing measures: Technical features within and across grades. *School Psychology Review, 37,* 550–566.

McMaster, K. L. & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *Journal of Special Education, 41,* 68–84. doi:10.1177/00224669070410020301

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

Mercer, S. H., Martínez, R. S., Faust, D., & Mitchell, R. R. (2012). Criterion-related validity of curriculum-based measurement in writing with narrative and expository prompts relative to passage copying speed in 10th grade students. *School Psychology Quarterly, 27*(2), 85–95. doi:10.1037/a0029123

Michigan Department of Education (2017). *Michigan Student Test of Educational Progress (MSTEP) Guide to Reports*. Retrieved from https://www.michigan.gov/documents/mde/2017_M-STEP_GTR_598970_7.pdf

Michigan Department of Education. (2020). *Michigan Administrative Rules for Special Education (MARSE) With Related IDEA Federal Regulations.* Retrieved from https://www.michigan.gov/documents/mde/MARSE_Supplemented_with_IDEA_Regs_379598_7.pdf

Monroe, M., & Sherman, E. (1966). *Group Diagnostic Reading Aptitude and Achievement Tests*. Nevins Printing Co.

Montali, J. L., & Lewandowski, L. J. (1996). Bimodal reading: Benefits of a talking computer for average and less skilled readers. *Journal of Learning Disabilities, 29*, 271–279. doi:10.1177/002221949602900305

Muthén, L.K. and Muthén, B.O. (2017). *Mplus User's Guide, Eighth Edition*.

National Assessment Governing Board (2011). Writing Framework for the 2011 National Assessment of Educational Progress. Retrieved from http://www.nagb.org/content/nagb/assets/documents/publications/frameworks/writing-2011.pdf

National Center for Intensive Instruction (2017). *Academic Progress Monitoring Rating Rubric*. Retrieved from https://intensiveintervention.org

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

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards: English Language Arts*.

Olinghouse, N. G. (2008).Student- and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing: An Interdisciplinary Journal, 21*, 3–26.

Olinghouse, N.G. & Leaird, J.T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing: An Interdisciplinary Journal, 22 (*5), 545-565. https://doi.org/10.1007/s11145-008-9124-z

O'Rourke, L., Connelly, V. & Barnett, A. Understanding writing difficulties through a model of the cognitive processes involved in writing. In B. Miller, P. McCardle & V. Connelly (Eds.), *Writing development in struggling learners: Understanding the needs of writers across the lifecourse*. Koninklijke Brill.

Petscher, Y., & Kim, Y-S. G. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology, 49*, 107–129. doi:10.1016/j.jsp.2010.09.004.

Pitcher, T. M., Piek, J. P., & Hay, D. A. (2003). Fine and gross motor ability in males with ADHD. *Developmental Medicine and Child Neurology, 45*, 525–535.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163. doi:10.2307/271063

Ray, A. B., Graham, S., Houston, J. D., & Harris, K. R. (2016). Teachers' use of writing to support students' learning in middle school: A national survey in the United States. *Reading & Writing, 29*, 1039-1068. doi: 10.1007/s11145-015-9602-z

Ritchey, K. D., & Coker, D. L. (2013). An investigation of the validity and utility of two curriculum-based measurement writing tasks. *Reading & Writing Quarterly, 29*(1), 89–119. doi:10.1080/10573569.2013.741957

Ritchey, K. D., McMaster, K. L., Al Otaiba, S., Puranik, C. S., Kim, Y. S. G., Parker, D. C., & Ortiz, M. (2015). Indicators of fluent writing in beginning writers. In K. D. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications* (pp. 21–66). Springer.

Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-analysis of criterion validity for curriculum-based measurement in written language. *The Journal of Special Education, 51*, 72–82. doi: 10.1177/0022466916670637.

Santi, K. L., Barr, C., Khalaf, S., & Francis, D. J. (2015). Different approaches to equating oral reading fluency passages. In K.D. Cummings & Y. Petscher (Eds.), *The Fluency Construct* (p.223–265). Springer. doi:10.1007/978-1-4939-2803-3_9

Satorra, A. & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. Psychometrika 75: 243. doi:10.1007/s11336-009-9135-y

Schultz, K., & Fecho, B. (2000). Society's child: Social context and writing development. *Educational Psychologist, 35*, 51–62.

Smarter Balanced Assessment Consortium (2014). *Performance task writing rubrics*. Retrieved from www.smarterbalanced.org

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.

Swanson, H.L. and Berninger, V.W. (1995). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology, 63,* 358-385. https://doi.org/10.1006/jecp.1996.0054

Troia, G. A. (2018, August). *Relations between teacher pedagogical content knowledge and student writing outcomes*. Research paper given at the biennial meeting of the Special Interest Group on Writing of the European Association for Research on Learning and Instruction.

Troia, G. A., Olinghouse, N. G., Mo, Y., Hawkins, L., Kopke, R. A., … & Stewart, K. A. (2015). Academic standards for writing: To what degree do standards signpost evidence-based instructional practices and interventions? *The Elementary School Journal, 116* (2), 291–321. doi:10.1086/683984

Truckenmiller, A. J., Eckert, T. L., Codding, R. S., & Petscher, Y. (2014). Evaluating the influence of feedback on elementary aged students' fluency growth in written expression: A randomized controlled trial. *Journal of School Psychology, 52*(6), 531-548. doi: 10.1016/j.jsp.2014.09.001

Truckenmiller, A. J., McKindles, J. V., Petscher, Y., Eckert, T. L., & Tock, J. L. (2020). Expanding curriculum-based measurement in written expression for middle school. *Journal of Special Education, 54*, 133-145. doi:10.1177/0022466919887

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.

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

Weissenburger, J. W. & Espin, C. A. (2005). Curriculum-based measures of writing across grade levels. *Journal of School Psychology, 43,* 153–169. doi:10.1016/j.jsp.2005.03.002

Whitaker, D., Berninger, V., Johnston, J. & Swanson, H.L. (1994). Intraindividual differences in levels of language in intermediate grade writers: Implications for the translating process. *Learning and Individual Differences, 6* 107-130. https://doi.org/10.1016/1041-6080(94)90016-7