2015

The Effect of a Metalinguistic Approach to Sentence Combining on Written Expression in Eighth Grade Science for Students who Struggle with Literacy

Lynne Telesca
University of Central Florida

THE EFFECT OF A METALINGUISTIC APPROACH TO SENTENCE COMBINING ON WRITTEN EXPRESSION IN EIGHTH GRADE SCIENCE FOR STUDENTS WHO STRUGGLE WITH LITERACY

by

LYNNE TELESCA
B.S. State University of New York at Geneseo, 1993
M.S. Gallaudet University, 1995

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the College of Education and Human Performance
at the University of Central Florida
Orlando, Florida

Summer Term
2015

Major Professor: Barbara J. Ehren

ABSTRACT

Recent data indicate that less than 50% of American secondary students are able to demonstrate grade-level proficiency in reading, writing, and science (National Center for Educational Statistics [NCES], 2007, 2011, 2012a, 2012b). Secondary students' are expected to develop advanced literacy skills, especially in writing, in order to be ready for college and careers. Students are expected to develop these advanced literacy skills, within all academic subjects. In other words, they are expected to develop disciplinary literacy skills. The statistics are alarming overall, but they are particularly alarming in the area of science. Students need strong literacy skills, including written expression, to be prepared for employment opportunities in science fields, which currently are being filled by graduates of other industrialized nations, who have a more advanced skill set. This loss of occupational opportunity poses a threat for the U.S. to remain globally competitive in science innovation and advancement, which ultimately secures economic prosperity. Despite these staggering concerns, there is little research conducted to evaluate effective instructional methods to develop complex writing skills in academic disciplines such as science.

To address this critical issue, the present study examined the effects of a metalinguistic approach to the writing intervention of sentence combining with eighth-grade students who struggle with literacy. The researcher conducted the study in a typical science classroom in an urban American school setting. The focus of the intervention was to increase students' metalinguistic awareness of science text, to improve written sentence complexity in science, as well as the written expression and determination of comparison and contrast of science content. The study employed a quasi-experimental design. The participants consisted of an experimental group (two classes) who received the treatment during typical science instruction and a comparison group (three classes) who did not receive treatment, but participated in their typical science instruction. There were four participating teachers and 84 participating students. The researcher conducted the study over a period of seven weeks within regularly scheduled science classes. Twenty intervention sessions were conducted for a length of 20 minutes each, totaling 400 minutes or 6.6 hours.

Hierarchical repeated measures ANOVA and hierarchical repeated measures MANOVA analyses revealed that the experimental group performed significantly better than the comparison group on their ability to determine similarities and differences (compare and contrast) related to science content, with a medium effect. The experimental group achieved a slightly higher marginal mean over the comparison group on their ability to combine sentences, with a small effect. Multiple statistical analyses revealed a trend of higher marginal means in favor of the experimental group over the comparison group on several measures of written sentence complexity on both the science compare and contrast writing prompt (small-medium effect) and the science expository essay (medium to large effect). One experimental class also demonstrated higher scores in their overall sentence correctness on science expository essay as compared to all the other classes.

These findings suggest that sentence combining, utilizing a metalinguistic approach, may hold promise as an effective writing intervention in a content area classroom, for secondary students who struggle with literacy. Furthermore, the findings suggest that a metalinguistic approach to sentence combining can be successfully embedded within a content area class, which may result in increased concept knowledge and writing skills in that academic discipline. Implications for practice and future research directions are discussed.

To my husband John,

"I'm on the edge of glory; I'm on the edge with you."

ACKNOWLEDGMENTS

This doctoral degree and dissertation would not have been possible without the contributions of many people. There are many who deserve my thanks and recognition. First, thank you to Lake Speech and Language Center, especially Elizabeth Lobaina and William Pavone. I would never have been able to work part time and complete my coursework without your flexibility and support. I will always be grateful. To my friends, aunts, uncles, and cousins in Florida and back home: thank you for your encouragement and love during every step of this journey. Thank you to my dear friend Melissa Mayor for her support and help during this project. I will miss you. Thank you to my godmother Lynne who continues to remind me how precious life is and to embrace each moment. Thank you to my mom and dad for their unconditional love and unwavering confidence in me. To my sister Michelle, you are truly my Anna when I am having my Elsa moments. You reminded me never to let anyone dictate who I am or what my career aspirations should be. I love you.

I would like to thank all those at or affiliated with the University of Central Florida who have guided and supported me throughout this degree. I wanted to thank two professors outside my department, Dr. Lisa Dieker, who taught me the importance of the history behind my focal areas and helped me to form my mission statement, and Dr. Trey Vasquez, who helped me with my pilot study, which set the groundwork for this study. In my department, I was in a cohort of only myself, but I never felt alone. Thank you to Ruth, Erika, Pam, Mary, and Joanna for being my cheerleaders, as well as editing my work in a pinch. To my friend, Dr. Karen C. Davis, thank you for guiding me and telling me I could do this when I really needed it the most. A huge thank you goes out to my friend and peer mentor, Dr. Melissa Malani. Thank you for meeting with me at Panera before I started this program, sharing your work, giving advice, allowing me to tag along with you at the ASHA Schools Conference, and always listening and lending a hand. Thank you for everything; you are truly amazing. I would also like to thank special folks from the UCF clinic who guided and supported me as a Scholar in Residence specifically, Tom Ehren, Janet Whiteside, and Debra Knox. In addition, I want to express my gratitude to all my special friends at the Willow Schools, formerly the Bridges Academy. First, thank you to Jackie Egli for agreeing to partner and foster the first Scholar in Residence program. Second, to all the wonderful staff and teachers with whom I collaborated, especially Alicia, Mary, Laurie, Pam, Heather, Heidi, Erica, and Michael. I will never forget all your kindness and willingness to help. Last, to my great friend Principal Carla Brandt, the three years with you restored my faith that there are school administrators out there who truly care about their staff, students, and families. Working with you was a magical gift that I will cherish always.

I would like to thank those who willingly gave of their time to help me complete this study. To my friend Linda Rosa-Lugo, thank you for helping me assemble my research team and for looking out for me for the last four years, after nudging me to pursue this degree many years ago. To my amazing research testing team: Dialy's Reyes-Greene, Andrea Yang, Devyn Squires, Yaritza Reyes, and Nicole Johnson, thank you for your commitment and enthusiasm for this project. To my incredible research assistants Sasha Perez and Brittany Hines, I would never have made it through all the data analysis without the two of you. Thank you for your time, dedication, and hard work. To the administration, teachers, and students at the school where I completed this study, thank you for your flexibility and trust in me during such an incredibly busy and stressful time of the school year. It was an honor to work with you all. I hope that the information gained from this study will be benefit the school, teachers, and students. To each of my students, remember you are the “WRITEr.”, keep on writing.

To my committee, it has been a privilege to work with you all. Dr. Anthony Kong, I respect you so much not only as my professor, but also as an innovative researcher. Thank you for supporting my work and providing your unique perspective. Dr. Vicky Zygouris-Coe, thank you for introducing me to the world of literacies. You have always treated me as valued literacy professional and made me feel that I made a special contribution to literacy as a speech-language pathologist. I could not think anyone else that I would want more to be here with me at the end of my degree. To Dr. Debbie Hahs-Vaughn, I would never have made it through the statistical analyses without you. Thank you for your time, patience, guidance, and wisdom. To my committee chair, Dr. Barbara Ehren, your confidence and faith in me has been never ending. You have pushed me beyond a level of expertise that I thought was capable. I am a better therapist, professional, researcher, and future scholar because of you. Thank you for trusting me with the Scholar in Residence program and affording me that invaluable opportunity. Thank you for your bloody comb, which has improved my writing above and beyond my expectations. Thank you for your guidance through the land of adolescent and disciplinary literacy. It is a land that I will continue to visit and advocate for in my future career. Thank you for inducting me officially into the “Syntax is not for Sissies” club. Last, thank you for allowing me to flourish in a place where I could feel good again about being so passionate about our field. It is has been a great honor and blessing to be under your tutelage, and I will miss you greatly.

Finally, I would like to give special thanks to the three closest to my heart. Thank you to the best boy in the world, Cash. You have brought me such joy either by my side or at my feet. To my son, Jonah Martin, Mr. Goo, I want you to know that no matter what, being your mom and watching you grow into such an amazing, talented, funny, and sensitive young man has given me the greatest happiness and pride, more than any educational degree ever could. I am sorry for all the things I missed and all the times I needed to work instead of spending time with you. I hope that when you are older and think of this time you will remember two things. First, do not ever let anyone try to tell you who you are. Second, that with hard work and dedication, you can be anything that you want to be. Finally, to my husband and love of my life John, there are no words to express my gratitude. It has been a long, great sacrifice. You were willing to put our life on hold to allow me to achieve my dream. I would have never accomplished this without you, and I am everything I am because you love me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

This study explored the effects of a metalinguistic approach to the writing intervention of sentence combining with eighth-grade students who struggle with literacy. The focus of the intervention was to increase students' metalinguistic awareness in science to improve written sentence complexity in science, as well as the written expression and determination of comparison and contrast concepts of science content. The researcher delivered the treatment during regularly scheduled science classes. This chapter presents the current problem, purpose of the study, theoretical framework, research questions/hypotheses, significance of the study, limitations and delimitations, assumptions, and operational definitions.

Statement of the Problem

In the United States of America, a major concern in educational reform is that students who graduate from high school do not acquire the skills they will need to be ready for career, college, or life (Bill & Melinda Gates Foundation, 2008; National Governors Association [NGA] Center for Best Practices & Council of Chief State School Officers [CCSSO], 2010b). Despite past reform efforts, U.S. students have achieved minimal gains in reading, writing, math, and science, when compared to students of other nations (Martin, Mullis, Foy, & Stanco, 2012; Organisation for Economic Co-operation and Development [OECD], 2014). Because U.S. students do not have the skills to meet current workforce demands (Jellinek, 2012; Kena et al., 2014), employers in the fields of science, technology, engineering, and mathematics (STEM) are hiring students from other industrialized nations for job positions with greater salaries and advancement opportunities (Chen & Weko, 2009; STEM Education Coalition, 2013). The result of hiring students from outside the U.S. is a loss of revenue, innovation, and advancement, which threatens the ability of the U.S. to compete in the global economy (U.S. Department of Education 2014b; White House Office of Science and Technology, 2014).

Two significant educational reform organizations have addressed concerns about student college and career readiness. The first, the Partnership for 21st Century Skills (P21), founded in 2002, was comprised of the business community, educational leaders, and policymakers who defined and positioned 21st century skills at the forefront of America's kindergarten through twelfth (K-12) grade educational agenda. According to P21 (2014), 21st century skills are the skills students need to develop in order to be ready for college or careers in our rapidly changing technological society. At the core of 21st century skills are the use of the three R's, reading, writing, and arithmetic to develop critical thinking skills. Teachers target the three R's within core instruction of English, reading or language arts, world languages, arts, mathematics, economics, science, geography, history, government, and civics. Other areas interconnected with core-subject instruction are:

- interdisciplinary themes of global awareness, and the literacy of: finance, economics, business, entrepreneurialism, civics, health and the environment;

- instructional and innovative skills of: critical thinking, problem solving, communication, collaboration, innovation, and creativity;

- information, media, and technology skills;

- life and career skills of: flexibility, adaptability, initiative, self-direction, social and cross-cultural skills, productivity, accountability, leadership, and responsibility (P21, 2011).

The second group, the National Governors Association (NGA) Center for Best Practices and Council of Chief State School Officers (CCSSO), created the Common Core State Standards (CCSS, 2010). The authors created the standards as a means of establishing nationally shared expectations and focus (Porter, McMaken, Hwang, & Yang, 2011). The standards target the broad categories of Mathematics and English Language Arts (ELA). The main goal of the ELA standards is that students will be "college and career ready in literacy no later than the end of high school" (NGA Center for Best Practices and CCSSO, 2010c, p. 3). Instruction within the CCSS employs an integrated model of literacy, specifically the development of listening, speaking, reading, and writing skills. Listening, speaking, reading, and writing develop simultaneously through student engagement with narrative and expository information that increases in complexity over time. An additional area of focus in the ELA standards for secondary students (grades 6-12) is literacy in History/Social Studies, Science, and Technical subjects. These standards address the need for ongoing literacy instruction to develop the advanced or high literacy skills that are unique and specific to each academic discipline (Heller & Greenleaf, 2007).

Unfortunately, a significant number of secondary students who graduate from high school neither possess advanced literacy skills, nor possess basic literacy skills (Achieve Inc., 2005, American College Testing [ACT], Inc. 2000, 2005, 2007; Graham & Perin, 2007b; National Commission on Writing for America's Families, Schools, and Colleges, 2006). According to the 2013 National Assessment of Educational Progress (NAEP) in reading, only 36% of eighth graders and 38% of 12th graders were at or above proficiency, with no change in 12th grade performance since the 2009 assessment (NCES, 2011, 2014b). In the area of writing, the 2011 NAEP reported that only 27% of both eighth and 12th graders performed at or above the proficient level (NCES, 2012b). The 27% was a decrease in score for eighth grade from the 2007 NAEP writing assessment in which 34% of eighth graders were at or above proficiency. For 12th graders, the 27% indicated only a slight increase in score from 25% in 2007 (Salahu-Din, Persky, & Miller, 2008). Results of international assessments in reading literacy indicated that U.S. students' performance lagged behind other industrialized nations. According to the 2012 Program for International Student Assessment (PISA), 15-year- olds in the U.S. were ranked 23rd when compared to other nations in reading literacy (Kelly et al., 2013). Two years prior, U.S. students had ranked 14th in reading literacy on the PISA (Fleischman, Hopstock, Pelczar, & Shelley, 2010). In summary, our national assessments reveal that two-thirds or more of secondary students are below a proficient level in reading and writing. International assessments reveal that U.S. student reading performance has significantly dropped in the last two years. Consequently, the concern remains justified that a large percentage of secondary students not only lack the advanced literacy skills they need for college and career success, but may lose future college or career opportunities to students from other nations with stronger advanced literacy skills .

Like reading and writing achievement, secondary U.S. students are not excelling in achievement in science, which will hinder their ability to compete in the evolving global marketplace. Poor science achievement results in the loss of college and career opportunities for U.S. students in science and STEM related occupations. This loss of opportunity has been the impetus for the U.S. government's distress that the U.S. is not maintaining its competitive position in the global economy. To combat these concerns, President Barack Obama has deemed student proficiency in STEM fields an educational priority (U.S. Department of Education [DOE], 2014a). STEM proficiency is a priority because STEM occupations yield high revenue and impel innovation (Chen & Weko, 2009; White House Office of Science and Technology, 2014). Yet, even with STEM educational initiatives and funding, student performance in science is still waning. First, the 2011 NAEP in science revealed that only 34% of eighth-grade students earned at or above a proficient rating (NCES, 2012a). Second, on the 2012 PISA, the U.S. ranked 17th in science literacy (Kelly et al., 2013), which was a decrease compared to a ranking of 13th in 2009 (Fleischman et al., 2010). Last, the 2011 Trends in International Math and Science Study (TIMSS) revealed that eighth-grade students in the U.S. ranked 13th based on average science score and 13^{th} on the percentage of students who achieved at or above the advanced rating (Martin et al., 2012)

Although there have been significant attempts towards educational reform, secondary students' poor achievement in reading, writing, and science are not the only concerns. Secondary students are also expected to master the literacy of academic subjects, or disciplinary literacy. Specifically, the CCSS have established standards at the secondary level specifically for "Literacy in History/Social Studies, Science, and Technical Subjects" (NGA Center for Best Practices and the CCSSO, 2010c). Disciplinary literacy is the ability to develop understanding and to construct meaning of discipline through the specific literacy practices of the discipline (i.e., listening, speaking, reading, writing), as well as to decipher how disciplinary differences are socially constructed (Moje, 2008). Disciplinary literacy is also considered the apex of literacy development superseding basic and intermediate literacy skills (Shanahan & Shanahan, 2008). If student achievement in reading, writing, and an academic content area such as science is poor (NCES, 2013, 2014a, 2014b), then the chance that students will be able to develop disciplinary literacy skills is unlikely. The current pool of research-validated instructional practices to help reverse this negative trend and help develop disciplinary literacy skills is scarce (Brozo, Moorman, Meyer & Stewart, 2013; Ehren, Murza, & Malani, 2012). Consequently, the need for research on effective reading and writing instructional practices within different academic disciplines remains imperative (Faggella-Luby, Graner, Deshler, & Drew, 2012). With the national focus placed on the importance of STEM Education, research specific to disciplinary literacy in the area of science is even more of a priority.

Purpose of the Study

The primary purpose of this study was to examine whether a research-based writing intervention, sentence combining, implemented with a metalinguistic approach, was effective in improving the acquisition of knowledge and written expression for comparison and contrast in science for eighth-grade students who struggle with literacy. This study used a writing intervention in science, not only to develop literacy skills in science, but also to increase the linguistic knowledge specific to comparison and contrast of science content. As the writing demands continue to increase to meet the disciplinary literacy demands of the CCSS, there is a need for an empirical research base to inform and validate effective writing interventions in the disciplines (Harris & Graham, 2013; Mo, Kopke, Hawkins, Troia, & Olinghouse, 2014; Troia & Olinghouse, 2013). There is an even greater need for writing interventions that will be effective for students who struggle with literacy, but must meet the same writing expectations set for all students (Graham & Harris, 2013; Saddler & Asaro-Saddler, 2013). Researchers have documented the effectiveness of sentence combining on writing performance for students who struggle with literacy (e.g., Hillocks, 1986; Hunt 1965; Graham & Perin, 2007a; Saddler & Graham, 2005). However, there is a paucity of research related to sentence combining with academic text.

Theoretical Framework

The use of metalinguistic sentence combining as the core intervention was rooted in a few theoretical frameworks. First, the use of a metalinguistic approach is an integral part of sentence-combining instruction (Saddler, 2013). Sentence combining allows students, with teacher modeling, to explicitly and overtly think, discuss, and manipulate language patterns that would be inherent in the writing process, and the language intentions and implications of those patterns (Strong, 1986). However, most sentence-combining research studies do not explicitly outline the metalinguistic approach the teacher or researcher used to provide this modeling. For this study, the researcher implemented an explicit metalinguistic approach that included a metalinguistic script with specific cues and actions. The metalinguistic approach provided opportunities for students to develop their metalinguistic awareness, the development of the conscious awareness of the knowledge and skills related to the structural and content features of language (Tunmer, Herriman & Nesdale, 1988). Metalinguistic activities incorporate intentional reflecting, linguistic monitoring, and planning of all aspects of language including, phonology, morphology, syntax, semantics, and pragmatics (Gombert, 1992). According to Cazden (1974), metalinguistic awareness is critical for the development of reading and writing at the secondary school level. Metalinguistic awareness allows students to recognize and correct their linguistic errors (Paul & Norbury, 2012). Hence, the ongoing development of metalinguistic awareness is essential to continue writing development, especially for students who struggle with literacy (Perin, 2013; Shanahan, 2009). The use of metalanguage, or the explicit language to discuss the metalinguistic aspects of the language, is essential in the development of metalinguistic awareness (Basturkmen, Loewen, & Ellis, 2002).

Second, according to Noam Chomsky's theory of transformational-generative grammar (1965), there are two structures of the syntactic base of language. The first is the deep structure or competence, which consists of a series of innate phrase structure rules that generate sentences based on the speaker's actual meaning or intent. Chomsky believed that the deep structure is converted using transformational rules to produce a spoken or written sentence. This sentence is the surface structure, which is a concrete production or application of a speaker's competence. Chomsky's theory was influential in linguistic theory proposing the idea that a sentence was the combination of several kernel sentences or basic sentence patterns. This combining of sentences requires cognitive reordering for creating efficient sentences (Phelps-Gunn & Phelps-Terasaki, 1982). Sentence-combining instruction is a means of developing the use of transformational rules that increase sentence complexity (Strong, 1986). Although other research studies before 1900 used sentence-combining practices, sentence combining became a legitimate practice due to Chomsky's (1965) transformational-grammar theory (Connors, 2000). Chomsky's theory legitimized the belief that sentence-combining taps into the mind's natural process of combining kernel sentences to create many sentence possibilities that are not just innate (Saddler, 2013). This creation of sentences through the combining of kernel sentences was also believed to improve writing ability (Hunt, 1965).

Third, sentence-combining is a skill set that allows students to learn across the reciprocal processes of listening, speaking, reading, and writing, Oral language is the foundation for writing; listening skills are the foundation for reading (Anderson et al., 1985; Fromkin, Rodman & Hyams, 2014; Owens, 2008; Pence-Turnbull & Justice, 2012). There are numerous research studies that support the theory that each of the reciprocal processes contributes to the development of the other (e.g., Berninger, 2000; Berninger & Abbot, 2010; Catts, Fey, Tomblin, & Zhang, 2002; Graham & Herbert, 2010; Kamhi & Catts, 2012; Loban, 1976; Shanahan, 2006; Storch & Whitehurst, 2002). Accordingly, researchers have suggested that instructors teach reading and writing together (Applebee, 1977; Shanahan, 2009; Scott, 2012; Westby, 2012). Moreover, sentence combining may improve the understanding of academic content through listening, speaking, reading, and then writing about academic content (Bangert-Drowns, Hurley & Wilkinson, 2004; McLeod, Miraglia, Soven, & Thaiss, 2001; NGA Center for Best Practice and CCSSO, 2010a; Rijlaarsdam et al., 2014).

Fourth, sentence combining is also aligned with Halliday's language-based learning theory (1993) and Halliday's advanced literacy or disciplinary stage (2004). Halliday's language-based learning theory (1993) proposed that students learn through language, specifically, through the spoken and written registers. Sentence combining instruction uses both spoken and written sentences related to academic content to facilitate the learning of the linguistic aspects of academic content, through the manipulation of the language. In Halliday's (2004) advanced literacy or disciplinary stage (ages 9 years through 18 years) students have to interact with academic text that is abstract and complex. In order to comprehend and communicate with academic text, students need to become familiar with the grammar of each discipline (Fang, 2012a). Instruction such as sentence combining can help students engage and manipulate text according to the sentence constructions of that discipline.

The methodological structure of the intervention was based on three theoretical frameworks; namely, the zone of proximal development (ZPD, Vygotsky, 1978), the use of explicit instruction (Archer & Hughes, 2004; Rosenshine, 1986), and the use of a gradual release model (Pearson & Gallagher, 1983). The instruction incorporated the use of explicit instruction and gradual release as a means of facilitating students' skill development within their ZPD. A student's ZPD is the distance between their developmental level and their level of potential (Vygotsky, 1978). In order to achieve the ZPD, the instruction must be tailored to support the completion of written activities that students would struggle with independently. In this study, the researcher provided support with explicit instruction. Students who struggle with literacy require explicit instruction in specific areas language to improve their writing (Datchuk & Kubina, 2013; Olinghouse, Graham, & Harris, 2010). Explicit instruction incorporates the review of previous learning, presentation and practice of new material, detailed instruction, repetition, explanation, and systematic feedback (Rosenshine, 1986). Explicit instruction also provides a direct approach to instruction with supports or scaffolds that result in independent mastery (Archer & Hughes, 2004). The supports or scaffolding provided in this research study followed a gradual release model (Pearson & Gallagher, 1983). The researcher used a protocol that systematically decreased the amount of metalinguistic verbal cues and overt actions the researcher provided for the students. The final level of each phase was independent practice during which the students would complete sentence combining on their own, without any cues from the researcher or peers. The students were not completely independent, for they still had the aid of a visual support (i.e., mnemonic) to recall the steps of the sentence combining, even at the independent practice stage.

The researcher employed the following research-based practices that promote explicit instruction: (a) opening and reviewing with an advance organizer, (b) incorporating student enlistment, (c) prompting involvement, (d) checking for understanding, (e) correcting and expanding responses, (f) summarizing with a closing organizer, (g) personalizing the intervention, (h) using motivational strategies, and (i) providing feedback (Deshler, Alley, Warner & Schumaker, 1981; Hughes, 2011; Schumaker; 1989). The intervention also incorporated a mnemonic device for the students to remember the five steps needed for metalinguistic sentence combining (Deshler & Lenz, 1989). By providing a visual support such as a mnemonic, the researcher was able to control the cognitive difficulty or processing demands of the task, which in turn can increase student success (Swanson & Deshler, 2003).

Research Questions and Hypotheses

The researcher conducted the study to answer and provide evidence for the following four research questions and hypotheses:

<u>Question One</u>: Do eighth-grade students struggling with literacy who participate in metalinguistic sentence combining (MSC) instruction in science demonstrate an increase in their

sentence-combining ability as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

Hypothesis One: Eighth-grade students struggling with literacy who participate in MSC instruction in science will demonstrate a greater increase in score on the Test of Written Language-4 Sentence Combining Subtest form A/B than eighth-grade students struggling with literacy who participate in typical science instruction alone.

Question Two: Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase in specific aspects of sentence complexity in response to a science compare/contrast writing prompt as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

Hypothesis Two: Eighth-grade students struggling with literacy who participate in MSC instruction in science will demonstrate greater increases in measures of sentence complexity in response to a science compare/contrast writing prompt than eighth-grade students struggling with literacy who participate in typical science instruction alone. Sentence complexity will be measured using the following seven measures: (1) sentence length, (2) sentence connectives, (3) agentless passive voice, (4) number of correct word sequences, (5) number of targeted sentence connectives, (6) correct versus incorrect sequences count (CIWS), and (7) number and type of morpho-syntactical errors.

Question Three: Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase in specific aspects of sentence complexity when writing a science expository essay as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

Hypothesis Three: Eighth-grade students struggling with literacy who participate in MSC instruction in science will demonstrate greater increases in sentence complexity when writing a science expository essay than eighth-grade students struggling with literacy who participate in typical science instruction alone. Sentence complexity will be measured using the following eight measures: (1) sentence connectives, (2) words before the main clause, (3) agentless passive voice, (4) noun phrase density (5) verb phrase density, (6) prepositional phrase density, (7) correct versus incorrect sequences count (CIWS), and (8) number and type of morpho-syntactical errors.

Question Four: Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase their ability to determine similarities and differences (compare/contrast structure) related to science content as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

Hypothesis Four: Eighth-grade students struggling with literacy who participate in MSC instruction in science will demonstrate a greater increase in their ability to determine similarities and differences (compare/contrast structure) related to science content as measured by an increase in score on the compare and contrast double bubble map by Thinking Maps ® than eighth-grade students struggling with literacy who participate in typical science instruction alone.

Significance

This study aimed to add to the empirical data supporting the efficacy of a metalinguistic approach to the writing intervention, sentence combining, to improve knowledge and written expression of science content for students who struggle with literacy. The writing demands for secondary students have increased since implementation of the CCSS. More importantly, these writing demands are expected across all academic disciplines for all students (Graham & Harris, 2013). For secondary students who struggle in literacy and science, which is the majority, writing will be an arduous task. The U.S. proposed budget for 2015 has already allocated significant funding toward STEM education (White House Office of Science and Technology, 2014); therefore, the expectations in students' science literacy will remain high. To meet these expectations, strong written expression in science will be essential. Thus, intervention that targets the improvement of written expression in science is imperative. This study will add to limited empirical research in several areas. First, this study will provide data on effective writing intervention (1) with students struggling with literacy, (2) in an academic discipline, (3) in the discipline of science, and (4) that targets specific linguistic aspects of comparison and contrast in science. Second, this study will add novel empirical data to the existing research base on sentence combining, specifically the use of a metalinguistic approach and use of academic content. The use of metalinguistic approach to improve writing and determining comparison and contrast relationships in science may improve overall academic performance in science. Last, this intervention may also prove beneficial for improving knowledge and the written expression of the linguistic concept of comparison and contrast in other academic disciplines.

Limitations

The following are the possible limitations of the study:

1. The researcher did not select students randomly from the population; hence, the research design may increase threats to internal validity such as history (i.e., events that occur during the time of treatment that could affect the outcome), selection bias, maturation, and statistical regression (Edmonds & Kennedy, 2013).
2. The participating students were residents of Central Florida and may not be representative of students in other geographical areas. This may preclude the researcher's ability to generalize from the experimental sample to a defined population sample (Gall, Gall & Borg, 2006).
3. The participating students were students who struggle with literacy and chosen within the strict parameters set for this definition by the school district from which they came. This may also preclude the researcher's ability to generalize from the experimental sample to a defined population sample (Gall et al., 2006).

4. Although different forms of the TOWL-4 were used to assess sentence-combining skills, the participants' posttest scores may be affected by participating in a pre-test condition using the same instrument, which could threaten internal validity (Campbell & Stanley, 1963).
5. The science teacher was present in the intervention sessions and may inadvertently focus on aspects of the intervention while teaching outside the intervention time.
6. The intervention was conducted by the researcher who is a speech-language pathologist, which may limit the ability to predict if other educational professionals could implement the intervention successfully.

Delimitations

1. Student participants were required to meet the following inclusionary criteria:

(a) be enrolled in eighth grade in the participating school;

(b) be a student in an eighth-grade science class for students who struggle with literacy.

Assumptions

The researcher made the following theoretical assumptions:

1. Students who struggle with literacy have significant delays in their knowledge and use of age-expected syntactical structures, which negatively affects their reading and writing skills.
2. Students who struggle with literacy have limited knowledge of the specific syntactic structures such as conjunctions, passive voice, and embedded and expanded phrases or clauses that are prevalent in the literacy of science.
3. When students increase their metalinguistic skills in a content area, they will make gains in writing and understanding of content-area material.
4. When students practice writing in short intervals on a consistent basis, they will make gains in written sentence complexity and correctness.

Operational Definitions

The following terms were operationally defined for the purposes of the study:

1. *Adverbial clause* is a group of words that contain a subject and a predicate to provide information about time, place, and motivation for an action or a state (Justice and Ezell, 2008).
2. *Agentless passive voice* is oral or written expression that does not include a subjective noun agent (Justice & Ezell, 2008). The incidence of agentless passive voice in the writing samples was measured with Coh-Metrix 3.0 software (McNamara & Graesser, 2012).

3. *Correct Versus Incorrect Word Sequences (CIWS)* is a curriculum-based measure (CBM, Deno, 1985) of sentence-level writing skills. This measure counts the number of correct word sequences, which are defined as two adjacent words that are correctly spelled, capitalized, punctuated, and are grammatically and semantically acceptable within the context of the sentence. The amount of word sequences that are incorrect (i.e., does not qualify as a correct word sequence) are also counted. To calculate the CIWS count, the sum of the incorrect word sequences is subtracted from the sum of the correct word sequences (Breaux & Frey, 2009).
4. *Formative assessments* are measures used to gauge students' learning through informal methods (Garrison, Chandler & Ehringhaus, 2009). Graphic organizers are one method of formative assessment (Dodge, 2009).
5. *Kernel sentences* are simple sentences that do not include more than one clause and are a minimum of a noun and a verb (Andrews et al., 2004; O' Hare, 1973).
6. *Language sample analysis* is the tally and interpretation of the number and type of different language structures used in oral or written expression generated by a student; it includes semantic, morphologic, syntactic, and pragmatic features (Owens, Farinella, & Metz, 2014).
7. *Metalanguage* is the explicit use of words to focus on various aspects of the nature of language, such as the word choice, word structure, and phrase or clause structures that convey meaning and intent (Basturkmen et al., 2002).
8. *Metalinguistic cues* are verbal hints or signals that target the use of metalanguage to call attention to the language structures of phonology, morphology, syntax, semantics, and

pragmatics (Gombert, 1992). Metalinguistic cues are separated specifically into the subcategories of metalanguage: metaphonology, metamorphology, metasyntax, metasemantics, and metapragmatics (Paul & Norbury, 2012; Roth, Spence, Cooper, De La Paz, 1996).

9. *Metalinguistic sentence combining (MSC)* is the process of manipulating, merging, and rewriting sentences using explicit instruction and metalanguage to bring conscious awareness of the underlying linguistic knowledge needed to create those sentences (Saddler, 2013; Strong, 1986).
10. *Metamorphology* is the ability to think and talk about morphemes (i.e., derivational and inflectional morphemes), and their effect on meaning and grammaticality (Roth et al., 1996; Westby, 2004).
11. *Metaphonology* is the ability to understand and talk about how words are made of sounds represented by symbols or letters that affect meaning (Paul & Norbury, 2012, Roth et al. 1996).
12. *Metapragmatics* is the ability to think and talk about appropriate use of language in social situations, contexts, and schemas (Gombert, 1993; Paul & Norbury, 2012; Westby, 2004). In relation to writing, metapragmatics is the discussion of how intent or genre may dictate the types of words, structure, and conventions needed (Troia, 2012).
13. *Metasemantics* is the ability to think and talk about the manipulation of meaning of words and sentences (Gombert, 1993; Roth et al., 1996) by analyzing vocabulary and word relationships (e.g., categories, parts, synonyms, antonyms, Westby, 2004).

14. *Metasyntax* is the ability to think and talk about grammatical categories (i.e., parts of speech), syntactic structures (i.e., cohesive devices, phrases, clauses), and word order that affect meaning and grammaticality, which are unique in each genre and discipline (Gombert, 1993; Roth et al., 1996; Westby, 2004).
15. *Noun phrase density* is a measurement to determine the use of a group of words that has a noun or a pronoun at its head (Justice & Ezell, 2008). Noun phrase density was measured in the writing samples as an incidence score calculated with Coh-Metrix 3.0 software (McNamara et al., 2014).
16. *Phrase* refers to a syntactic structure consisting of one main word and one or more words grouped around it (Justice & Ezell, 2008). The theoretical model of a sentence in Coh-Metrix 3.0 is morpheme groups or phrases. Syntactic difficulty increases as the number of phrases increases.
17. *Prepositional phrase density* is a measurement to determine the use of a group of two or more words that begins with a word that "links a noun or pronoun to another sentence element by expressing direction, location, time, or figurative location" (p. 189, Justice & Ezell, 2008). Prepositional phrase density was measured in the writing samples as an incidence score calculated with Coh-Metrix 3.0 software (McNamara et al., 2014).
18. *Relative clause* is a noun and a predicate that modifies an independent clause. It is a dependent clause that cannot stand alone (Justice & Ezell, 2008).
19. *Sentence connectives* are cohesive links between ideas and clauses that provide clues about text organization. The sentence connectives were measured with Coh-Metrix 3.0 software

are: (1) causal (e.g., because, so); (2) logical (e.g., and, or); (3) adversative (e.g., although, whereas); (4) temporal (e.g., first, until); and (5) additive (e.g. and, moreover; McNamara et al., 2014).

20. *Sentence length* is the number of words in a sentence. Sentence length was measured with Coh-Metrix 3.0 software (McNamara et al., 2014).

21. *A student who struggles with literacy* is defined as an adolescent who has scored below a minimum achievement level score of 3, or below a developmental score of 228 (171-227) on the 2014 Florida Comprehensive Assessment Test (FCAT) in reading. Students who are below a level 3 are believed to demonstrate a below satisfactory level (achievement level of 2, developmental score range of 213-227) to inadequate (achievement level of 1, developmental score range of 171-212) satisfactory level of success with the challenging content of the Next Generation Sunshine State Standards (Florida Department of Education, 2014).

22. *Verb phrase density* is a measurement to determine the use of a group of words that have a main verb or clause with any attached auxiliary forms or modifiers (Justice & Ezell, 2008). Verb phrase density was measured in the writing samples as an incidence score measured with Coh-Metrix 3.0 software (McNamara et al., 2014).

23. *Words before the main clause* refers to the mean number of words before the main verb of the main clause in sentences as measured with Coh-Metrix 3.0 software. An increase in the number of words before the main clause signifies an increase in the complexity and cognitive

load of a sentence. Another term for the use of words before a main clause is "left-embeddedness" (McNamara et al., 2014).

Chapter Summary

This chapter presented a problem statement, purpose of the study, theoretical framework, research questions/hypotheses, significance of the study, limitations and delimitations, assumptions, and operational definitions. This study examined the effect of a metalinguistic writing intervention delivered during regularly scheduled science classes with eighth-grade students who struggle with literacy. Specific research questions were posed to investigate gains in sentence combining, written sentence complexity of a science comparison and contrast prompt and science expository essay, as well as the determination of comparison and contrast relationships using science content.

CHAPTER TWO: LITERATURE REVIEW

This chapter presents the rationale for metalinguistic sentence combining in science as a writing intervention for eighth-grade students who struggle with literacy. The review begins with three major topics relevant to the significance of the study: workforce literacy, literacy for college and career readiness, and scientific literacy. After that, the review presents pertinent research literature that contributed to the theoretical framework of the intervention. First, the review discusses literature in the area of syntax, as well as syntax in relation to the reciprocal language processes and disciplinary literacy. Then, because one goal of this study is to explore the impact of a writing intervention on comprehension of linguistic concepts related to science content, the review reports research on the relationship among writing, reading, learning, and academic content. Next, the review documents metalinguistic skills and awareness, along with literature on writing within a metalinguistic framework. The chapter concludes with a review of the evidence base of this study's intervention, sentence combining.

Workforce Literacy

This section begins with an historical overview of the definition of workforce literacy. The research intervention targeted the improvement of writing skills that are not only required in school, but that are also required in the workforce. Therefore, this section continues with discussion about the writing demands in today's workforce and concludes with data that documents the current writing crisis in today's workforce that supports the need for an intervention such as the one designed in this study.

History and Definition

Johnston and Packer (1987) conducted research to inform future federal policies related to workforce development or what they described as, "Workforce 2000." The goal was to predict what workforce skills would be necessary for the U.S. to remain economically competitive. The report predicted that the workforce would grow and change significantly, increasing the number of older workers, female workers, minority workers, and disadvantaged workers. What would also increase is the expected skill set required of these workers. In order to combat these changes and maintain productivity, Johnston and Packer listed obtaining higher education and acquiring higher educational skills as major issues to be addressed in order to promote further economic prosperity. However, the issues of higher education and its acquired skills also meant the need for higher literacy and numeracy skills as well. The need for higher levels of literacy and numeracy was due to the fast growing job market in high-performing environments that demanded advanced language, math, and reading skills to enhance productivity (Levin, 1994). This report was the catalyst for the revision of government initiatives that supported and defined what workforce literacy means today.

One government initiative was the National Literacy Act (1991), an amendment to the original Adult Education Act of 1966 and Economic Opportunity Act of 1964. This initiative established a number of programs to ensure that by the year 2000, all adults would be literate and "possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship" (Irwin, 1991). Congress determined that illiteracy was intergenerational and closely associated with the lack of supply of workers needed for skilled labor positions (National Literacy Act, 1991). Congress sought to equalize the economically disadvantaged by funding programs that would provide instruction in English to adult workers whose inability to read, write, or do arithmetic impacted their ability to obtain employment (United States Department of Education, 1991). The 1991 National Literacy Act further clarified the definition of literacy as meaning:

> "an individual's ability to read, write, and speak in English, and compute and solve problems at levels of proficiency necessary to function on the job and in society, to achieve one's goals, and develop one's knowledge and potential" (p. 1)

The National Literacy Act (1991) described workforce literacy as the basic skills training needed to develop literacy including (a) English as a second language instruction, (b) communication skill building, (c) interpersonal skill building, (d) reading and writing skill building, and (e) computation and problem solving. However, these areas referred to skill building for those who were unemployed or already underprepared in the workforce, not the specific skills required for workforce readiness for those trying to obtain employment in a high-performing environment.

The government defined the workforce readiness skills needed for a high-performing work environment through the work of the Secretary's Commission on Achieving Necessary Skills (SCANS; Kane, Berryman, Goslin, & Meltzer, 1990). The SCANS (1990) determined and documented the skills and knowledge needed to be successful in different work scenarios. The commission deduced that the workforce requirements had increased to sustain a fast-paced technological society (Kerry et al., 1990). The SCANS (1991) released its results and separated workforce skills into three foundational skills and five competencies. The foundational skills included: (1) basic skills- reading, writing, arithmetic, mathematics, speaking and listening, (2) thinking skills, and (3) personal qualities (e.g., responsibility, integrity). The five competencies listed were productive use of (1) resources, (2) interpersonal skills, (3) information, (4) systems, and (5) technology. SCANS continued to collect data and modify its recommendations until the end of the 20th century.

As the 21st century approached, the need for stronger workforce skills continued to grow with urgency. To remain competitive, employers in the 21st century sought better, faster, cost-effective products and services (Business Roundtable, 1999). The Center for Workforce Preparation (CWP, 2002), announced that 21st century employers required exemplary workforce skills upon hire. The traditional practice of hiring entry-level workers who would develop necessary skills over time was foregone. In the 21st century, entry-level workers would have more responsibilities than ever before, requiring advanced educational credentials, specialized training, and established high literacy skills (Comings, Reder, & Sum, 2001). These high literacy skills included the ability to write effectively (CWP, 2002).

Writing Demands in the Workforce

Each revision of the SCANS resulted in a more detailed and complex definition of the required workforce writing skills. The SCANS (1991) specified that most jobs would require the ability to request, explain, illustrate, or convince through written correspondence, instructions, charts, graphs, or proposals. The SCANS (1991) further indicated that workers needed to be able apply their writing skills in contextual and relevant problem solving situations. In 1992, the SCANS published a second document based on their field research that specified that writing in the workplace should be for the purposes of informing, persuading, and clarifying. According to SCANS (1992), workplace writing would be for a range of audiences, in a variety of formats, with neatness and grammaticality. After continued research through job analyses, the final SCANS (1999) report noted that workforce writing:

> communicates thoughts, ideas, information, and messages in writing; records information completely and accurately; composes and creates documents such as letters, directions, manuals, reports, proposals, graphs, flow-charts; uses language, style, organization, and format appropriate to the subject matter, purpose, and audience; includes supporting documentation and attends to level of detail; and checks, edits, and revises for correct information, appropriate emphasis, form, grammar, spelling, and punctuation (p. 2).

This definition had incorporated previous SCANS components, as well as emphasized the language structure, form, and mechanics that were expected in the workforce. The SCANS (1999) definition of writing as a workplace skill was standardized by the American College Testing Incorporated (ACT, Inc., 2000) based on its documented commonalities across 35 other governmental and international writing frameworks, including the Occupational Informational Network (O*NET). Currently, O*NET provides a system to characterize and provide the requirements of each occupation in the United States (US Department of Labor, n.d.). ACT, Inc. (2000) also developed a five point behaviorally anchored scale for writing. At the lowest level, level 1, the writing expectation would be to record or copy simple phrases or lists of words to communicate information. At the middle level, level 3, the writing expectation would be to use appropriate vocabulary, style, and tone to compose workplace documents, reports, or essays. At the highest level, level 5, the writing expectation would be to create documents, articles, proposals, or presentations that synthesize and compare/contrast complex information. The last level is even beyond the SCANS because it requires complex cognitive processes to organize the information to be presented. The SCANS writing expectations established a continuum of expected skills that are used in today's current fast-paced technological workforce.

Due to the fast-paced technological advancements of the internet, electronic communication (e.g., email, texts, blogs), and social media (e.g., Facebook, Twitter), written communication has become a preeminent form of communication (Beaufort, 2006; Graham, 2013). All forms of writing are ubiquitous in the job market including email, PowerPoint presentations, technical reports, and memorandums (NCOW for America's Families, Schools, and Colleges, 2004, 2005). These forms of workforce writing come with underlying elements. First, the writing needs to be clear, accurate, and concise (ACT Inc., 2000; Smagorinsky, 2006). Second, workforce writing must cross multiple authors, purposes, and contexts through the appropriate and flexible use of grammar, rhetoric, and logical expression (Graham & Perin, 2007b; NCOW for America's Schools and Colleges, 2003). Third, workforce writing is imperative for the dissemination of business information such as budget proposals, company transactions, evidence reports, office memorandums, technical reports, policy changes, procedural instructions, or information updates (Casner-Lotto & Barrington, 2006). More importantly, a written document in the workforce becomes a permanent record of daily business proceedings (NCOW for America's Families, Schools, and Colleges, 2006). Thus, workforce writing is "high stakes writing" that maintains a business' image, policy, transactions, and even legal decisions (Coker & Lewis, 2008).

Survey research conducted on employers has corroborated that high level writing skills are expected. NCOW (2004) conducted a survey of 64 business leaders, from companies responsible for a total of almost four million employees, about the importance of writing in the workplace. The survey found that writing was considered a "threshold skill for employment and promotion" (p. 3). Writing was a required skill for both salaried and hourly employees. Business leaders expected salaried employees to write in over 30 to 80% of job fields and 15 to just over 30% of hourly positions. They also indicated that workers with poor writing on their applications, cover letters, or a writing sample would not likely to get an interview. They further commented that a worker's writing was a reflection of professionalism and the ability to attend to detail. Incidentally, more than half of the business leaders of salaried employees indicated that writing skills were reviewed as part of the process of a worker's promotion and advancement. Conversely, they also reported that lack of writing ability was more likely to be a factor in job termination.

Writing Crisis in the Workforce

Current research on writing in the workforce indicates that many workers do not have the writing skills that employers require (Graham et al., 2014b). This finding mirrors concerns expressed by Merrill Sheils in 1975 who proclaimed in a passionate cover article for Newsweek titled *Why Johnny Can't Write*, that the U.S. education system was "spawning a generation of semi-literates" (p. 1). Sheils indicated that commerce, industry, and professions depended on the clarity of the written word for formal written communications, and that the U.S. education system and its philosophies were to blame. Sheils stated that language skills were developed throughout life and that these language skills, particularly writing, needed consistent attention throughout school. The result had been a pool of college students who do not understand the importance of formal English; therefore, they are unable to articulate their thoughts effectively. By not teaching writing skills, Sheils affirmed that the workforce of the next generation might not have the opportunity to develop writing skills that are at a level required for American law, politics, commerce, and literature. Sheils ended with a plea for reinstating the importance of writing and its relevance. Authors have heralded Sheils (1975) as one of the forerunners of informing the public of a national writing crisis, and its possible effect on the U.S. economic future (e.g., Hourigan, 1994; Roach, 2009; Southern Regional Education Board, 2013; Varnum, 1986).

Results of recent surveys have revealed that Sheils' claims about the future of writing in the workforce remain relevant. On the NCOW survey (2004), only 35% of employers indicated that one-third of employees possessed the writing skills that employer's value. Consequently, 40% of employers needed to provide writing training or retraining for salaried employees and 20% or less for hourly employees. Moreover, the NCOW (2004) reported that the cost per year for this writing training was about $950 per employee, yielding extrapolated costs for large corporations of up to 3.1 billion or higher per year. In 2005, another survey revealed that 90% of white-collar workers and 80% of blue-collar workers had identified writing as an important skill in the workforce, but that employers listed at least 30% of its workers as failing to meet writing expectations (NCOW for America's Families, Schools, and Colleges, 2006). Subsequently, a 2006 survey of over 400 U.S. employers listed writing in English and written communications within the top five of "very important" skills for job success (Casner-Lotto Barrington, 2006). Yet on this same 2006 survey, employer respondents reported writing in English as deficient for 72% of high school graduates, 46% for college graduates, and 28% or greater for existing workers with college diplomas. A more current survey, continued to reveal the same trends. The American Management Association's 2014 Critical Skills Survey stated that 74.6% of employers listed written and oral communication skills as becoming even more important in the next three to five years. However, on this same survey, employers rated 51% of their employees' written and oral communication skills as average or below. In sum, the survey data over the last decade has proven that employers and employees have agreed on the importance of writing skills in the workforce, but that the actual writing skills of those preparing for entering or existing in the workforce continue to decline.

Summary

As technology has advanced, so have workforce expectations. Now more than ever, workers need to have strong literacy skills to maintain employment in any career. Writing is an area of literacy that has steadily increased in need and complexity. Writing is no longer a skillset that employers expect workers to develop; it is a requirement. Research has shown that our current workforce is deficient in the writing skills needed for today's competitive workplace. Students who are graduating from high school and college are also not meeting workforce expectations.

The next section discusses literacy for college and career readiness, which includes the literacy skills students need to prepare for the demands of workforce literacy. It provides evidence regarding the increase in rigor in writing standards for secondary students to prepare them for college and career demands. Current data does not indicate positive trends for student writing at the secondary school level to combat the writing crisis in the workforce.

Literacy for College and Career Readiness

This section discusses how the definition of literacy for college and career readiness has emerged over time. It also provides information on adolescent literacy, including high or advanced literacy and disciplinary literacy, as they are essential components of literacy for college and career readiness. Also within this section, the current writing demands for secondary students are outlined, along with the current writing crisis that exists in secondary school education. The demands and crisis documented support the need for writing intervention that targets complex information in an academic subject area.

History and Definition

The skills for college readiness have become aligned with the skills expected in the workforce (Somerville and Yi, 2002). As a result, the term college readiness has been combined with the term career readiness to become college and career readiness. Moreover, the rigor of 21st century expectations is not only for college, but also for the workforce, especially for those who may later pursue further career development (American Diploma Project, 2004; National Commission on the High School Senior Year, 2001).

Like Sheils' (1975) publication *Why Can't Johnny Write*, Gardner's (1983) famous publication, *A Nation at Risk*, sought to generate reform by stating grim educational statistics and providing revolutionary recommendations. Some argue that Gardner propelled education to the top of the national political agenda (Hess & McGuinn, 2002; Johanningmeier, 2010; Tharp & Gallimore, 1988). Gardner (1983) declared that our nation was at significant economic risk due to an undisciplined educational system that lacked high expectations. His two concerns were that students were not striving for excellence or developing the crucial “higher-order intellectual skills” (p. 9) needed for future college or careers.

Like Garner (1983), many researchers over the years continued to believe that the current educational system of menial standards had yielded unskilled low-achieving students (Koret Task Force for K-12 Education, 2003; National Commission on Excellence in Education; 1983). Similarly, college testing agencies have provided evidence that students are not excelling enough to achieve college and career readiness (Achieve, 2005; ACT, Inc., 2005, 2006, 2009, 2013b; The College Board, 2014). ACT, Inc. (2004) emphasized that college and career readiness meant that a student not only achieved the necessary skills, but also did not need additional remediation. However, studies have found that from 28 to 40% of students in four-year colleges are enrolled in at least one remedial course in writing; the rates for students in community colleges surpass 50% (National Conference of State Legislatures, 2015). Based on Achieve (2005) survey data, only a little more than half of high school students and employers felt that high school graduates were prepared with skills to advance beyond entry-level jobs. Even more alarming, a more recent survey revealed that only 25% of college professors felt that freshman college students were well prepared for college as compared to 89% of high school teachers who felt that students were prepared (ACT, Inc., 2013). College test score performance sheds light on students’ college preparation. The ACT, Inc. (2013) reported that only 31% students met college readiness benchmarks. The College Board (2014) reported that on the SAT 43% of students met their college and career readiness benchmark (i.e., total sum of reading, math, and writing score of 1550). Overall, these data reflect students' poor college-entry test performance and the need for remediation, confirming that high school teachers misperceive students' college readiness. Collectively, this has been referred to as the "college and career readiness gap" (National Center for Public Policy and Higher Education & the Southern Regional Education Board, 2010). Many researchers believe that this gap was the result of a misalignment of skill expectations for all academic areas and a lack of curricular coherence across high schools and colleges (Achieve, 2005; Callan, Finney, Kirst, Usdan, & Venezia, 2006, National Center for Public Policy and Higher Education, 2009; Somerville & Yi, 2002; Venezia, Callan, Finney, Kirst, & Usdan, 2005; Wise, 2008). The desire to close the college and career readiness gap was one impetus for the development of the CCSS (NGA for Best Practices & CCSSO, 2010b).

Gardner's (1983) second key concern was that high school graduates lacked higher-order intellectual skills. The CCSS defined these higher order skills as the literacy for college and career readiness (NGA for Best Practices & CCSSO, 2010a). More specifically, these are the listening, speaking, reading, writing, and language skills needed to succeed in college and career. Furthermore, according to the CCSS, a student who is college and career literate (a) demonstrates independence, (b) builds strong content knowledge, (c) responds to the varying demands of audience, task, purpose, and discipline, (d) comprehends as well as critiques, (e) values evidence, (f) uses technology and digital media strategically and capably, and (g) comes to understand other perspectives and cultures.

Adolescent Literacy

Adolescent literacy encompasses the coordination of existing and developing literacy skills that will allow a student to engage in lifelong learning in a rapidly changing world (Carnegie Corporation for Adolescent Literacy, 2010). One important aspect of adolescent literacy is the use of higher-order literacy skills or advanced literacy skills (National Council of Teachers of English, 2006). Bereiter and Scardamalia **(**1987) had expressed that historically there were two types of literacy, low literacy and high literacy. Low literacy had traditionally been basic reading skills for religious purposes such as reading the bible or reciting prayers. High literacy skills were beyond basic reading and writing skills and were considered the verbal reasoning abilities of the educational elite. High literacy required higher-order cognitive skills such as problem solving, self-regulating, executive structuring, and intentional learning. Moreover, higher literacy is the understanding of how reading, writing, language, content, and social appropriateness work together by using this knowledge in effective ways" (p. 1, Langer, 1999).

According to the International Reading (now Literacy) Association (2012), the most recent definition of adolescent literacy is "the ability to read, write, understand, interpret, and discuss multiple texts across multiple contexts" (p. 2). More specifically, an adolescent student must use advanced literacy skills to not only have to retain and make meaning of complex text (Meltzer, 2001), but also with variety of texts across academic disciplines and evolving media formats (Alvermann, 2002). These disciplinary skills are referring to another important aspect of adolescent literacy, which is disciplinary literacy.

Disciplinary literacy is the ability to understand and construct meaning within an academic discipline (e.g., science, history, English, math) through the development of the specific literacy, social, and cognitive practices of a discipline (Fang, 2012; Moje, 2008). Disciplinary literacy is considered the highest level of literacy, above advanced literacy skills (Shanahan & Shanahan, 2008). The use of disciplinary literacy skills allows adolescents to gain specialized knowledge that is not only relevant to each subject area, but for the literacy expectations for later college and careers in today's world (Zygouris-Coe, 2012).

Halliday (2007) described three stages of language development. In the first stage at infancy, children construct classes and develop the ability to generalize proper names and common names. In the second stage (ages 4-6), children transition from everyday spoken grammar to the grammar of literacy. The grammar of literacy allows a child to acquire educational forms of knowledge through reading and writing. The last stage (ages 9-12) is the "disciplinary literacy stage", when children move from the grammar of literacy to the grammar of literacy in the content areas. During this stage, students must learn to reconstruct language in a more theoretical mode, developing the ability to understand grammatical metaphor. Fang (2012b) has described literacy development literacy development is the braiding of the strands of everyday language, abstract language, and metaphoric language or the merging of Halliday's three stages. Fang further defined disciplinary literacy as "the ability to engage in social, semiotic, and cognitive practices consistent with those of content experts" (p. 1).

The literacy skills used by content area experts, or disciplinary literacy skills, should not be mistaken for what has been referred to as content area literacy. Content area literacy skills are the use of literacy skills to comprehend and use academic content in order to achieve curricular standards (Fang et al., 2006; Lenz, Ehren, & Deshler, 2005). In a content literacy approach, students access content information using the same literacy practices they would use in any academic discipline (Moje, 2008; Shanahan & Shanahan, 2012). In 2007, Heller and Greenleaf had recommended that literacy pedagogy extend beyond the use of the generic reading and writing strategies of a content area literacy approach. They furthered that a new approach to literacy pedagogy was needed because each academic area had its own unique lexicon, format, and stylistic conventions. Therefore, teachers would need to build students' understanding of how to develop the specific literacy practices of each discipline or develop students' disciplinary literacy (Shanahan & Shanahan, 2012). In disciplinary literacy, the text of the discipline dictates the literacy processes needed to master its content (Brozo et. al, 2013). Each academic discipline uses distinct literacy processes or language in order to engage, synthesize, and analyze disciplinary knowledge in a distinct way (Fang & Schleppegrell, 2010; Shanahan & Shanahan, 2008). Developing a student's disciplinary literacy skills provides the linguistic flexibility, agility, and accuracy that is imperative in the development of specialized knowledge (Fang, 2012b).

Disciplinary literacy not only focuses on the literacy processes needed to develop specialized knowledge, yet focuses on the unique communication patterns of an academic discipline (Ehren et al., 2012; Fang, Schleppegrell, & Moore, 2014). These unique patterns differ not only across each discipline, but differ from informal everyday conversational registers (Schleppegrell, 2007). Informal conversational language cannot convey the precision needed for complex disciplinary information (Fang & Schleppegrell, 2008). Thus, students need to be able create and use complex specialized texts (Fang, 2012b), which will require students to create and use complex specialized language. Complex text is referring to not only reading disciplinary text, but writing disciplinary text. At the secondary level, writing is even more important for it becomes the documentation of students' understanding of disciplinary knowledge through the students' use of complex written disciplinary language. The next section discusses the current writing demands at the secondary level.

Writing Demands in Secondary Education

In the realm of standards-based education, writing is included as a core skill along with listening, speaking, and reading. The majority of states in the U.S. have revised their state educational writing standards and assessments to increase writing achievement (Graham & Harris, 2013). The Common Core College and Career Readiness Anchor Standards in Writing (NGA Center for Best Practices and the CCSSO, 2010c) are an example of the current writing expectations for students to achieve by high school graduation. Generally, the College and Career Readiness Anchor Standards for writing address writing processes, contexts, purposes, components and conventions (Mo et al., 2014). According to the NGA Center for Best Practices and the CCSSO (2010c), students are expected to use the writing processes of planning, revising, editing, rewriting, or other approaches. With regard to the context of writing, students must use writing that is appropriate for task, purpose, and audience. This writing includes the use of technology, to not only publish writing, but to interact and collaborate with others. With regard to the purpose of writing, students must write across a variety of genres. They need to write arguments with reasoning and sufficient evidence, write informative/explanatory texts to convey complex ideas and information clearly and accurately, and write narratives to develop real or imagined experiences. For the purpose of research, students must write to develop questions, gather relevant information, demonstrate understanding through analysis and reflection, and cite evidence with credible sources. Starting in sixth grade, students use writing skills to demonstrate their knowledge across academic subjects, namely, science, social studies/history, and technical subjects. Within each academic subject, students must recognize the task, purpose, and audience within that discipline, using deliberate information, structures, and formats. For a student who is ready to graduate high school, writing is the key means of persuasion or argument, documentation of knowledge, and narration of a real or imagined experience.

According to Mo et al. (2014), there has been a shift in the expectations at the secondary level with regard to writing components and conventions, but little explanation of a logical order in achieving these goals. By the secondary level, students must be versatile in the components of content-related organization, structure, style, lexicon, and detail. At this level, students are also expected to have solid command of writing conventions (i.e., spelling, capitalization, punctuation) with practice focusing on improvement of these skills through the editing process.

The increase in writing demands has left many researchers concerned that most secondary students will not be capable of meeting these demands (Dockrell, 2014; Graham &

Harris, 2013; Scott, 2014; Troia & Olinghouse, 2013). Currently, the majority of students have poor achievement in writing, which lends credence to these concerns (Graham, 2013; Kelly et al., 2013, NCES, 2012b). Researchers (e.g., Applebee & Langer, 2011; Gillespie et al., 2014; Graham et al., 2014a, 2014b) have continued to collect data to prove that the educational system is still experiencing a writing crisis even after implementation of higher writing academic standards.

Writing Crisis in Secondary Education

Typically, writing instruction at the secondary level has been mainly in English language arts classes (Graham & Harris, 2013). Writing outside of this subject consisted of abbreviated compositions (Applebee & Langer, 2011; Troia & Olinghouse, 2013). Writing has now become a requirement across all academic areas, which is a new responsibility for teachers and students (Graham, Capizzi, Harris, Hebert, Morphy, 2014; Rijlaarsdam et al., 2014). More importantly, students' writing skills across academic areas will be considered as a measure of overall student achievement and a reflection of teacher effectiveness and school performance (Applebee & Langer, 2011; Dockrell, 2014). Therefore, schools are trying to implement changes to meet these new writing standards in an expeditious manner (Graham, Early, & Wilcox, 2014). Unfortunately, there are long-standing concerns about student and teacher preparation that may negatively affect achievement in students' writing (Graham, Gillespie & McKeown, 2012).

The majority of secondary students are writing below grade level (NCES, 2012b; Salahu-Din, Persky, & Miller, 2008). Researchers have speculated on various reasons contributing to poor student performance in writing, including minimal explicit writing instruction, limited writing practice, and lack of extended writing assignments (Perin, 2013). The National Commission on Writing (NCOW, 2003) had recommended that the amount of time students spent on writing needed to double, meaning schools needed to allot time specifically for writing. This increase in time would include writing in all classes across the curriculum, as well as writing assignments completed at home. Applebee and Langer (2011) reported that in middle and high school, students wrote an average of about 1.6 pages of content per week in English and 2.1 pages of content per week total across all other subject areas. The majority of writing (approximately 80%) was what Applebee and Langer described as "writing without composing" (p. 15), meaning fill-in-the blank, short answer exercises, or copying of information directly. More recently, Graham et al. (2014a) conducted a survey (*n* = 285) on middle school teachers and Gillespie et al. (2014) conducted a survey of high school teachers (*n* = 211) that measured the amount of time spent on writing. According to Graham et al. (2014a), middle school teachers reported that students wrote an average of 45 minutes per week in class and about 30 minutes per week out of class. Fifty percent of the middle school teacher participants assigned writing at least weekly, with writing that consisted of short answer responses, notes, and worksheets. According to Gillespie et al. (2014), the median writing time in high school was 30% of class time and two days per week for homework. The writing that high school students completed most often was note taking to support their learning. For both the middle school (Graham et al., 2014a) and high school (Gillespie et al., 2014), the majority of teachers indicated that students wrote summaries, descriptions, and journal entries at least one time per month,.

Most teachers indicated that longer assignments such as essays, research reports, or procedural writing were only assigned one time per year. There is little mention on any secondary level surveys as to the amount of explicit instruction the teachers provide on writing. Related survey data from teachers in grades 4-6 had indicated that teachers only spent about 15 minutes per day on writing, which may not be explicit instruction (Gilbert & Graham, 2010).

Results of these surveys indicate that writing recommendations made by the NCOW in 2003 are still not in effect today. A lack of "deeper writing" is still evident, specifically in content area classes, despite the new writing standards (Applebee & Langer, 2011). Writing, such as note taking and short answer questions, does not allow the deep understanding of topics beyond basic content comprehension, such as problem solving, analysis, or critical thinking (Applebee & Langer, 2009). Graham and Perin (2007b) have documented that increased writing practice may not yield student achievement in writing, especially if that writing practice is not developing higher level thinking skills to connect with academic content. Many teachers believe that with the time constraints and curricular pressures, any writing beyond note taking or worksheets may be impractical (Zumbrunn & Krause, 2012). However, teachers also recognize the limitations of students only practicing this type of writing. Qualitative research has reported that teachers do not think there is enough time to cover required content area material, model writing, complete writing, or provide written or verbal feedback about writing (Applebee & Langer, 2011). Thus, they utilize other writing practices, such as note taking and answering short-answer questions.

Graham and Perin (2007b) stated that quality instruction, meaning the use of research based practices, is the key factor for student writing gains. However, there is a plethora of research indicating that teachers may be without the preparation, skill set, motivation, or confidence to provide effective instruction (Graham, Gillespie, & McKeown, 2012; Street & Stang, 2008; Zumbrunn & Krause, 2012). Many content area teachers are not required to take courses specifically on how to teach writing while they are obtaining their college degrees (Graham et al., 2014a; Kiuhara et. al, 2009), nor do they receive the appropriate professional learning once they are working (Gillespie et al., 2014; Harris, Graham, Friedlander,& Laud, 2013; Street & Stang, 2009). Teachers who do not know how to conduct high-quality writing instruction will most likely provide only what they feel prepared to do (Gilbert & Graham, 2010; Harris & Graham, 2013). Furthermore, some teachers outside of English language arts still attest that teaching writing is neither their job, nor their responsibility (Gillespie et al, 2014; McLeod et al., 2001).

In addition, Coker and Lewis (2008) opined that academic writing tasks should resemble real-life writing tasks. Such tasks would include content information for purposes that are similar to what are used in college and careers, incorporating current media and technology (Alvermann, 2002; Applebee & Langer, 2009; Moje, 2008). These types of writing activities allow students to develop writing in a context that is anchored in discipline specific practices (Heller & Greenleaf, 2007; Meltzer, 2001; Perin, 2013). For example, the teacher would model the necessary writing process and then the students would research topics using multimedia to produce a written product. These assignments could be in written forms such as a persuasion blog or news article for social studies, a lab report or research request for science, a poem or biographical piece for English, or a procedural statement for mathematics. Students need to be equipped with the skills that will prepare them to write in a specific discipline, regardless of whether they choose to have a career in that discipline (Moje, 2007).

Summary

For students to be prepared for their future, they will need to develop literacy for college and career readiness to be prepared to meet the literacy expectations of the workforce. In today's fast-paced technological world, writing has become a prevalent and powerful form of communication (Graham, Gilbert, McKeown, 2012; Harris & Graham, 2013). Information that may have once been exchanged orally is now exchanged in an expedited written form that requires complex writing skills and agility (Graham, 2013). Writing today carries high stakes in that it is a permanent record or reflection of the writer; a writer's written product may be used as the basis for crucial decision making for school, college, or careers (Coker & Lewis, 2008; Graham, Harris, & Herbert, 2012). Students who struggle with writing, which is the majority, are at a major disadvantage, for writing competency is vital to future success (Street & Stang, 2008). One academic area that is vital for our nation's future prosperity is science. The next section discusses science and crucial aspects related to the literacy of science. It addresses how scientific literacy has continued to change as science has advanced over time.

Scientific Literacy

This section documents past science educational standards and notes how the present educational standards (i.e., The Next Generation Science Standards (NGSS)], have altered the definition of scientific literacy. A strong influence on scientific literacy has been governmental pressure generated by U.S. students lagging behind students of other industrialized nations in STEM education, innovation, and acquisition of high paying STEM occupations. The government has urged the educational system to support STEM advancement, which has yielded higher expectations in science and science literacy. This section continues with the higher expectations for science writing in secondary education and concludes with a discussion about the writing crisis that exists in science in secondary schools, despite educational reform to meet the government's expectations and demands. This aspect of the literature review supports the need for and importance of writing intervention in the areas of science.

History and Definition

Historically, scientific literacy has had a variety of meanings (DeBoer, 2000). Like other areas of literacy, the definition of scientific literacy and science education reform are highly correlated. During the post Sputnik era (late 1950's), the concern was that America was lagging behind in scientific innovation, particularly in space technology, which eventually could become a threat to national security (Bybee, 1997; Hiatt, 1986). The priority in education became a shift from liberal to technological education in order to accelerate and expand science through the creation of future scientists (Bybee & Fuchs, 2006; Dow, 1997). During the post-Sputnik era, the definition of scientific literacy was synonymous with science education's goals, which was to increase knowledge of the scientific method and broad content information across several fields of science (DeBoer, 2000; Hurd, 1997). The definition of scientific literacy began to transform in the 1970's, which led to the development of the National Science Education Standards in 1996. This transformation emphasized the need for scientific knowledge in order to be an informed citizen that could tackle everyday technological, economic, social, cultural, and environmental concerns that impact our future society (DeBoer, 2000; Villaneuva & Hand, 2011). Therefore, to provide such solutions for the future, the definition of scientific literacy widened to include not only broad content and scientific method, but solving problems, making choices and judgments, developing theories, providing facts, and sharing these results through oral and written discourse (McFarlane, 2013; Yore, 2007).

The National Science Education Standards (NSES, 1996) were developed to ensure that all students could achieve the new definition of scientific literacy (National Research Council [NRC], 1996). According to the NCES, scientific literacy enables people to "use scientific principles and processes to make personal decisions about scientific issues that affect society" (p. ix). To achieve scientific literacy, the standards promoted excellence and equity in science. In addition, these standards promoted a more active process including a new emphasis on inquiry and the value of life-long learning (McFarlane et al., 2013). This inquiry would involve the processes of investigation, interpretation, collaboration, and dissemination of information instead of simple recall of information (Marx & Harris, 2006). The ultimate goal of the NCES was that a person with scientific literacy could identify scientific issues and express opinions that were scientifically informed (NRC, 1996).

Scientific Literacy in the 21st Century

Many researchers have proposed that the goals of NCES were thwarted by the No Child Behind Act's (2001) attention to math and reading achievement, with a lack of focus on science (Bybee, 2010; Gonzalez & Kuenzi, 2012; Greenleaf et al. 2011a; Kuenzi, 2008; Marx & Harris, 2006; National Research Council, 2012; Pearson et al., 2010). Concerns about science education and its future quickly resurfaced when the National Academies Committee on Science, Engineering, and Public Policy (2005) published its study on America's competitiveness in the newly evolved global marketplace. The authors of this report concluded that in order for the U.S. to continue to compete, prosper, and remain economically secure in the 21st Century, science, technology, and engineering must continue to advance. The report documented that U.S. innovation was lagging behind other nations, thereby in danger of losing its competitive edge in the global economy (Bybee & Fuchs, 2006; National Academy of Sciences, 2011). In response, the U.S. government quickly developed legislation, proportioned funds, and garnered attention to what has become known as STEM, or the acronym for science, technology, engineering, and mathematics (Dugger, 2010; Sanders, 2009).

STEM has been defined as the integration of science, technology, engineering, and mathematics into a new cross-disciplinary subject that develops from pre-school to higher education and careers (Gonzalez & Kuenzi, 2012; Kuenzi, 2008). With regard to education, accomplishments in STEM are more likely when students participate in STEM coursework and opportunities early in life (Science Pioneers, 2014; Wai, Lubinski, Benbow, & Steiger, 2010) and later in high school (ACT, 2005). With regard to careers, projected growth of STEM occupations in this decade will be close to double that of occupations outside of STEM, which do not yield higher salaries like STEM occupations (Carnevale, Smith, & Melton, 2011; Langdon, McKittrick, Beede, Khan, & Doms, 2011). Yet, in the last decade, the number of graduates and new employees in STEM fields in the U.S. has not increased (Chen & Weko, 2009; STEM Education Coalition, 2013). Students of other nations are securing profitable STEM job positions over American graduates (Kuenzi, 2008). Consequently, the U.S. has continued to implement its plan to make STEM education a priority to prevent further loss of revenue (National Science and Technology Council, 2013). In 2015, President Obama increased the budget to support STEM education and research. He has commissioned the reorganization and evaluation of outcomes to determine further strategic STEM investments (Whitehouse Office of Science and Technology, 2014).

In addition to government focus and funding on STEM, the National Research Council (NRC, 2012) proposed a new framework for science education to address the need for increased scientific competency and advancement available for all students. The NRC's new framework (2012) would develop coherence of learning in science, utilizing 21st century skills, through the process of inquiry. The framework's eight practices of inquiry are: (1) asking questions and defining problems; (2) developing and using models; (3) planning and carrying out investigations; (4) analyzing and interpreting data; (5) using mathematics and computational thinking; (6) constructing explanations and designing solutions; (7) engaging in argument from evidence; and (8) obtaining, evaluating, and communicating information. The NRC's framework stressed a balance between content and inquiry. While engaging in inquiry, students would use common linguistic functions and language practices that span different types of science, which progress in complexity from elementary to secondary education. The linguistic functions at the primary school level may be identifying similarities and differences, determining cause and effect, and making predictions. At the secondary school level, linguistic functions are the semantic basis of the cited evidence required for explanations and arguments. Beyond the language skills inherent in science, the NRC (2012) indicated that the new science standards should "reflect high academic goals for all students' science and engineering learning-as literacy" (p. 279). In other words, science's inherent linguistic functions and language practices were now an integral part of the latest definition of scientific literacy.

The new standards that adhere to the NRC's framework are the Next Generation Science Standards (NGSS). The NGSS have three dimensions: disciplinary core ideas (content), scientific and engineering practices, and crosscutting concepts (NGSS Lead States, 2013e). These standards blend these three dimensions to intersect science practice, content, and connection. Unlike past standards, the NGSS do not specifically redefine scientific literacy. Instead, they refer to different types of science literacy, such "earth science literacy," or "space literacy." Within each area of science literacy, the standards outline mastery of the practices, ideas, and concepts inherent to that specific literacy (NGSS Lead States, 2013f; NRC, 2012).

In order to work in sync with the CCSS, the NGSS identified essential reading and writing skills needed for science (NGSS Lead States, 2013d). In relation to reading in the NGSS, students must develop an appreciation of science norms and conventions. These norms and conventions include using evidence, attending to precision and detail, making and assessing intricate arguments, synthesizing complex information, and following detailed procedures. In relation to writing, students write in science to assert and defend claims, demonstrate concept knowledge, and convey thoughts and experiences that occur within the science learning process. Students also need to be able to gain knowledge from elaborate scientific diagrams and data. Students also need to be able to organize science information to plan and execute oral presentations. To further specify the listening, speaking, reading, and writing skills needed within each type of science literacy, the NGSS has aligned its literacy expectations with the CCSS for Literacy in History/Social Studies, Science, and Technical Subjects (NGSS Lead States, 2013d; Quinn, Keller, Moulding, & Eberle, 2012).

Writing Demands in Science

The writing demands in science have increased significantly with the introduction of new science standards and the CCSS for Literacy in History/Social Studies, Science, and Technical subjects (Stage, Cheuk, Dero, & Hampton, 2013). Students are now required to document their knowledge beyond rote memorization and repetition of diluted content (NGSS Lead States, 2013a). Instead of stating, describing, or defining, secondary students need to analyze, argue, and provide evidence (NRC, 2012). Students must construct deeper understanding not only by doing science, but also by communicating their findings in writing like scientists (Billman & Pearson, 2013; Shanahan, 2012a, 2012b).

For scientists, writing is the primary means of communication in the scientific community (NRC, 2012). Yore et al. (2013) stated that scientists write to document their observations, provide explanations, make claims, and support their work with evidence. Scientists write to describe the steps of a process, assess the replicability of a procedure, provide visual displays to accompany the text, and connect findings to everyday life (Perin, 2013). Hence, writing is how scientists accomplish two goals related to the nature of science, that is, to argue and present claims with their peers, and to provide explanations and awareness to the public. These two goals, as stated by Yore et al. (2013) and Perin et al. (2013), are mirrored by the two types of writing specified in the CCSS Literacy in History/Social Studies, Science, and Technical subjects (Stage et al., 2013).

According to the CCSS Literacy in History/Social Studies, Science, and Technical Subjects writing standards, (National Governors Association [NGA] Center for Best Practices & Council of Chief State School Officers [CCSSO], 2010c), secondary students in science must be able to write argumentative and informative text in the subject areas. For an argument, a student will introduce a claim, distinguish it from an opposing claim, provide credible sources to present evidence for the claim, and end with a concluding statement related to the argument. For the informational text, a student will need to introduce the topic, and then organize the topic clearly with formatting, graphics, and multimedia. The student should provide facts, definitions, details, or quotations and conclude with a supporting statement. Furthermore, all students' writing in science must use a formal style and an objective tone. The student must also use discipline specific vocabulary to clarify a topic with discipline- specific language structures such as transitions to create cohesion. In order to meet these writing demands, secondary level students must have a high level of writing competency beyond the traditional note taking or abridged report writing that has been in existence in secondary science instruction (Carnevale et al., 2011; Shanahan, 2012b). Consequently, students who struggle with writing in secondary school will face new challenges developing writing that meets these sophisticated writing standards (Graham & Harris, 2013; Mo et al, 2014; Troia & Olinghouse, 2013).

Writing Crisis in Science

Student performance on large-scale national (NAEP, TIMSS) and international science assessments (PISA) continues to plummet (Fleischman et al., 2010; Kelly et al., 2013; Martin et al., 2012, NCES, 2012a). These assessments, unlike other science assessments, require students to provide written explanations for their answers for almost half of the test items (National Center for Education Statistics, 2007). Students in the U.S. do not exhibit the same abilities to write in science as students from other industrialized nations. There have been significant roadblocks in both science and science literacy education in the U.S., which may be impeding an upswing in achievement in science writing.

The first roadblock to improving science writing has been the lack of focus on science as a whole. Researchers have noted that during the last two decades the government had focused their attention on reading and math education with little focus on science (Bybee, 2010; Gonzalez & Kuenzi, 2012; Marx & Harris, 2006; Pearson, 2010). This lack of focus may not have allowed students to develop the background knowledge of science in elementary school to be successful in secondary school. In addition, students who struggle with literacy may have had increased time in remediation for reading, with fewer learning opportunities in science than students who do not struggle in reading (NRC, 2012). In many elementary schools, this increase of reading remediation has decreased the amount of time allotted to science class, which narrowed science instruction to the core content information with no time remaining to spend on science writing (NGSS Lead States, 2013b).

The second roadblock to improving science writing has been lack of science teachers who majored in science in college. Sadly, less than half of existing secondary school science teachers majored in science in college (Gonzalez & Kuenzi, 2012; National Science and Technology Council, 2013). The recruitment of teachers with intensive science training at the college level has become an educational priority for governmental agencies (STEM Education Coalition, 2013; Whitehouse Office of Science and Technology, 2014). Writing in science will require science teachers to embrace a new role not only as content area teachers, but also as teachers who are teaching discipline-specific writing as a form of science communication (Quinn et al., 2012). The science curricula and standards of the past focused on memorization of content, as opposed to learning content through inquiry and text based literacy tools such as formal writing (Greenleaf et al. 2011a; NGSS Lead States, 2013a). Science teachers without extensive science background may not be equipped with the expertise to teach the literacy of the written science register (Lemke, 1990). According to researchers and scientists, language practice is a legitimate part of science literacy (Shanahan, 2012a, 2012b) that must be taught competently and explicitly (Hand et al., 2003; Yore et al., 2003). Science instruction must combine rhetorical technicality with reasoned argument (Fang et al., 2014) and blended use of informal and formal writing immanent in the multidimensional nature of science (Hand et al., 2003). Without the tool of written language in science, students will not develop this academic style of written language needed for their future (Yore et al., 2006, 2007). Even more pressing, this type of writing practice in science will become increasingly necessary as the literacy requirements related to STEM continue to escalate in specialization and complexity (Ehren, Lenz, & Deshler, 2014).

The third roadblock to improving science writing has been its complexity. The nature of science and scientific method requires precise language to reveal valid empirical evidence (Edmonds & Kennedy, 2012; Yore et al., 2006). The written pattern, sequence, and detail in science writing is important to extract information accurately (Halliday, 2004; NRC, 2012). Science employs a specialized language that is technical, dense, abstract and unlike everyday discourse, hence difficult to read and write (Fang, 2012; Schleppegrell, 2007). Science also commands an authoritative, unopinionated, declarative tone (NRC, 2012: Yore et al., 2003). This tone yields a complex sentence structure of multiple word sentences, embedded phrases and clauses, relative, adverbial, and nominal clauses, increased distance between related grammatical elements, and non-canonical order such as passive voice (Fang & Schleppegrell, 2008; 2010; Scott & Balthazar, 2010, 2013; Scott & Koonce, 2014; Shanahan, 2012b). Science language must also be flexible, reciprocal, and multimodal (Hand et al., 2003; NRC, 2012; Yore et al, 2003). The language must be flexible in its use of formal and informal structures when bridging information gaps between the scientific community and the public. The language must be reciprocal meaning one must listen and read, then speak and write for tasks in science such as gathering evidence and then providing arguments. Last, the language must be multimodal and include a mixture of words, diagrams, charts, symbols, mathematical equations, with cohesive written language structures.

Summary

The development of consistent expectations and standards that resemble workforce related skills is imperative in all subject areas, especially in a constantly evolving area, such as science (McFarlane et al., 2013; Yore et al., 2003). Researchers, scientists, and government officials all recognize the value of quality science education and writing instruction. Writing compromises a large portion of scientists' work, thus it needs to remain an important instructional target in science classrooms (Metz, 2015).

This section concludes discussion of the pertinent literature in the three areas of literacy that are relevant to this study: workforce literacy, literacy for college and career readiness, and scientific literacy. What is common across all three areas of literacy, is that over the course of history, the writing demands and expectations have increased, but the writing skills of those who must meet those higher demands and expectations have not increased at the same rate. This lag in writing skills, or what has been referred to as the writing crisis, continues to remain a concern. Effective writing intervention in the schools can help combat these concerns.

The next section presents areas of research that are pertinent to the theoretical framework of the intervention: the core areas of syntax, the relationship of writing, the reciprocal language processes, and metalinguistics. In relation to the core areas, the next section also discusses the subareas of disciplinary literacy, academic learning, and science. Last, this section concludes the chapter with a review of the research basis of sentence combining.

Syntax

Syntax is an area that has slowly garnered more research attention, particularly its role in the development of reading and writing skills needed to tackle complex educational text (Berninger et al., 2011; Scott, 2014). In order to make sense of complex text, students need to be able to decipher as well as create complex sentences, through the negotiation of syntactic structures (Schleppegrell, 2013). Therefore, syntax is an area that should be incorporated into existing literacy instruction for all students, especially those who struggle (Eberhardt, 2013).

Syntax is defined as the arrangement of words in sentences (Owens, Farinella, & Metz, 2015). Other definitions specify that syntax is the form or structure of a sentence and the rules that govern sentence organization (Pence-Turnbull & Justice, 2013). These rules determine the word order or word combinations that are linguistically acceptable and make sense (Owens, 2008).

Syntactic knowledge is the conscious understanding of syntactic categories (Mokhtari & Thompson, 2006). Syntactic knowledge develops as language develops (Arndt & Schuele, 2013). As syntactic knowledge increases, a person’s ability to comprehend and produce language of greater complexity increases (Scott, 2004). A skill distinct from syntactic comprehension and production is syntactic awareness (Cain & Oakhill, 2007). Syntactic awareness is the ability to manipulate and reflect on the structure of language (Lightsey & Frye, 2004). Syntactic awareness is also a metalinguistic skill that taps into syntactic knowledge through memory processes (Cain, 2007). Syntactic knowledge and memory processes allow one to flexibly combine words into complex sentences (Brea-Spahn, 2014), assign sentential structure (Tunmer et al., 1988), and detect and correct sentential errors (Scott, 2009a).

Syntactic knowledge and awareness are evident in the processing and production of language in all its modalities-listening, speaking, reading, and writing (Scott& Koonce, 2014). Paul & Norbury (2012) have outlined receptive syntactic abilities, or listening skills, as a function of language comprehension. Hence, syntactic development is first the understanding of phrases, then simple sentences, and then complex sentences. Paul & Norbury furthered that a child’s syntactic comprehension is measured by the accuracy of a child’s response, first nonverbally such as gestures, identifying pictures, following directions, and then later in their verbal or written responses. Later in childhood, strong syntactic skills become essential for the more complex processes of reading and writing (Arndt & Schuele, 2013; Scott, 2004). A child who is suspected of having difficulty with language, a language impairment, or other disability, may struggle with these syntactic comprehension tasks and develop syntactic skills at a slower rate (Scott & Koonce, 2014). Students who struggle with language may not understand the differences among correct and incorrect syntax rules, simple and complex sentences, and the phrasal and clausal structures in complex sentences (Scott & Balthazar, 2013).

Syntax and the Reciprocal Language Processes

With regard to speaking and syntax, researchers have long documented spoken syntax development, with more focus on the early language developmental milestones (Scott & Stokes, 1995). However, more research has recently emerged related to school-aged students and adolescents (Nippold, 2010). Traditionally, language samples, or the collection, dictation, and analysis of language, focused on spoken conversational samples for young children (Owens, 2008; Pence-Turnbull & Justice, 2013) and school-aged students and adolescents (Nippold, Mansfield, Billow, & Tomblin, 2008; Scott, 1988). Recent research has focused on other spoken genres, such as narrative (Nippold et al., 2014), expository (Nippold et al., 2008, Nippold, Hesketh, Duthie & Mansfield, 2005; Nippold & Scott, 2010; Scott & Windsor, 2000), and persuasive (Nippold, Ward-Lonergan, & Fanning, 2005) tasks. According to Nippold (2010), as children develop spoken discourse, the syntax of their sentences gradually increases in length and complexity. Syntax varies with context or genre, with the most complex syntax in persuasive tasks, then expository, narrative, and conversation, respectively (Nippold, 1993;

Nippold et al., 2014). Syntactic complexity may also vary with level of knowledge, interest, and motivation to talk (Nippold, 2010).

Research has shown that students who struggle with syntax will perform similarly on spoken tasks with regard to length of response and the number of words (Gilliam & Johnston, 1992; Scott & Windsor, 2000). However, students who struggle with syntax lack density and complexity of their spoken language, especially in narrative and expository tasks (Nippold, 2007; Nippold, Hesketh et al., 2005; Nippold, Ward-Lonergan, et al. 2005). The spoken language of a student increases in syntactic complexity into adolescence. Adolescents begin to use more propositions, subordination and embedding (i.e., relative, nominal, and adverbial clauses), cohesive devices, non-canonical syntax order, and low frequency syntax structures (Gummersall & Strong, 1999, Nippold, 2010; Thompson & Shapiro, 2007). According to Scott and Balthazar (2013), students who struggle with syntax will lack these attributes in their spoken syntax. In addition, their spoken language may be characterized by use of fragmented sentences that lack cohesion, clarity, precision, and variety. These students' sentences also tend to contain significant morphological or semantic errors that violate syntactic rules.

With regard to reading and syntax, there has been a significant amount of research with varying results. Some researchers maintain that syntactic knowledge provides the foundation for linking language forms and meaning; therefore, syntactic knowledge is correlated with reading fluency and comprehension (Mokhtari & Thompson, 2006; Tunmer, 1984; Walker, Mokhtari, & Sargent, 2006). This correlation allows students to make sense of what they are reading

(Lightsey & Frye, 2004). Other researchers have confirmed that it is syntactic awareness that is essential to reading comprehension (Menyuk & Flood, 1981; Scholl & Ryan, 1980). Having the ability to make syntactic judgments allows a reader to parse syntactic structures at the sentence level and monitor their reading comprehension (Cairns, Schlisselberg, Waltzman, & McDaniel, 2006; Scott, 2009a). Lack of syntactic skills has been suspected as the reason why a student with appropriate decoding and phonemic awareness skills still struggles with reading (Catts, Adlof, Hogan, & Ellis-Weismer, 2005; Nation & Snowling, 2000; Scarborough, 2001). Students who struggle with reading have been reported to have sentence processing deficits (Leikin, 2002); limited spoken syntactic repertoire (Nation & Snowling, 2000, 2004); and trouble detecting and correcting syntactical errors (Mokhtari & Thompson, 2006; Scott, 2004). Yet, a definitive connection between syntax and reading has been difficult to establish, which may be the reason why syntax and reading has been a less popular research topic as opposed to decoding or phonemic awareness skills (Lightsey & Frey, 2004). Speculation that is more current is that syntax is not a single ability, but a set of multiple skills that is difficult to distinguish from other language skills needed for reading (Cain & Oakhill, 2004; Mokhtari & Thompson, 2006; Scott, 2009a). Syntactic skills are mediated by other language or related skill areas such as phonology, morphology, vocabulary, memory, pragmatics, social interaction, and executive functioning (Berninger & Abbott, 2010; Cain, 2007; Singer & Bashir, 1999; Scarborough 1991; Troia, 2012).

With regard to writing, research has indicated that like reading, writing is a complex language process that incorporates a variety of skills, including syntax (Nelson, Roth, Scott, Van Meter, & Troia, 2006). Writing is the active coordination of orthographic, graphomotor, and linguistic skills (Puranik, Lombardino, & Altman, 2008). Furthermore, Berninger and Abbott (2010) indicated that writing is a combination of transcription processes (i.e. handwriting and spelling) and multi-levels of language such as semantic word choice, sentence-level syntax, and text composition. Berninger and Abbott (2010) revealed from their study that students with developed syntactic awareness had better ability to transfer their thoughts into carefully crafted sentences. Moreover, syntax as well as spelling and transcription contributed uniquely to sentence composition for students in fourth grade.

Like Berninger and Abbott (2010), researchers have attempted to document distinct milestones for written syntax. Many mention an evident change in development at around age 9 or fourth grade. One of the first documented studies by Kellogg Hunt (1965) concluded that students in grade 4 use all syntax structures, but that the later school years is the period where students master the ability to manage syntax structures that increase complexity within a single sentence. In contrast, another longitudinal study conducted by Loban (1976) recorded oral and written language development for students from kindergarten to age 13. Loban's conclusion was that written and oral language developed in parallel, with written language patterns mirroring oral language patterns in complexity about a year later. Conversely, Perera (1984) examined speaking and writing, but specified a more distinct difference in later elementary school-age development than the parallel theory of Loban (1976). Like Berninger (2011), Perera (1984) believed that early writing was consumed with the processing of mechanics, thus early writing is lacks the complexity of early speaking. Like Hunt (1965), Perera (1984) furthered that at age 9 years, a student's writing begins to surpass speaking in its lexical and structural density. This period, age 9 years, also corresponds with Halliday's (2007) "disciplinary literacy stage" where there is an increase in abstract relationships and higher-level thought (Farrall, 2013).

A difference between speaking and writing is that speaking can remain in a linear code to communicate language complexity, utilizing pauses and clauses combined with conjunctions (Biber, 1988; Halliday, 1985). However, written language must include more embedding in order to pack a greater density of ideas into a single sentence (Cazden, 1974). Further, writing requires a degree of planning and organization to anticipate a reader's questions and limitations, as opposed to the context and spontaneous repair that occurs in spoken communication (Farrall, 2013). The written context requires the use of obligatory complex structures that convey the message to the reader (Scott, 2004). All researchers agree that complex written language development is marked by specific characteristics or increases in the number of words, number of correct and complete sentences, and number of sentence types. The development of complex writing is also marked by an increased (a) variety of word order, (b) conjunctions, (c) subordinate and embedded clauses, (d) sentence combinations, and (e) complex vocabulary containing derivational morphemes (Loban, 1976; Nelson, 2013a; Nippold, 2010; Scott, 1988; Scott, Nelson Anderson, & Zelinski, 2006; Scott & Stokes, 1995). In addition, a student's writing skills can be judged for syntactic productivity and syntactic complexity. Syntactic productivity is measured by the total number of words or utterances produced in repose to a particular language task (Nippold, 2010). Likewise, syntactic productivity can also be judged according to writing conventions, which is believed to be related to syntactic awareness (Nelson & Van Meter, 2007). Syntactical complexity is typically measured by the mean length of terminal units (MLTU) or t-units (i.e., main clause and subordinate clauses) as well as the number of subordinate clauses embedded in other subordinate clauses or clausal density (Hunt, 1965; Farrall, 2013).

Past research has presented the measures of length of T-unit (MLTU) or clausal density as the best measure in distinguishing children with language disorders from typically developing peers with regard to written syntax skills (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004; Nelson & Van Meter, 2007; Scott & Windsor, 2000). However, these differences have not always been consistent across studies, with some studies revealing similar performance in written syntactic complexity for students with and without language disorders (Scott, 2009b). Other factors such as text genre (Berninger et al., 2011; Farrall, 2013) or age and genre (Sun & Nippold, 2012) have been better predictors of a student's written syntactical complexity than presence or absence of a language disorder. In general, narrative writing tasks are easier for younger students; older students are more successful with expository tasks. Within the domain of expository or informational writing tasks, writing for the purpose of persuasion poses the most difficulty, but also improves with age.

Although measures of syntactic complexity such as MLTU may not always distinguish writers who struggle, such as students with language disorders, one aspect of syntax is irrefutable. Students who struggle with writing tend to have more syntactical errors in their writing (Scott, 2014). They may also exhibit lower syntactic productivity and complexity (Nelson et al., 2006; Puranik et al., 2008). Consequently, they may omit content, omit key sentence constituents, use fragments and run-on sentences, misplace clauses, or confuse morphological endings (Nelson, 2013a). Students who struggle with writing may also have difficulty with other aspects of writing, such as spelling, mechanics, handwriting, executive functioning, working memory, and text generation (Berninger & Winn, 2006; Dockrell, 2014; Farrell, 2013; Singer & Bashir, 1999). The ability to produce accurate and competent writing is a pervasive weakness for many students, thus many students are considered “struggling writers” regardless of whether or not they have a diagnosed disability (Dockrell, 2014; Graham & Harris, 2013; Rijlaarsdam et al., 2014).

Syntax and Disciplinary Literacy

Writers who struggle have difficulty extending their learning, even more when extending across text of a variety of disciplines (Mo et al. 2014; Rijlaarsdam et al., 2014). Disciplinary or academic text has complex syntactic features (Scott & Balthazar, 2010). Schleppegrell (2004) has stressed three characteristics of academic texts: field, tenor, and mode. Field is related to the topic, which is usually reflected in lexical choices that are technical and abstract, such as long nominal groups. These texts also have a high lexical density due to a greater proportion of nouns, verbs, and adjectives. Tenor is related to the writer’s stance. Informational text is often authoritative. Tenor is reflected in use of declarative sentences, modal verbs, passive voice, and contrastive conjunctions attached to supporting clauses. Last is mode, which is the organizational structure that a discipline commands. This organization is hierarchical and accomplished through extensive clausal embedding and subordination, and theme/rheme structures (Fang & Schleppegrell, 2010; Scott & Balthazar, 2010).

Syntax in Science

There are specific syntax structures that are common in science writing. First is nominalization, which is the transformation of verbs into a noun to describe a process (e.g., evaporation) (Shanahan & Shanahan, 2008). Nominalization allows one word to represent an explanation sequence (Fang & Schleppegrell, 2010). Second, text in science is rarely in the first person. It is declarative and typically in the passive voice to establish authority (Fang, 2005; Fang, Schleppegrell, & Cox, 2006). Third, science uses restrictive and nonrestrictive relative clauses, object complement clauses, and adverbial clauses to create logical relations (Fang, 2006; Fang & Schleppegrell, 2008). Science text uses conjunctions to signal sequential, conditional, causal, and comparative relations (Schleppegrell, 2007; Troia, 2009). Science texts utilize subordinate clauses in which the subject or its relative pronoun is omitted. Also in science text, there is a large distance between the main noun and verb of a sentence and nonfinite verbs are not marked for tense or number (Scott & Balthazar, 2010). In addition, in relation to theme/rheme in science, the most important information is usually placed at the end of the sentence (Fang & Schleppegrell, 2008). Therefore, sentences in science contain adverbial clauses that precede the main noun, or nominal, or object complement clauses just after the main noun, but before the main verb (Fang, 2005; Otero & Graesser, 2002; Schleppegrell, 2007).

Summary

This section reviewed the area of syntax with regard to reciprocity, disciplinary literacy, and the discipline of science. The information presented in this section is the foundation for a syntactically based intervention, such as sentence combining. This study explored the effect of syntax on writing and concept development in a discipline. The next section presents literature that supports these hypotheses.

Writing, Reading, and Learning Connection

The language processes of listening, speaking, reading and writing are reciprocal and interrelated, which can yield positive gains on each other (Berninger & Abbot, 2010; Dockrell, 2014; Catts, Fey, Zhang & Tomblin, 1999; Paul & Norbury, 2012; Shanahan, 2006; Westby, 2012). However, the relationship between writing and reading is uniquely similar. Reading and writing, unlike listening and speaking, are facilitated through the written word, which has distinct semantic, syntactic, and pragmatic structures different from spoken language (Shanahan, 2006). The unique similarities of reading and writing support the furtherance of research that writing promotes the development of reading and learning academic content (Rijlaarsdam et al., 2014).

Writing and Reading

Writing and reading are similar in that they are dependent on shared knowledge and common cognitive abilities (Berninger et al., 2010; Shanahan, 2006). According to Fitzgerald and Shanahan (2000), this knowledge can be separated into four categories: (1) domain knowledge, (2) metaknowledge, (3) knowledge of universal text attributes that underlie reading and writing, and (4) procedural knowledge. Domain knowledge is the ability to obtain new content knowledge based on prior knowledge. Metaknowledge is monitoring one's own meaning making through knowledge of the function and purpose of reading and writing, and its interaction. Universal text attributes are the phonemic, morphemic, syntactic, pragmatic, and orthographic features of a text that underlie reading. Last is procedural knowledge, which is knowledge to access, use, and generate information during reading and writing.

Research on the writing-reading relationship has focused on establishing distinct correlations (Fitzgerald & Shanahan, 2010). Consequently, it has been concluded that writing and reading are neither symmetrical, nor the inverse of each other (Berninger et al., 2006); rather they influence and enhance each other (Graham & Hebert, 2011; Shanahan & Lomax, 1986, 1988). This writing/reading relationship is the reason why past research, as well as current research, has recommended writing and reading pedagogies that are interconnected (e.g., Applebee, 1984; Klein, 1999; Smith, 1988; Stotsky, 1982). Biancarosa and Snow (2004) recommended intensive writing as one of the fifteen instructional elements proposed in *Reading Next*. They indicated that elements of academic writing such as grammar and spelling reinforce reading comprehension that requires higher-level reasoning and critical thinking. Graham and Gilbert (2010, 2011) conducted a meta-analysis that examined the effects of writing on reading. In Graham and Gilbert's report, *Writing to Read* (2010), they proposed three main recommendations to connect writing and reading instruction. The first recommendation was that students should write about texts that they read through responses, summaries, notes, or answer questions (average weighted effect sizes = .40 on norm-referenced tests, .51 on researcher created tests). The second recommendation was that students should learn writing skills and processes that help create text such as spelling, which aids word reading; spelling and sentence construction, which aids reading fluency; and sentence and paragraph construction, which aids reading comprehension (average weighted effect sizes = .18 on norm-referenced tests, .27 on researcher-created tests). The third recommendation was that students needed to increase their amount of writing and production of their own texts (average weighted effect sizes = .30 on norm-referenced tests).

Writing to Learn

The philosophy of writing to promote learning has a long history. Writing to learn theory has often been associated with Vygotsky's theory of "inner speech" (1987) and its developmental process. By school age, a child's language of social communication develops into the mediation and regulation of semiotics and thought. Children use explicit knowledge to make sophisticated semantic and syntactic choices that structure thought and organize experience for learning (Bazerman et al., 2005; Herbert, Gillespie, & Graham, 2013). Britton, Burgess, Martin, McLeod, and Rosen (1975) proposed a model of three functional types of writing: (1) transactional, for communicating information; (2) poetic, for creating beautiful objects; and (3) expressive, for exploring and reflecting upon ideas. It was the last category that spawned the idea of writing to learn. Similar to Vygotsky (1987), Britton et al. (1975) argued that expressive writing becomes the tool for learning as it develops formality and informational or persuasive purpose. Emig (1977) further supported this view and pointed out that as opposed to the natural process of speaking, writing is both a process and a product that create a form and source of learning. Emig believed that writing was the "the symbolic transformation of experience through the specific symbol system of verbal language that is shaped into a graphic product" (p. 124). Both Britton et al. (1975) and Emig (1977) influenced other researchers who promoted writing as a process with nonlinear sub-processes (e.g., planning, monitoring, drafting, revising, editing); these processes vary with writer and the writer's purpose (Applebee, 1984; Flower & Hayes, 1981; Klein, 1999).

Writing Across the Curriculum

As writing to learn theory gained popularity, this theory was also employed to promote writing as a learning tool across academic content areas or programs labeled as Writing Across the Curriculum (WAC, McLeod, 1992). WAC programs were implemented at the college level and consisted of two major approaches (Bazerman et al, 2005). First, was that writing is for knowledge transformation or discovery as opposed to just knowledge telling (Bereiter & Scardamalia, 1987). Second, was that the development of academic discourse or writing in a discipline is to develop reasoning and methods of proof within that discipline. The WAC program encouraged writing to learn and learning to write in all disciplines, a change in teaching methods, and a supportive administrative structure that would build consensus across disciplines (McLeod & Miraglia, 2001). There is limited research that has looked at the exclusive characteristics of an academic discipline and its impact on learning in that discipline (Faggella-Luby et al., 2012). Some research has been done to investigate how writing in an academic content area has yielded gains in learning in that content area (Applebee, 2000). A meta-analysis conducted by Bangert-Drowns, Hurley, and Wilkinson (2004) on writing to learn studies concluded that writing to learn produced small, positive effects on school achievement as compared to conventional instruction. Graham and Perin (2007b) also indicated a small positive effect (i.e., effect size of .23) for writing to learn activities, which were consistent across studies. Other researchers such as Peha (1995) and Marzano (2012) have promoted aspects of writing to learn in the academic content areas as the key component for deepening understanding and improving retention of content.

At present, writing to learn practices have more documented evidence as an effective practice in science (Fry & Villagomez, 2010), particularly within the last decade (Holliday, Yore & Alvermann, 1994). Science researchers value writing as a form of learning in science education (Mason, 2001). Some science researchers have proposed design principles for writing in science (e.g., Keys et al., 1999; Klein & Kirkpatrick, 2010; Villaneuva & Hand). Moreover, researchers have documented that writing about science text facilitates comprehension and learning, as it provides a means for recording, connecting, analyzing, personalizing, and manipulating key ideas from the text (Gilbert & Graham, 2010). Research conducted by Hand, Prain, and Yore (2001) revealed that students who wrote in science explained their ideas better on subsequent tests and performed better on higher order thinking tasks. Some researchers have linked writing in science class, with gains in science comprehension at the elementary level (Cervetti, Barber, Dorph, Pearson & Goldschmidt, 2012; Guthrie, Anderson, Haug, & Ødegaard, 2014; Romance & Vitale 2012a, 2012b; Solomon, Alao, & Rinehart, 1999) and at the secondary level (Geier et al., 2008; Greenleaf et al., 2011b; Hand, Wallace & Yang, 2004; Keys, Hand, Prain & Collins, 1999). Additional research to support that writing improved learning in science includes Akkus, Gunel & Hand 2007; Balgopal &Wallace, 2009; Bullock 2006; Hand et al., 2004; Hohenshell & Hand, 2006; Klein, Piacente-Cimini, & Williams, 2007.

Summary

This section has presented research literature that supports the hypothesis that an intervention in writing may yield gains in an academic subject area such as science. This section has highlighted the connection between writing, reading, and learning academic content. The next section presents the last theoretical piece, metalinguistics, which uniquely defined the sentence-combining intervention used in this study.

Writing in a Metalinguistic Framework

This section provides background to support writing within a metalinguistic framework. The researcher used a metalinguistic approach to the writing intervention that is rooted in this literature. Last, the next section ends with discussion of the main intervention, sentence combining. Sentence-combining intervention focuses on syntax, promotes learning through language reciprocity, and improves reading comprehension skills.

Metalinguistics

Metalinguistics is considered a branch of metacognition (Gombert, 1993). According to Flavell (1979), metacognition is the conscious process of “cognition of cognitive phenomena” (p.

906), or the ability to access our own cognitive process (Paul & Norbury, 2012). Some researchers believe that metacognition is a precursor to the development of "Theory of the mind" (Westby, 2014). Theory of the mind is the theory that as part of cognitive development, children develop recognition of other's thoughts and emotions to make inferences about them, as well as reflect on their own thoughts and emotions (Wellman, 1990). Students must employ their awareness of these mental processes during language tasks.

Subsequent to metacognitive development is the development of the process of thinking about language, or metalinguistics (Gombert, 1992). Metalinguistic skills are the skills that allow one to think and talk about language overtly (Finestack, 2013; Flood & Salus, 1982). It is when language becomes the object of thought, rather than the transmission of thought, that one consciously reflects on the nature and properties of language (Cazden, 1976; Van Kleeck, 1982). In other words, it is the ability to reflect on language rules explicitly by recalling implicit linguistic knowledge (Sutter & Johnson, 1990). Metalinguistic skills concentrate on recognizing, differentiating, evaluating, correcting, explicating, and relating linguistic forms to enhance verbal and nonverbal communication (Menyuk & Flood, 1981; Van Kleeck, 1982). This communication is enhanced by the conscious awareness of language, or metalinguistic awareness (Gombert, 1992; Paul & Norbury, 2012). Metalinguistic awareness uses conscious awareness to access linguistic knowledge in order to manipulate language for a variety of goals (Peets, 2014).

Earlier research had questioned when metalinguistic skills develop, if it preceded literacy or was a consequence of literacy (Flood & Salus, 1982; Tunmer, 1984). Some believed that metalinguistic awareness develops in children at preschool age, but can only take the simpler form of listening and talking about language (Finestack, 2013; Menyuk, 1976). Others believed that becoming literate is subsequent to reflecting consciously on language, which could only occur in middle childhood (Halliday, 1993; Tunmer, 1984). Other researchers believed that metalinguistic skills were completely separate and unrelated to language proficiency (Bialystok, 1986). Regardless of these opposing views, metalinguistic abilities were proven as an important mechanism that mediates language learning and development (Chen & Jones, 2013; Tunmer et al., 1988; Van Kleeck, 1982). Moreover, language develops simultaneously with the development of cognition and attention (Bialystok, 1986; Cazden, 1976; Kamhi & Koenig, 1985; Sutter & Johnson, 1990).

Earlier research had also specified that metalinguistic awareness be possibly related to the ability to separate form from content (Kamhi & Koenig, 1985). Today, metalinguistic awareness is viewed as the reflective ability to manipulate structural features of language that contribute to its meaning (Brea-Spahn, 2014). More specifically, one uses the parallel processes of segmental, morphemic, and sentential information to manipulate language structures at the phoneme, word, sentence, and conversational level (Finestack, 2013). Therefore, metalinguistic awareness and skills include the interaction of all the language components; namely, phonology, morphology, syntax, semantics, and pragmatics (Menyuk & Flood, 1981). Students who have strong metalinguistic awareness and skills are typically more successful with not only the language processes of speaking and reading, but reading and writing (Armbruster, Echols, Brown, 1983; Burkhalter, 1996; Ebbels & van der Lely, 2000; Ebbels & van der Lely& Dockrell, 2007;

Finestack, 2013; Hirschman, 2000; Hodgson, 1992; Lightsey & Frey, 2004; Scholl & Ryan, 1980).

Writing researchers have asserted that writing is the most complex of the language processes. Writing requires simultaneous coordination of a variety of linguistic abilities (Hillocks, 1986; NCES, 2012; Troia & Graham, 2003; Witte & Faigley, 1981). Metalinguistic awareness and skills are among these linguistic abilities. Finestack (2013) stated that metalinguistic awareness and skills should be taught explicitly, presenting patterns and principles related to the language context. Specifically, students should be taught how to use language as a tool for thinking, while in the context of writing (Lightsey & Frey, 2004; Menyuk & Flood, 1981). Hirschman (2000) had incorporated metalinguistic awareness into writing practice to help students who struggle with language surface the necessary cognitive processes to a conscious level in a modality that is not dependent on auditory memory. In relation to grammar specifically, Burkhalter (1996) believed that if metalinguistic awareness was taught explicitly with the teaching of grammar, then it would be easier to use that awareness when writing. This explicit discussion of grammatical structures has proven to be for effective in the instruction of complex written language (Scott & Balthazar, 2010). During instruction, the teacher should facilitate this explicit discussion using planned metalanguage (Basturkmen et al., 2002). The teacher's metalanguage provides the opportunity to attend to form, while engaged in meaning-focused language use (Ellis, Basturkmen, & Loewen, 2002).

Myhill et al. (2013) ascertained that grammar skills are a subset of metalinguistic skills. Other researchers have used the term metalinguistic awareness and skills synonymously with the terms grammatical awareness, grammaticality, or simply grammar (Cairns et al., 2006; Scholl & Ryan, 1980; Sutter & Johnson, 1990). Within these definitions, grammar is a broader term that includes the organization of all areas of language including phonology, morphology, semantics, syntax and pragmatics (Fromkin, Rodman, & Hyams, 2013; Yule, 2010). Grammar has also been defined in a narrow sense, or a more traditional sense, as the rules that govern a language's structure and parts (Andrews et al., 2006). Traditional grammar instruction incorporates the explicit and decontextualized teaching of grammar rules, recitation, or drill exercises, and little implicit teaching (Hillocks, 2005). The method of "traditional grammar instruction" is one that researchers have constantly questioned (Burkhalter, 1996; Haynes, 1978; Vocke et al., 2012). A significant number of studies indicate that traditional grammar instruction has little benefit or can detract from student's writing (Andrews et al., 2006; Biancarosa & Snow, 2004; Graham & Perin, 2007a, 2007b; Haynes, 1978; Hillocks, 1986; Maize, 1954; Wyse, 2001).

Contrary to "traditional grammar instruction", numerous studies have supported that embedded grammar instruction yields positive gains in writing. Researcher have reported gains when

- grammar instruction was prior to, during, and after writing (Hillocks, 1986);
- grammar instruction was an explicit, intensive part of writing instruction, which is essential for struggling writers (Graham, 2006; Hirschman, 2000; Saddler & Asaro-Saddler, 2009);

- grammar knowledge and use was executed in real writing practice (Crawford & Royer, 1935 ; Hillocks, 1986; Maize, 1954; Symonds, 1931; Weaver, Bush, Anderson, Bills, 2006);

- grammar was targeted when writing in a meaningful academic context (Maize, 1954; Mellon, 1969; Scott & Balthazar, 2010; Symonds, 1931; Vocke et al., 2012; Wyse, 2006);

- grammar was incorporated through the reciprocal language processes listening, speaking, reading, and writing (Mellon 1969; Miller & Ney, 1968; Saddler & Graham, 2005; Vavra, 1987);

- grammar was addressed at the sentence level to improve receptive and expressive sentence construction skills (Datchuk & Kubina, 2013; Eberhardt, 2013; Hunt, 1965; Saddler & Graham, 2005; Scott, 2009a; Scott & Balthazar; 2013; Scott & Koonce, 2014);

- grammar was a part of instruction to build and use metalinguistic awareness and knowledge when writing (Myhill, Jones, Lines & Watson, 2012; Myhill et al., 2013).

Sentence Combining

Sentence combining (SC) is research-based intervention that targets writing within a metalinguistic framework through the facilitation of the metalinguistic processes needed for conscious construction of written sentences (Scott & Nelson, 2009). SC is also a grammatically based intervention. SC exercises were developed by several applied linguists who drew upon the systematic process of combining "kernel" sentences in a manner similar to that used by Chomsky (1965) in his transformational grammar theory (Combs, 1977)**.** SC is the process of manipulating and rewriting kernel or short declarative sentences into new syntactically complex forms (Strong, 1976). In this intervention, writers increase their knowledge of sentence structures and concepts through routine and repetitive sentence formation and reformation (Saddler & Graham, 2005). In addition, during SC, a writer explores alternative sentence structure combinations that may or may not convey the same meaning or intent (Hillocks, 1986). Subsequently, the practice of SC can become a strategy for students to embellish, edit, or revise their writing (Saddler & Preschern, 2007).

SC has been praised as the better alternative to traditional grammar instruction (Andrews et al., 2006; Graham & Perin, 2007a, 2007b; Hillocks, 1986). SC has a large research base (Scott, 2009) which has resulted in SC becoming the most commonly recommended intervention in relation to syntax and writing (Eberhardt, 2013; Farrell, 2013). SC has been documented to improve written (a) sentence construction (Datchuk & Kubina, 2008; Scott, 2009a), (b) syntactic maturity (Combs, 1976; Hunt, 1965; Mellon, 1969; O'Hare, 1973), and (c) written quality (Saddler, Asaro, & Behforooz, 2008; Saddler, Behforooz, & Asaro, 2008; Saddler & Graham, 2005). Researchers have also deemed SC as a valid method of assessing, developing, and documenting growth in writing (Hunt, 1977; Nelson, 2013a; Scott et al., 2006).

In the past, SC was criticized for its lack of formal or conscious analysis of syntactic structure and its inductive practice of making judgments about grammar (Burkhalter, 1996). Weaver (1996) added that although SC utilized the practice of teaching grammar in context, it did not require the ability to explain unacceptable parts of sentences or then explain how to correct them using grammatical terms. Conversely, Strong (1986) opposed such claims and stated that these misunderstandings were common with SC, due to the lack of understanding of how to conduct SC activities. Strong proposed a metalinguistic approach to SC, where the aim would be to develop the awareness of the connection between spoken and written language, through reading sentences and thinking aloud, thus shaping one's prose to match one's intention. In sum, students would use the reciprocal language processes to determine the form needed to convey content for specific use. Moreover, Saddler (2012) stated that the two most important parts of SC were the initial creation of a sentence combination and then the evaluation of it. Saddler, like Strong (1986), supported the use of a metalinguistic approach with emphasis on explicit modeling of the language decisions students needed to make, as well as the how and why of these language decisions.

Two meta-analyses have established the credibility of SC as a valid method to improve writing quality over traditional grammar instruction. The first conducted by Hillocks (1986) indicated that SC treatment had gains yielding an effect size of .35 (small effect) as opposed to traditional grammar instruction with a negative effect size of -.30. Similarly, a second meta-analysis conducted by Graham and Perin (2007a, 2007b), indicated that SC treatment studies yielded an average weighted effect size of 0.50 (medium effect), as opposed to traditional grammar instruction with a negative effect of -.032. Additionally, a systematic review of 18 SC studies that met criteria for review concluded that for students between the ages of 5 and 16, SC was effective for improving syntactic maturity (Andrews et al., 2006).

The two most recently published SC studies replicated a SC protocol from a study conducted by Saddler and Graham (2005), which was part of the meta-analysis conducted by Graham and Perin (2007a). For both studies, the researchers employed a single-subject research design working with fourth-grade students with disabilities (SWD). In the first by Saddler, Behforooz, and Asaro (2008), the subjects participated in a peer-assisted SC instruction and achieved a range of 87.5-100% of non-overlapping data (PND) on sentence-combining tasks, as well as syntactic complexity (average length of t-units) and rating on a writing rubric of the quality of story writing. In the second by Saddler, Asaro, et al. (2008), the subjects participated in generalization training, specifically, parallel SC writing tasks and the use of a peer-editor checklist for story writing. The subjects achieved 100% PND data on taught sentence-combining constructions, as well as on the quality rating and number of revisions on story writing.

The relationship between SC instruction in writing and its effect on reading was a popular debate in the 1970's to 1990's (Stotsky, 1983). Stotsky (1975) believed that although there was no solid research-base to support the connection between SC instruction in writing and reading, the connection between syntactic knowledge and reading comprehension was inarguable. Neville and Searls (1985, 1991) believed that there was a relationship, but that it was difficult to prove with older students because standardized reading assessments were not commensurate with the level of syntax complexity evidence in the student's writing. This notion was further supported by the speculations that certain reading measures may be insensitive to the benefits of SC in writing made by both Fusaro (1993) and Wilkinson and Patty (1993).

The ambiguous results documented across the SC literature prevented the solidification of the theory that SC instruction in writing can also improve reading (Stotsky, 1983). Straw and Steiner (1982) conducted a study where the experimental group who participated in SC instruction made gains in listening comprehension ($p < .001$) and on a cloze (i.e., fill in the blank) reading assessment (*p* < .001), but not on a standardized reading assessment (*p* >.05). Similarly, Fusaro (1993) conducted a meta-analysis of 24 studies on the effect of SC instruction on writing and reading achievement. Fusaro reported that SC had a beneficial effect on reading comprehension when the criterion measure was a fill in the blank cloze test (mean effect size of .199), but SC did not have a beneficial effect on standardized reading comprehension tests (mean effect size = -.046). In contrast, other studies have revealed the opposite results. Evans, Venetozzi, Bundrick, and McWilliams (1986) revealed positive gains in reading comprehension on a standardized measure (*p* < .0081) for twelfth grade students who participated in SC instruction in writing. Likewise, Wilkinson and Patty (1993) found that their subjects, fourth-grade students who participated in SC instruction in writing, performed with significantly higher scores on the Stanford Achievement Test (SAT) Reading Subtest at posttest (p < .02), but found no significance on two different fill in the bank cloze reading tests (p > .05).

There has been minimal research on SC instruction in writing and reading in the past two decades; however, current research is emerging. Scott (2009) has defended that the practice of talking and writing about sentence complexity may help students fluently recognize and deconstruct complex constructions that are essential for reading comprehension. Students who struggle with these skills will lag behind in reading in later elementary and secondary school, when reading text increases significantly in abstractness and complexity (Scott & Koonce, 2014). Students who struggle with reading are often suspected of having syntactic difficulties, regardless of whether or not they are diagnosed with language impairment (Scott & Balthazar, 2013; Scott & Koonce, 2014). Based on these ideas, Scott and Balthazar (in press) have utilized a language treatment protocol for children ages 10–14 years with specific language impairment called "Building Complex Language (BCL)." This protocol targeted production and comprehension of complex sentences, which will reveal newer data in relation to reading comprehension (Scott, 2014).

Chapter Summary

Over the last half century, the required skill set expected for U.S. workers has steadily increased. Today's workforce must sustain innovation to maintain economic prosperity in a technologically advanced society. Workforce and educational reform movements have continued to defend the right of all individuals to possess the skills they need to be successful in today's workforce. These skills include higher levels of literacy skills. Writing has become the most essential, pervasive, powerful, and permanent form of literacy. Yet, writing is the area of literacy that continues to receive the least attention, despite documented declines in pedagogical practice and student performance.

Current educational standards require that students have complex writing that can target a variety of audiences, genres, and purposes across academic disciplines. This level of writing is the anchor standard for college and career readiness, which is needed for highly specialized subjects such as science. Currently, the development of science is a major focus of government-funded educational reform; science education and science literacy are expected to improve as a result. Research, governmental, and educational agencies continue to develop the most effective pedagogical practices to foster academic growth in science. Therefore, effective writing education in science, as well as all other academic disciplines, is a preponderant factor that must remain at the forefront of these movements.

The literature has defended the need for advanced writing skills to be successful in college, career, and in life. For students struggling with writing, which are the majority of students, research needs to continue to explore new writing practices to meet higher educational standards. There is a significant gap in the research data on the effect of writing interventions in a specific academic discipline. The intervention conducted in this study incorporates several research-based practices. First is the use of a metalinguistic approach, which may aid students who struggle by developing a conscious awareness of the language underpinnings of writing in science. Second, is the use of sentence combining, which is an inherently metalinguistic intervention. In addition, sentence combining targets syntactic structure at the sentence level and embeds grammar instruction into a meaningful context. Moreover, sentence combining also uses the reciprocal processes of listening, speaking, reading, and writing. Metalinguistic and syntactic awareness will develop across the language processes, which may ultimately lead to better content learning. Last, the incorporation of explicit instruction in this intervention provides further scaffolds for students who struggle with literacy, so that they can learn, apply, and generalize their skills. This study aims to lend empirical support for the use of a metalinguistic writing intervention, metalinguistic sentence combining (MSC), with adolescent students who struggle with literacy, to improve their writing skills in science, as well as their understanding of similarities and differences related to science content. This study also aims to support the use of an explicit metalinguistic approach, which has been a factor missing from any sentence combining research thus far.

CHAPTER THREE: METHODS

This study investigated the effects of a metalinguistic approach to the writing intervention of sentence combining to increase students' metalinguistic awareness in science to improve written sentence complexity in science as well as the written expression and determination of comparison and contrast concepts of science content, with eighth-grade students who struggle with literacy. The study was conducted from mid-February to early April, spanning the third and fourth quarter of the school year. The methods employed in the study are reported as follows: (a) research design, (b) setting, (c) student and teacher participants, (d) sampling procedures (students and teachers), (e) instrumentation, (f) assessment, (g) procedures, (h) intervention, (i) data analysis, and (j) fidelity of implementation.

Research Design

This study employed a quasi-experimental pretest posttest hierarchical design, with students nested in classrooms. Quasi-experimental designs are designs in which there is no random assignment of participants to each condition (Campbell & Stanley, 1963). This lack of randomization can result in stronger threats to internal validity due to confounding variables that may influence treatment outcomes (Edmonds & Kennedy, 2013). Quasi-experimental designs require the researcher to make plausible inferences that require logic, design, and enumerative data analysis to compensate for potential confounding variables when interpreting results

(Shadish, Cook, & Campbell, 2002). However, quasi-experimental designs can reveal useful information, particularly for educational research (Gall et al., 2006). It is often difficult to have true experimental control in natural social settings such as a school; however, quasi-experimental studies can provide information about the effectiveness of an intervention within its natural setting and conditions (Campbell & Stanley, 1963). The intervention was designed to provide intensive writing instruction to students who are struggling with literacy in the standard curriculum. The researcher was able to collect a convenience sample (Edmonds & Kennedy, 2013) of students struggling with literacy, from a middle school in Central Florida. The researcher conducted the intervention with an entire class during their typical science instruction. The school had only a select number of eighth-grade science classes designated by the school administration to be for students who struggle with literacy. Therefore, the sample could not be randomized. The specific sampling procedures are discussed later in this chapter.

Setting

This study took place in a middle school (grades 6-8) in Central Florida. The participating classes were the five eighth-grade science classes with students who struggle with literacy. A students who struggles with literacy was defined as a student who has scored below a minimum achievement level score of 3, or below a developmental score of 228 (171-227) on the 2014 Florida Comprehensive Assessment Test (FCAT) in reading. The school administration grouped these students together in specific science classes that utilize the same curriculum, but move at a slower pace. In these classes, the teachers were aware that further support in the area of literacy related to the subject content would be needed. The experimental classes received intervention in their assigned classroom during regularly scheduled classes on the school campus. The comparison classes received their typical science instruction. All testing was conducted in a quiet classroom environment, also on the school campus, during regularly scheduled class time. No intervention or assessment took place outside of regularly scheduled class time or off the school campus.

District Demographics

District demographic data from 2014 reported a population of 191,599 students from preschool to grade 12. Students enrolled in the district were identified as 37% Hispanic, 29% Caucasian, 27% Black, 5% Asian, less than 1% Native American, Pacific Islander, or ‘other’. The district had 17.5% of students enrolled in the Exceptional Student Education (ESE) Program, which included both special education and the gifted education programs. Of these ESE students, 62.5% were in special education and 37.5% were in gifted education. Ten percent of students in the district were identified as English Language Learners (ELL). Forty percent of students in the district were enrolled in the Free or reduced lunch (FRL) program (Florida Department of Education [FLDOE], 2015).

School Demographic Data

School demographic data from 2014 reported an estimated population of 974 students in grades six through eight. There were 334 sixth graders (34%), 333 seventh graders (34%), and

307 eighth graders (32%). There were a total of 464 (48%) male students, and 510 (52%) female students. In eighth grade specifically, there were 149 (48.5%) male students, and 158 (51.5%) female students. There were 476 (48%) students in the school who identified as Hispanic, 217 (22%) students who identified as Black, 209 (21%) students who identified as Caucasian, 47 (4%) students who identified as Asian, and 5% of students who identified as Middle Eastern, Native American or 'other'. Approximately 137 (14%) of students in the school were enrolled in the ESE program. There were 83 (60.5%) ESE students in special education, and 54 (39.5%) ESE students in gifted education. The total number of students in the ELL program was 102 (10%) students. Approximately 672 students (64%) were enrolled in the FRL program. School and demographic (2014) data are presented in Table 1 (FLDOE, 2015).

Table 1: *School and District Demographic Percentages (2014)*

	Gender		Caucasian	Hispanic	Black	Asian	Other	ESE	ELL	FRL
	Male	Female								
Grade 6	46%	54%	19%	49%	25%	4%	3%	-	-	-
Grade 7	46%	54%	26%	43%	23%	6%	2%	-	-	-
Grade 8	51%	49%	20%	55%	18%	5%	2%	-	-	-
School	48%	52%	22%	49%	22%	5%	2.3%	8%	10%	64%
District	51.5%	48.5%	29%	37%	27%	5%	2%	17.5%	10%	40%

Note. ESE=Exceptional Student Education, ELL=English Language Learner, FRL = Free and reduced lunch program

Students

All student participants were enrolled in the selected middle school. The school had five eighth-grade science classes for students who struggled with literacy. A total of 84 eighth-grade students, across the five different eighth-grade classes, participated in the study. These classes were similar, with a range of 13-20 students. The number of study participants was 27% of the total population of eighth-grade students at the school. The students ranged in age from 12 to 16 years old. The students across the classes also varied in English Language Learner status, Exceptional Student Education status, and eligibility for the Free or reduced lunch program. Exceptional student eligibility labels across the experimental and comparison groups included 1) specific learning disability, 2) language impaired, and 3) other health impaired.

Demographic data were collected for all the participating students. Fifty-five percent of the sample was male students and 45% was female students. The majority of students across groups identified as 13 years of age (26%), 14 years of age (57%), and 15 years of age (15%). There was one student who was 12 years of age (1%), and one student who was 16 years of age (1%). Both groups presented with a large number of students who participated in the free and reduced lunch program (FRL, 80%), and FRL was similar between treatment groups (experimental group 78%, comparison group 81%). The number of English Language Learner (ELL) students (experimental group 30%, comparison group 30%) and Exceptional Student Education (ESE) students was also similar between treatment groups (experimental group 16%, comparison group 21%). Table 2 presents the demographic data for all student participants and by treatment group.

Table 2: *Demographic Characteristics for Student Participants*

Variable	Experimental Group (*n* = 37)		Comparison Group (*n* = 47)		Total Sample (*n* = 84)	
	n	%	*n*	%	*n*	%
Gender						
Male	19	51	27	57	46	55
Female	18	49	20	43	38	45
Age						
12	0	0	1	1	1	1
13	10	27	12	26	22	26
14	20	54	28	60	48	57
15	6	16	6	13	12	15
16	1	3	0	0	1	1
Socioeconomic Status						
No free or reduced lunch (FRL)	8	22	9	19	17	20
Free or reduced lunch	29	78	38	81	67	80
English Language Learner (ELL)	11	30	14	30	25	30
Exceptional Student Education (ESE)	6	16	10	21	16	19

Teachers

All participating teachers were employed as full-time science teachers who had at least one eighth-grade science class for students who struggle with literacy on their teaching schedule. The four eighth-grade science teachers who teach these five science classes agreed to participate. Three of the teachers taught one class each, and one teacher taught two of the classes. The experimental group was comprised of two teachers and the comparison group was comprised of three teachers. The teacher who taught two of the classes was assigned to both the experimental group with one class and the comparison group with the other class.

Sampling Procedures

The researcher received approval from the University of Central Florida Institutional Review Board (IRB; see Appendix A). The IRB board deemed that the researcher did not need to obtain written consent from either the teachers or the students because the study was considered exempt educational research. The researcher met with the teachers to explain the research study and to ask them to participate. The participating teachers and students were not compensated in any manner for participating in the study. The participating school and school district were also not compensated in any way for agreeing to participate in the study.

For quasi-experimental designs that use an experimental and a comparison group, the more similar the experimental and comparison groups are in their composition and pretest scores, the more effective control the researcher has of the equivocality of interpretation of the nonrandomized groups (Campbell & Stanley, 1963). The main threat to internal validity with inequitable groups is the possibility that the group differences on the posttest were due to preexisting group differences instead of the treatment effect (Gall et al., 2006). Prior to the study, the researcher collected demographic information, reading and science test scores, as well as other pertinent information related to student instruction. In addition, the researcher collected demographic information about the teachers. The researcher assigned the five classes to either the experimental group who participated in the intervention or a comparison group who did not participate in intervention. This purposeful assignment was to achieve even distribution of various factors that could negatively impact internal validity such as history, maturation, selection bias, or a combination of other treatments (Edmonds & Kennedy, 2013). The experimental group students, experimental class teachers, and comparison class teachers were able to infer their group assignment based on the presence or absence of the researcher in their classes. The comparison students were not aware of their assignment of condition. Last, the researcher conducted baseline equivalency testing of each measure to determine the baseline distance between the pretest mean scores of the experimental and comparison groups. Baseline equivalency testing is discussed in more detail in the data analysis section of this chapter.

Students

Reading Scores

The majority of the participating students were assigned to one of the five science classes chosen for the study because they were students who struggle with literacy. Specifically, a student who struggles with literacy was defined as a student who scored below the achievement score of a level 3 on the Florida Comprehensive State Assessment (FCAT) in reading the prior year. However, there were a total of 12 students (14%), who were placed in these classes who achieved a score of level 3 or higher on the FCAT reading, and one student (1%), who achieved a level 4. These students were placed in these classes because the students had enrolled in school after the school year had already begun. By that time, the other science classes had reached the class size limit (22 students) and were not available to these students. Ten students who participated in the study did not have 2014 FCAT reading scores in their school record. The experimental and comparison classes had an even number of students with FCAT reading scores of level 1 and 2. The comparison group had a slightly higher percentage of a score of level 3 on the FCAT reading (9%) than the experimental group (5%). The comparison group also had one student who achieved a score of level 4 on the FCAT reading assessment. Overall, the distributions of scores in reading were considered equitable between groups. Table 3 presents the descriptive data of the FCAT reading levels of each group. Figure 1 is a bar graph that represents the frequencies of reading level for each group.

Table 3: *FCAT Reading Levels (2014) by Treatment Group*

	Experimental Group (*n* = 37)		Comparison Group (*n* = 47)	
2014 FCAT Reading Level	*n*	%	*n*	%
1	16	19	17	20
2	14	17	14	17
3	4	5	8	9
4	0	0	1	1
5	0	0	0	0
No score reported	3	4	7	8

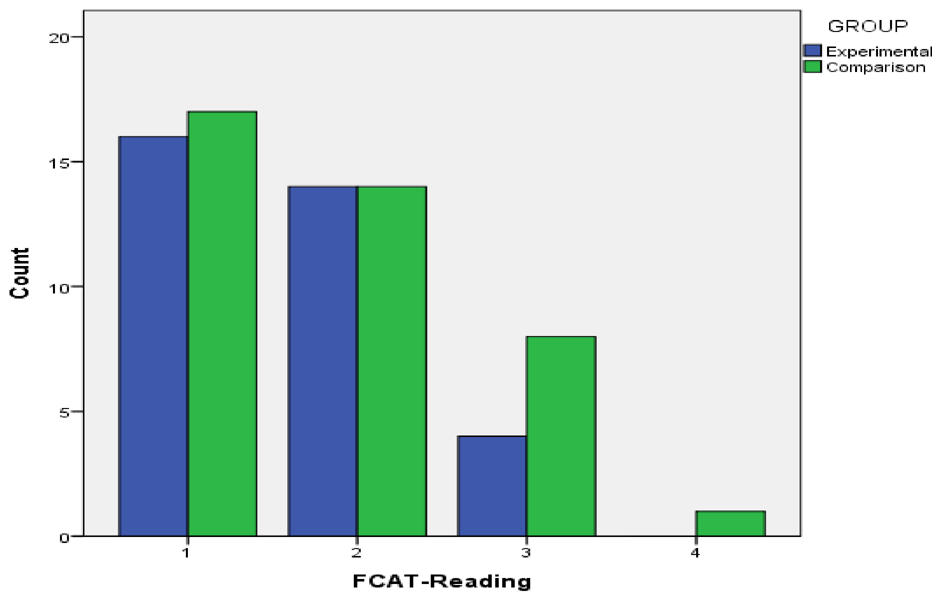


Figure 1: *FCAT Reading Levels (2014) Frequencies per Treatment Group*

Science Scores

A month prior to the intervention, all eighth-grade students had taken the Florida Science End of Course (EOC) Midterm Exam in Science. The test consisted of 25 multiple-choice questions worth four points each with maximum score of 100 points. The results revealed similar performance on this exam for both the experimental and comparison groups. Scores were only available for 79 students. Three students were missing data from the experimental group and two students from the comparison group as a result of being either absent or enrolled in a different district at the time of the exam. The experimental group results were M = 38.18, SD = 11.449, with a minimum score of 8, a maximum score of 60, and a range of 52. The comparison group results were M = 35.91, SD = 11.043, with a minimum score of 12, a maximum score of 64, and a range of 52. Overall, the distribution of scores in science was considered equitable across groups. Table 4 presents the results of the Science EOC Midterm Exam. Figures 2 and 3 represent histograms depicting the Science EOC scores for the experimental and comparison groups, respectively. The shape of the histograms suggested relatively normal distribution of scores for both groups.

Table 4: *Science End of Course Midterm Exam Scores (2015) for Study Participants*

	n	Missing Data	Mean	Standard Deviation	Minimum	Maximum	Range
Experimental	33	3	38.18	11.449	8	60	52
Comparison	46	2	35.91	11.043	12	64	52

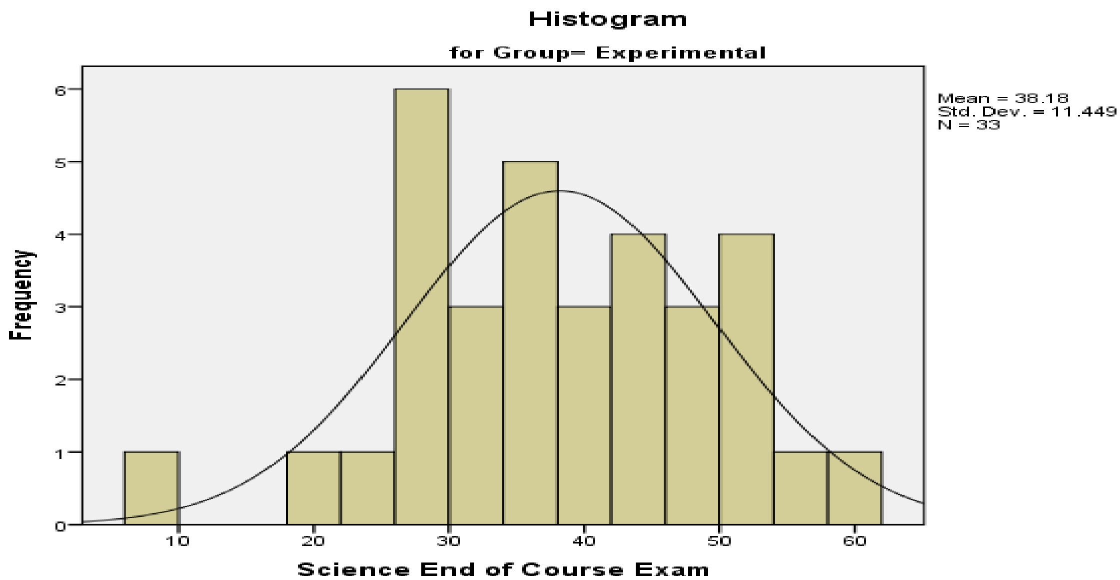


Figure 2: *Science End of Course Midterm Exam Scores- Experimental Group*

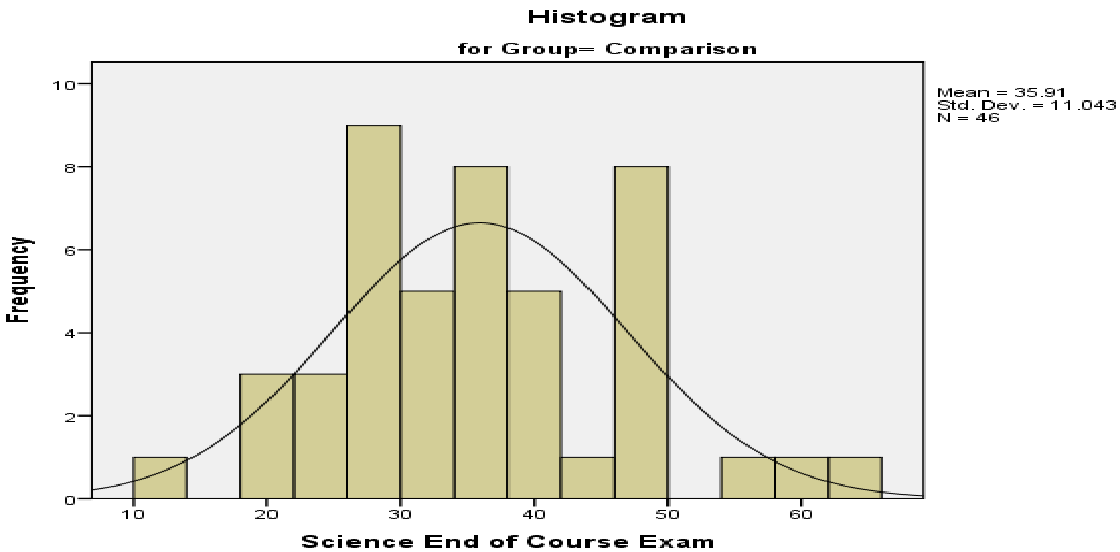


Figure 3: *Science End of Course Midterm Exam Scores- Comparison Group*

Other Instructional Factors

Two other factors related to student instruction were considered in group assignment: the amount of writing instruction and the amount of reading instruction that the students received during a typical school day. These two factors were considered to insure that group assignment yielded equitable treatment groups. Large differences between the groups could have threatened internal validity (Campbell & Stanley, 1963).

The first instructional factor that the researcher considered to insure equitable distribution of treatment groups was writing instruction. All students received a period per day of language arts instruction, which targeted writing skills. The school currently uses the SpringBoard ® English-Language Arts curriculum. SpringBoard® is a print and online program for all students in grades 6–12. The curriculum utilizes the critical features of: (1) academic vocabulary; (2) text-dependent questions to guide interactions with the text; (3) language and grammar instruction connected to in-context writing and text-related assignments in argumentative, informative, and narrative modes; (4) evidence-based written responses and oral discussions; and (5) a balanced selection of engaging nonfiction, literature, and non-print texts (College Board, 2015). The weekly writing assignments published in the SpringBoard® curriculum utilize the writing process method of planning, monitoring, drafting, revising, and editing (Flower & Hayes, 1981). After thorough review of the SpringBoard® curriculum and consultation with school administration and teachers, the researcher determined that there was no evidence to support that any form of sentence combining instruction was being addressed within the curriculum. Therefore, since all eighth-grade students participated in one language arts class per day, this instructional factor was considered equitable for all students in the study.

The second instructional factor that the researcher considered was the amount of reading instruction the students received during a typical school day. The amount of reading instruction could vary if the student was receiving intensive reading, ESE disability services, or ELL services. In relation to intensive reading, the majority of participating students were enrolled in intensive reading because they had achieved a score below a passing score of 3 on the 2014 FCAT reading assessment. Students who had scored a level 2 on the 2014 FCAT reading assessment received one period per day of intensive reading and students who scored a level 1 on the same assessment received two periods of intensive reading per day. In relation to ESE disability services, students may have also received “support facilitation services” in reading from a special education teacher. Support facilitation means that the special education teacher came into either the student's math or language arts classroom to provide further support in reading, which was specified on their Individualized Education Plan (IEP). There were also a few ESE students who were enrolled in a learning strategies class, which may have provided strategies to help with reading. In relation to ELL services, students do not receive any outside support services. However, these students do receive accommodations that may be related to reading to aid in their learning of academic content.

In summary, the following student factors were considered when assigning the classes to treatment groups: (a) class size, (b) time of day, (c) number of ESE students, (d) number of ELL students, (e) number of students receiving ESE in class support facilitation services, (f) number of students receiving learning strategies instruction, (g) number of students receiving one period per day of intensive reading, (h) number of students receiving two period per day of intensive reading, and the (i) number of students participating in the free and reduced lunch program. Table 5 outlines all the student factors considered for group assignment

Table 5: *Student Factors Considered in Treatment Group Assignment*

	Experimental Group (*n* = 36)			Comparison Group (*n* = 48)			
Student Factor	1	2	**Group average**	1	2	3	**Group average**
Class size	19	17	**18**	20	13	15	**16**
Time of day	a.m.	p.m.		a.m.	midday	p.m.	
ESE students	5 (26%)	2 (12%)	**19%**	1 (5%)	8 (62%)	2 (13%)	**26%**
ELL students	7 (37%)	4 (24%)	**31%**	4 (20%)	2 (15%)	5 (33%)	**23%**
Receiving ESE support facilitation services	2 (11%)	2 (12%)	**11.5%**	1 (5%)	3 (23%)	4 (27%)	**18%**
Receiving Learning Strategies instruction	1 (5%)	0 (0%)	**2.5%**	0 (0%)	0 (0%)	1 (7%)	**2.3%**
One period of intensive reading per day	7 (37%)	7 (41%)	**39%**	8 (40%)	3 (23%)	3 (20%)	**28%**
Two periods of intensive reading per day	6 (32%)	8 (47%)	**40%**	5 (25%)	6 (46%)	7 (47%)	**31%**
FRL	17 (89%)	12 (71%)	**80%**	17 (85%)	11 (85%)	11 (73%)	**81%**

Teachers

With regard to the teachers, the following characteristics were considered when assigning the teachers to treatment groups: (a) bachelor's degree major, (b) master's degree major, (c) years of teaching experience, (d) years of teaching experience in science, and (e) the amount of coursework or professional development in writing. Table 6 outlines all the teacher factors considered for group assignment.

Table 6: *Teacher Factors Considered in Group Assignment*

	Experimental Classes		Comparison Classes		
Teacher Factor	1	2	1	2	3
Bachelor's degree	Exercise Science	Chemical Engineering*	Education with minor in chemistry	Chemistry	Chemical Engineering*
Master's degree	Single subject teaching-Secondary Science	Middle School Science Education*	-------------	-------------	Middle School Science Education*
Teacher certification	Middle grades 5-9 integrated	Chemistry grades 9-12*	Chemistry and Spanish grades 6-12	Temporary middle school	Chemistry grades 9-12*
Years teaching	9	6*	23	.5	6*
Years teaching science	5	6*	23	Less than 1	6*
Coursework/ professional development in writing	One or more courses in college and one professional development course	One course in college and one professional development course*	One or more professional development courses	None	One course in college and one professional development course*

Note. * *indicates the same teacher assigned to both groups.*

Study Groups

Experimental Group

The experimental group consisted of 36 students from two different science classes, taught by two different teachers. One class had 19 students and the other had 17 students, yielding an average of 18 students. One class was scheduled during the morning (4th period) and the other class was scheduled during the afternoon (6th period). The combined group of students represented an average of 19% ESE, 31% ELL, and 80% FRL. An average of 11.5% of students received ESE support facilitation services in another class and an average of 2.5% of students received instruction in a learning strategies class. An average of 39% of students received one period of intensive reading per day and an average of 40% of students received two periods of intensive reading per day. The experimental group participated in the intervention during their typical science instruction for a total of 400 minutes (20 minutes for 20 sessions) at the beginning of their regularly scheduled science class.

Comparison Group

The comparison group consisted of 48 students from three different science classes taught by three different teachers. One class had 20 students, one had 13 students, and the other had 15 students, yielding an average of 16 students. One class was held during the morning (4th period), the second class at midday (5th period), and the third class during the afternoon (6th period). The combined group of students was an average of 26% ESE, 23% ELL, and 81% FRL. An average of 18% of students received ESE support facilitation services in another class (i.e., not science) and an average 2.3% of students received instruction in a learning strategies class. An average of 28% of students received one period of intensive reading per day, and an average of 31% of students received two periods of intensive reading per day. The comparison group participated in their typical science instruction.

Instrumentation

Standardized Assessment Measures

The TOWL-4 (Hammill & Larsen, 2009) was used to assess sentence combining skills. The Sentence Combining Subtest of the TOWL-4 is a 22-item measure to assess a student's ability to integrate the meaning of several short sentences into one grammatically correct written sentence. Each item is worth one point if the student provides one of the acceptable answers stated in the test manual. A raw score is determined based on the total of all the accumulated points, and then converted into a standard scaled score and percentile rank. A standard score of 10 is the mean, with a standard deviation of plus or minus three. In other words, a standard scaled score of 13 or more points is above average, 8-12 points is considered in the average, and 7 points or lower is below average. Hamill and Larsen (2009) established both reliability and validity for the TOWL-4. In order for a test of this kind to be considered minimally reliable, the reliability correlation coefficient must approximate or exceed .80 in magnitude; coefficients of .90 or higher are considered the most desirable (Linn & Miller, 2005). The authors established reliability for the TOWL-4 Sentence Combining Subtest through an average coefficient range of

(r = .87) indicating internal consistency or that this subtest is a consistent, repeatable measure (Campbell & Stanley, 1963). Furthermore, the authors established content validity (r = .70) indicating that this subtest is an adequate measure of sentence combining (Lomax, 2007), as well as construct validity (r = .88), indicating that generalizations can be made from this subtest on the theoretical construct of sentence combining (Edmonds & Kennedy, 2013).

Formative Assessment Measures

A formative assessment is a question or task that aligns with student learning goals and determines a students' present level of understanding, to make instructional decisions (Furtak & Ruiz-Primo, 2007). The use of a variety of formative assessments can help evaluate students' content learning (Dodge, 2009). Teachers can monitor students' progress in writing across several dimensions such as ideation, organization, word choice, and conventions through the use of formative assessments (Graham & Harris, 2011). Moreover, formative assessments in science that require a written response have the potential to support student understanding of scientific content and processes (Keys et al., 1999).

The use of written prompts, expository essays, and graphic organizers has been documented in the research literature as effective formative assessments (Dodge, 2009, Graham & Harris, 2011). The researcher used these three types of formative assessments for this study. The first formative assessment was a science compare and contrast writing prompt. The students were asked to provide two similarities and two differences between two science concepts. An example of a prompt was "tell two similarities and two differences between reflection and refraction of light." The data from the prompt were analyzed to determine if there was a difference from pretest to posttest in the use of more complex syntactic structures found in science such as sentence length, targeted connectives, embedding, agentless passive voice, and written productivity or correctness.

The second formative assessment was a science expository essay. According to Nippold (2012), the use of expository measures encourages the use of complex content-related text. The students were asked to write an essay about a science topic using specific details or facts about the topic. For example, the pretest essay asked the students to tell about their favorite science lab. The essay data were analyzed to determine if there was a difference from pretest to posttest in the use of more complex syntactic structures found in science such as connectives, embedding, agentless passive voice, and written productivity or correctness, as well as a change in the use of noun, verb, or prepositional phrases.

The third formative assessment was a compare and contrast double bubble map by Thinking Maps ®, which was used to measure the students' ability to determine the similarities and differences between two science concepts. The use of Thinking Maps® was implemented at the participating school and was used by some of the science teachers as a formative measure of content knowledge. The research on Thinking Maps® has indicated that the maps encourage students to focus on the processes used to produce the "correct" answer through strategic thinking (Long & Carlson, 2011). The use of a graphic organizer such as the double bubble map has also been indicated as a tool that allows students to generate knowledge, with less reliance on the language needed to convey knowledge (Marzano, 2010). For this study, the students had to write down answers that demonstrated their ability to determine the similarities and differences between the two science concepts presented, without the cognitive demand of putting this information into complete sentences. The students were instructed to list three differences among each the concepts on either sides of the map (6 answers), and then three similarities down the middle of the map (3 answers). A completed map would have nine answers or bubbles filled (see Appendix B). The researcher collected data to determine if there was a change in the total amount of correct answers they provided on the map from pretest to posttest.

It is important to note here that the use of the double bubble map was recommended to the researcher by the teachers as a way to determine “science knowledge without language.” The science teachers’ misperception was that the double bubble map is a non-language or nonlinguistic task is also a misperception that appears in the current research literature. The research literature does not acknowledge either the metalinguistics aspects of a Thinking Map, or the need for accurate language within a Thinking Map. According to the research on Thinking Maps®, the purpose of using the double bubble map is to think and organize ideas related to comparison and contrast (Hyerle, n.d.). The research on Thinking Maps® also explicitly states that using the map develops the process of using metacognitive awareness of text structures. The research does not state any information about the metalinguistic awareness and skills that are developed or that the thought processes are related to the *language* of the concepts targeted. In addition, other researchers have referred to the use of graphic organizers, such as a double bubble map, as a nonlinguistic representation of knowledge (Long & Carlson, 2011; Marzano, 2010). According to Marzano (2010), the term "nonlinguistic" is referring to the actual drawing of the structure of a graphic organizer or map. However, Marzano does not explain that there is an underlying linguistic or language base in the creation and structure of the graphic organizer or map. Marzano (2010) does indicate that certain crucial information must be highlighted in the graphic organizer, but glosses over the fact that the organizer itself represents the organization of the linguistic aspects or language base of the depicted concept. For example, with the compare and contrast double bubble map, the map or graphic organizer represents the semantic knowledge of what is similar (placed down the center of the map) and what is different (placed on the outer sides of the map). The structure also dictates that the language contained within that organization structure clearly depicts similarities and differences.

Due to misperceptions about the compare/contrast double bubble maps, as well as other possible bias (i.e., wanting the students to perform well), the researcher was concerned that the teachers may be lenient when grading the science compare/contrast double bubble maps. Therefore, in order to prevent any threat to the validity and reliability of scoring the science compare/contrast double bubble maps the research created a comprehensive answer key and had the maps scored by two unbiased raters. The researcher employed a four-step process to create the answer key. First, after the pretesting and posttesting, the four eighth-grade science teachers were given a random sample of maps to grade. The researcher blinded the identity of the students and their group assignment prior to the grading process. The teachers were instructed to give one point for each bubble that had a correct answer, without any partial credit. The total score would be zero points minimum and nine points maximum. Next, the researcher tallied which answers the teachers unanimously marked as correct or incorrect and which were not. To achieve final consensus for the answer key, a fifth science teacher from a different school (with a Master's degree in teaching and years of experience teaching science and language arts) determined which of the pooled answers were acceptable and unacceptable. This fifth teacher focused on the language needed to convey the correct answer (i.e., the use of negatives, antonyms or synonyms, or specific verbs, adjectives, adverbs that clarify a similarity or a difference). After that, the researcher created a formal answer key. Finally, the two unbiased raters graded the science compare/contrast double bubble maps according to answer key.

A jury of five teachers, four science teachers, and a school administrator with a teaching degree in science, assisted in the selection of the content used for both the science compare and contrast written prompt and the science compare and contrast double bubble map by Thinking Maps ®. The researcher compiled a list of 12 science compare/contrast benchmarks that would covered according to the district's eighth-grade science scope and sequence plan. Specifically, six science compare/contrast benchmarks that would be covered within one month of the pretesting and six science compare/contrast benchmarks to be covered within one month of the posttesting. Next, the researcher gave the list to the jury who determined which of the benchmark goals would be covered within those time periods. The researcher compiled each of the teacher's selections, then chose four compare/contrast topics at random to be used as part of the assessment process. Two compare/contrast concepts were used for the pretest and posttest science compare/contrast writing prompts and two for the science compare/contrast double bubble maps. For the pretest science compare and contrast writing prompt, the students were required to write two similarities and two differences between reflection and refraction of light. For the posttest, the students were required to write two similarities and two differences between physical and chemical properties of matter. For the pretest double bubble map, the students had to write the similarities and differences between potential and kinetic energy. For the posttest, the students were required to write the similarities and differences between an observation and an experiment.

Curriculum Based Measures

A Correct Incorrect Word Sequence count (Breaux & Frey, 2009) was used to analyze the sentences from the science compare and contrast writing prompt and the science expository essay. According to Breaux and Frey (2009), Correct-Incorrect Word Sequences (CIWS) has traditionally been used as a curriculum-based measure of grammar and writing conventions (i.e., capitalization, punctuation, and spelling). A "correct word sequence" is considered two adjacent words that are written with proper capitalization, punctuation, spelling, grammar, and meaning (Videen, Deno & Marston, 1982). Credit is given for initial capitalization and ending punctuation. The researcher and research assistant tallied the total number of incorrect sequences and the total number of correct sequences, and then subtracted one from the other to get the CIWS count. Past research has established validity coefficients ranging for CIWS counts from .56 to .80 and inter-scorer agreement reported between .88 and .92 for middle school expository writing using CIWS counts (Espin et al., 2000; Espin, De La Paz, Scierka, & Roelofs, 2005). The use of CIWS counts has been documented as not only as a reliable indicator of progress monitoring of writing performance (Amato & Watkins, 2010; Breaux & Frey, 2009), but as sensitive to changes in writing proficiency over time (Espin et al., 2005).

Coh-Metrix 3.0 (McNamara, Louwerse, Cai, & Graesser, 2013) is a computerized system for calculating computational cohesion and coherence metrics for written and spoken texts. Text is categorized into the text genre categories of narrative, information, or science text prior to analysis with Coh-Metrix 3.0. The "science" genre category was used for this study. The students' written samples were typed into the program, and then the program generated a variety of measures and indices. The Coh-Metrix measures were counts such as the amount of words, the amount of sentences or average sentence length in the sample (Graesser et al., 2004). For this study, Coh-Metrix 3.0 was used to determine the measure of mean sentence length for the science compare and contrast writing prompt. The Coh-Metrix indices calculated numbers or incidence scores in relation to word and sentence frequency, parts of speech, logical operators (i.e., clause, phrase, and sentence types; location; and length), connectives, and readability (Graesser et al., 2004). For both the science compare and contrast writing prompt and science expository essay, an incidence score for the total the average of all connectives was computed. An incidence score is the number of occurrences per 1000 words (McNamara, Graesser, McCarthy, Cai, 2014). To measure the written sentence complexity for the science compare and contrast writing prompt the following were computed to determine changes in the use of connectives: (a) causal, (b) logical, (c) adversative/contrastive, (d) temporal, and (e) additive connectives. For both the science compare and contrast writing prompt and science expository essay, certain aspects of syntactic complexity were computed that are characteristic of the complex sentence structures found in science. The first was the incidence of left embeddedness, or the mean number of words before the main verb clause and the second was the incidence of agentless passive voice. For the science expository essay, the incidence of noun phrases, verb phrases, or prepositional phrases was used to determine any change from pretest to posttest. Noun, verb, and prepositional phrases were used in the intervention and were inherent in the science content throughout all phases of the intervention protocol.

Informal Measures

Language sample analysis

Language sample analysis, although an informal measure, is as an effective way to document adolescents' written language performance in its natural context (Nippold, 2010). The researcher and research assistant completed language sample analysis to determine the number targeted connectives and morpho-syntactical errors. For the science compare and contrast writing prompt, the number of times each of the four targeted connectives (both, like, but, however) occurred was tallied and totaled. For the morpho-syntactical errors, a key of common morpho-syntactical errors was compiled based on previous research. Students who struggle with literacy, or have impaired language, may demonstrate difficulty with maintaining subject-verb agreement or verb tense, using copula or auxiliary verbs, and marking morphemes such as plurals and possessives (Dockrell, 2014; Eisenberg, 2013; Scott, 2004, 2012; Scott & Balthazar, 2013; Scott & Windsor, 2000). Each incidence of a morpho-syntactic error that occurred in the students' writing was counted. The researcher and research assistant looked for morphemes that were used incorrectly, inconsistently, or were omitted. Each type of error was totaled, and then a percentage was determined based on the total number of all types of errors. Morpho-syntactical errors were totaled for both the science compare and contrast writing prompt and science expository essay. The morpho-syntactic categories tallied were: (a) plurals, (b) possessives, (c) prepositional phrases, (d) articles, (e) subject-verb agreement (singular and plural form), (f) present tense verbs, (g) past tense verbs, (h) future tense verbs, and (g) verbs in the perfect tense.

Teacher writing survey

The researcher developed two different surveys to be given to experimental and comparison teachers, the Experimental and Comparison Teacher Pre-Survey and the Experimental and Comparison Teacher Post survey. Both surveys asked questions related to the amount of student writing time, the type of student writing activities, the type of writing activities that are modeled by the teacher, the type of writing required for test taking, and the type of writing errors made by students. The teachers were also asked their perception of the importance writing in science class. In addition to questions, the survey collected demographic data about the teachers' (a) college degree and major, (b) years of teaching experience, (c) years of teaching experience in science, and (d) number of completed courses or inservices in the teaching of writing. The pre-survey consisted of 10 multiple-choice questions and one open-ended question (see Appendix C). The post-survey consisted of questions asking the same information in an open-ended format. The format of the survey was changed to document similarities or differences across responses when given choices or the ability to respond freely, as well as to prevent familiarity of the survey from pre to post-survey for the teachers (see

Appendix D). The researcher pooled all the results to determine the total percentage of each type of response at pretest and posttest.

Social validity surveys

The experimental teachers completed the Experimental Teacher Social Validity Post Survey. This survey was created by the researcher. This survey consisted of six open-ended questions used to gauge the experimental teachers' perception of the MSC intervention (see Appendix E). For example, the experimental teachers were asked questions such as "Do you think MSC had a positive impact on student writing in science?" or "What are your thoughts regarding the amount of time MSC required in science class?"

The students were given a Student Pre-Survey and the Student Post-Survey that was created by the researcher. Both surveys asked students to rate agreement and disagreement on a Likert scale of 5 (strongly agree) to 1 (strongly disagree), as well as answer two open-ended questions. The pre-survey (see Appendix F) asked general questions about students' perception of writing. The post-survey (see Appendix G) asked the same questions related to the students' perception about writing, and then some specific questions about their perception of the intervention. For example, some of the statements students were asked to rate on the Likert scale were "I like to write", "I think writing is important in science," and "I think metalinguistic sentence combining (MSC) helped me to read and learn the science concepts better." The researcher pooled all the results to determine changes in extent to which students agreed or disagreed with each statement from pre-survey to post-survey. The total percentage of each type of response at pretest and posttest was compared.

Intervention Protocol

Sentence-Combining Exercises

The intervention protocol was comprised of two elements, sentence-combining exercises and a metalinguistic approach. The metalinguistic approach used with the protocol is described later is this chapter. **S**trong (1986) and Saddler (2012, 2013) have published recommendations for sentence-combining protocols, which were used when creating this intervention protocol. The general recommendations for sentence-combining protocols utilized in this study were (1) the use of kernel sentences to combine into complex sentences; (2) the use of specific written prompts; (3) an errorless approach to instruction; and (4) use of an explicit, scaffolded, interactive, and metalinguistic instruction. Saddler (2012) has also specified that sentence combining is a process of (1) combining, (2) changing, (3) adding, and (4) rearranging (Saddler, 2013).

The first recommendation by Strong (1986) and Saddler (2013) for sentence-combining protocols is the use of sentence-combining exercises that involve the combining of kernel sentences. The researcher developed the exercises with the predetermined kernel sentence sets that the students combined. The kernel sentences contained information for comparing, contrasting, and elaborating science text from the previous or current science unit.

The second recommendation for sentence-combining protocols by Strong (1986) and Saddler (2012, 2013) is to use specific written prompts to guide the sentence-combining exercises. Prompts can be classified as “open”, “structured”, “cued”, and “closed.” Open prompts teach stylistic options that the students can choose in context. Structured prompts allow students to make choices from a limited set of correct choices. Cued prompts give specific hints to help to teach students make targeted transformations (Saddler, 2012). Closed prompts limit student choices, so students use correct syntactic structures (Strong, 1986).

The researcher used four different types of prompt combinations. The first was the use of a structured open prompt to help the students remember the targeted conjunctions or connectives that could be used for comparing and contrasting. The researcher provided the students with a "visual word bank" to help them recall the four conjunctions. For comparison, or stating how information was similar, the word bank listed "both" and "like." For contrast or stating how information was different, the word bank listed " but" and "however." The researcher selected the connectives "both" and "but" because they were frequently used in the science textbook, and the connectives "like" and "however" because they are complex syntactic forms that typically develop during adolescence. Students who struggle with literacy use a limited range of conjunctions in their writing (Scott & Koonce, 2014) and may rely on the same conjunctions throughout a writing passage (Scott, 2004). The use of the structured open prompt or visual word bank allowed the students to attempt to use a variety of conjunctions.

The second type of prompt used to guide sentence-combining exercises was a type of closed prompt called a "restrictive" prompt. The students were not able to use the word "because" as the connector between the sentence combinations. The restriction on the word "because" prevented the students from implying cause and effect relationships. Although cause and effect is prevalent in science text (Fang et al., 2006), compare and contrast was the focus of the exercises. Additionally, students with syntactic weakness will demonstrate overuse of certain conjunctions such as "because" (Scott & Balthazar, 2013). The students were reminded not to use the word "because" as part of the written instructions for each intervention session's set of exercises.

The third type of prompt used to guide sentence-combining exercises was the closed prompts of underline and parenthesis cue. Strong (1986) and Saddler (2012, 2013) have both proposed the use of the combination of underline and parenthesis cues to target the use elaborative and embedded information. When the students were given three sentences the third sentence became either a relative or an adverbial clause that would be combined with the other two sentences. The researcher underlined the portion of the third kernel sentence that must remain as an embedded or elaborative clause. A parenthesis cue with the relative pronoun or subordinating conjunction needed to combine the third sentence was placed at the end of the underlined portion of the third sentence. The relative pronouns used were "that", "which", and "who" for relative clauses. The subordinating conjunctions used were "when", and "after" for adverbial clauses. These closed cues allowed for practice of sentence embedding or sentence expansion without the cognitive, semantic, and syntactic demands that would be required without these cues. For example, when given the three sentences, (1) Elements cannot be altered by physical changes, (2) Physical changes cannot alter compounds, (3) <u>Physical changes happen in the environment</u> (that), the last sentence would become a relative clause. The sentence combination, combined with one of a targeted conjunction for similarities (e.g., "like") could be

“Compounds, like elements, cannot be altered by physical changes that happen in the environment” (elaborative relative clause). Another option using the conjunction “both” could be **“**Physical changes that happen in the environment cannot alter both elements and compounds” (embedded relative clause).

The third recommendation for a sentence-combining protocol by Strong (1986) and Saddler (2013) was to use an errorless approach of instruction. For all of the exercises, the students were encouraged to explore various options and ways that they could combine the sentences. The exercises promoted problem solving and language processing through interaction with the text. The focus of the instruction was not on one specific answer, but the options, manipulations, and decisions made to create grammatically correct and accurate sentences. In addition, beginning at intervention session four, the students were given the opportunity at the beginning of each intervention session to edit their sentences from the previous session. The editing process utilized in the protocol is discussed later in this chapter. The use of the editing process allowed the students to refine the two processes of changing and rearranging as specified by Saddler (2012). The process of editing also called further attention to the specific semantic, morpho-syntactic, and pragmatic choices made by students through their use of writing conventions (i.e., capitalization, punctuation, spelling), and word choices. Students who struggle with literacy have trouble comprehending and evaluating their writing errors (Singer & Bashir, 2004). Overall, the use of the errorless approaches of exploring and revising (editing) allowed students to build their confidence in selecting syntactical options and to control their writing by revising (O’Hare, 1973).

The last recommendation for a sentence-combining protocol by Strong (1986) and Saddler (2013) is that the instruction be interactive, explicit, scaffolded, and metalinguistic. Saddler (2013) specified in order to be interactive, the teacher should demonstrate what the students need to do while discussing why and how to make various sentence combinations. After the demonstration, the students should be guided to formulate multiple solutions with verbal and visual support. Then, the students should create solutions and discuss them with others. The discussion of solutions should be collaborative and could be facilitated by peer-supported interaction. Peer interaction has been documented as effective in the writing process (Graham, Harris, & Larsen, 2001; Graham & Perrin, 2007a; Rijlaarsdam et al., 2014; Saddler & Graham, 2005). Specific verbal and visual supports as well as peer-supported learning are within each phase of the intervention protocol, which is discussed later in this chapter.

The last aspect of sentence-combining instruction also mentioned instruction that was explicit, scaffolded, and metalinguistic. Students who struggle with literacy may need more repetitive and explicit instruction that targets metalinguistic awareness (Graham et al., 2001; Hillocks, 1986; Troia, 2011). The researcher developed a metalinguistic script that accompanied the instruction to provide explicit instruction and repetition using definitions, questions, and explanations (Saddler, 2010). These definitions, questions, and explanations surfaced the metaphonological, metasemantic, metamorphological, metasyntactical, and metapragmatic aspects of the sentences that may not be apparent to students who struggle with literacy (Paul & Norbury, 2012). Although Strong (1986) and Saddler (2013) described these metalinguistic aspects in their sentence-combining protocol recommendations, details on how to explicitly target metalinguistic aspects are absent from sentence-combining research study literature. The use of an explicit metalinguistic script was a unique aspect of this research study and is discussed later in this chapter.

Comparison and Contrast

The expository genre of comparison and contrast was used as the semantic purpose of the intervention. Researchers have suggested that instruction that focuses on a single text expository structure is more effective on students' comprehension of that expository text structure (Williams, Hall, & Lauer, 2004). The genre of comparison and contrast is difficult for students because it requires the ability to determine the similarities and differences, organize the information, and then relay the information in written form (Rijlaarsdam et al., 2014). The written form must express the similarities and differences in a clear, concise, and coherent manner, through appropriate linguistic choices such as the word order, the use of transition words between clauses and sentences, and the arrangement of sentences (Saddler, 2012). All of these linguistic aspects are necessary to convey to the reader what is similar and different.

In addition, determining similarities and differences is inherent in the logic and organization of other expository text structures (Strong, 1986) that used across academic subjects. For expository text that requires synthesis, argument, or persuasion, the writer must first organize and integrate information based on similarities and differences. The writer must then note these similarities and differences when forming conclusions, stating claims, and citing supporting or negating evidence. Detecting similarities and differences is also essential for detecting if a change has or has not occurred. If change has occurred, differences may be apparent. If a change has not occurred, then similarities may still be apparent. For the expository genres of cause and effect and problem and solution, change or lack or change is at the core of these concepts. Specifically, there needs to be an understanding of what is different or what has remained the same. An effect is a change that is a result or consequence of an action or other cause. A problem may be something that has changed (i.e., is different) or has not changed (i.e., is still the same); the solution may be to change or not make a change.

The expository structure of compare and contrast was used for this study for two reasons. First, the use of comparison and contrast as an effective instructional technique is recognized by research and by the participating middle school. A meta-analysis conducted by Haystead and Marzano (2009) revealed that the effect of the instructional technique of identifying similarities and differences yielded a percentile gain of 20% and had medium effect size of d = .52. Fifty-two different research studies had met criteria and were included in the analysis. The school administration had informed the researcher that they were using the Marzano Teacher Evaluation Model Learning Map (Marzano, 2014). This evaluation system was implemented by the Florida Department of Education as a research-based system to assess teacher effectiveness. One skill listed for students to practice and deepen knowledge is for the teachers to have students examine similarities and differences. This is an instructional practice that was expected of all teachers at the participating school.

The second reason why compare and contrast was targeted is because it is recognized as an essential expository structure in science by the 2013 NGSS and by the 2014 Florida Next Generation Sunshine State Standards, which were the standards followed by the participating middle school. In the 2013 NGSS, it has been stated that students need to analyze and interpret data by determining similarities and differences in their findings (NGSS Lead States, 2013f). Determining similarities and differences is considered within the NGSS crosscutting concept of providing explanations of stability and change in natural or designed systems (NGSS Lead States, 2013c). The NGSS also included comparing and contrasting of approaches, information, and findings within its Literacy Anchor Standards of the NGSS, which are in alignment with the CCSS (NGSS Lead States, 2013d). The Florida Next Generation Sunshine State Standards also list comparing and contrasting of a variety of science concepts as a benchmark skill throughout the eighth-grade physical science standards (FLDOE, 2014).

Science Content

The participating school used the Florida Science Fusion Holt McDougal Text Book for their science classes. According to Dispezio, Frank, Heithaus, and Ogle (2012), the authors structured the book around the Florida Next Generation Sunshine State Standards (NGSSS). The authors' goals were to build students' science literacy to prepare students for the concepts that will be assessed on the eighth-grade Florida Comprehensive Assessment Test (FCAT) in science. The students used the textbook to facilitate learning through the application of scientific facts, procedures, and inquiry-based learning. The text was based on the 18 "big ideas" from the NGSSS. These big ideas were listed in each unit; the lexile level of the text (average lexile 1050, mid-range expected between grades 6-8); and the NGSSS benchmarks, vocabulary, and key concepts were included. The textbook also contained sample test questions and application activities to promote critical thinking.

The student's science textbook was reviewed by the researcher. The review revealed the use of various conjunctions to express comparison and contrast, with "both" being the most common conjunction used to refer to a similarity and "but" as the most common conjunction to refer to a difference. The use of relative and adverbial clauses was prevalent in the science text. In the discipline of science, elaborative or embedded elements such as relative or adverbial clauses are used to present lengthy information in a cohesive manner (Fang et al., 2006). In science text, these types of clauses help clarify the meaning of a sentence with additional factual information (Justice & Ezell, 2008). The syntactic structure of the sentence combinations were the same as the science text, likely aiding the comprehension of this type of complex science text (Scott & Balthazar, 2010). The development of better skills in reading and comprehending of science text could also improve comprehension of the science concepts, specifically, comparison and contrast.

The researcher, in collaboration with the eighth-grade science teachers, created all of the metalinguistic sentence-combining exercises. The experimental students completed two exercises during each of the 20 sessions, for a total of 40 sentence combinations. The kernel sentences were directly related to the science content of the science unit under study. The researcher implemented a two-step process to develop the exercises. First, the researcher created a set of kernel sentences that could be combined to articulate similarities and differences of the science concepts or to compare and contrast. The exercises were inclusive of six different science units that would be covered over the seven weeks of the intervention. In developing these exercises, the researcher referred to the eighth-grade science scope and sequence plan created by the teachers, as well as the science text that matched the unit topic. The science units that were included in the exercises were: (1) physical properties of matter, (2) atoms and states of matter, (3) physical and chemical changes, (4) elements, compounds and mixtures, and (5) atomic theory.

The researcher followed a two-step process to insure the accuracy of the science context. First, the researcher created a set of possible exercises prior to the study. A jury comprised of three science teachers reviewed these exercises. Then, the teachers provided two types of feedback for the researcher. One type was feedback about whether the topic would be covered within the scope and sequence of their science curriculum, specifically, if the concept was a compare/contrast concept that was a targeted benchmark of the curriculum. The other type of feedback was related to the accuracy of the science content, such as the vocabulary and sentence structure. The second step of the process was completed prior to each week of intervention. The researcher confirmed with the two experimental teachers which science benchmark they had just covered. After that, the researcher chose exercises from the preexisting set that matched the science content. The researcher then gave the teachers the exercises to affirm that the content was correct and the wording was appropriate. The teachers also gave the researcher suggestions of vocabulary or particular phrases that could be targeted to further the students' understanding of the science content or concepts covered in class.

It should be noted that initially the science teachers were concerned with the use of passive voice, multiclausal sentences, and complex conjunctions such as "like" and "however" in the exercises. Although these were the structures that were evident in the science textbook, the teachers revealed that this was not the level of language that they used in teaching. The science teachers stated that they typically try to keep the language simple, so that the students can grasp the concepts without the demand of understanding complex sentences. To alleviate the teachers' concerns, the researcher agreed to use only science material in the exercises that had already been covered by the teacher in class. Hence, the exercises served as a review of the science content material. In addition, the researcher assured the teachers that the use of complex syntax structures would be accompanied with explicit instruction to facilitate student success, and that research supports instruction of more complex syntax structures to improve listening, speaking, reading, and writing.

Explicit Instruction

The researcher incorporated various instructional practices during the intervention to provide explicit instruction. These instructional practices have been recommended by researchers to use with students who struggle with learning (Deshler et al., 1981; Deshler & Lenz, 1989; Hughes, 2011; Schumaker, 1989). First, the researcher supplied an advance organizer, closing organizer, and a mnemonic for each session to help the students recall the knowledge and steps needed for the intervention. Second, the researcher employed a metalinguistic script to facilitate the metacognitive and metalinguistic aspects of sentence combining in an explicit and repetitive way. Third, the researcher provided specific levels of support with a gradual release of responsibility (Pearson & Gallagher, 1983; Schumaker, Deshler, Woodruff, Hock, Bulgren, & Lenz, 2006). Fourth, the researcher allowed the students to edit their work to check for understanding and provide explicit feedback. In addition, the researcher implemented motivational techniques as a means of praise and encouragement. The explicit instructional components, metalinguistic script, intervention phases, instructional levels of support, and motivational techniques are described in the next sections.

Advance and post-organizer

According to Schumaker (1989), an advance organizer helps to prepare the students for the intervention by stating the purpose of the intervention and relating the purpose and benefit to the students. The purpose was stated in this manner: "We will take simple sentences and combine them into mature, complex sentences to improve our writing. We read, think, and then, write." The advance organizer should also review previous learning and the goals and expectations of the current lesson. Each intervention session began with the advance organizer. A script of the entire advance organizer was on the front cover, back cover, and inside of the students' folders, so that the students could follow along visually while the researcher read or stated the advance organizer at the beginning of each session.

First, the researcher reviewed the students' role during the intervention and the purpose of the intervention (see Appendix H). The researcher stated, "You are the writer. You decide how you are going to say what you need or want to say. You are telling the reader how to read your message. You are the writer." Then, specific to the discipline of science, the researcher stated, "For this class we are going to be science writers. Science writers have to read, think, and discover the clues that tell about similarities and differences, and then write clear, correct, and mature sentences." Shanahan (2013) has stated that readers have to follow a writer's lead, for the writer "initiates the conversation" (p. 335). The researcher incorporated the theme of the power of a writer using a crown icon. This theme was carried throughout the intervention.

Second, the researcher reviewed the definitions of parts of speech as background knowledge needed for the intervention. A key function of this review was to guide the students to select the rules that were most appropriate for meeting the demands of a task (Deshler & Lenz, 1989). In addition, students who struggle with literacy need explicit reminders and practice with basic grammatical rules in relevant contexts (Eberhardt, 2013; Fey, Long, & Finestack, 2003; Gersten & Baker, 2001; Rogers & Graham, 2008). This need was addressed by utilizing the imagery of a crown of knowledge needed for success. The students were told to "Put Your Crown On." The tips of the crown stated NVAA, which stood for nouns, verbs, adjectives, and adverbs, which were the four parts of speech targeted. (see Appendix I). Third, the researcher stated or reviewed the specific goal of the day, which was "Combine two sentences into one sentence to compare or contrast science information." The last part of the advance organizer was the researcher stating or reviewing the mnemonic to help the students remember the steps for

MSC. The researcher then read aloud the instructions at the top of the MSC exercise worksheet along with the sentences the students would be combining that day (See Appendices J, K, and L). The following instructions were given:

- Combine these two sentences into one sentence (Phase One and Phase Two), or Combine these three sentences into one sentence (Phase Three).
- You cannot use the word “because.”
- You must use all the science information in your new sentence, but you can add or change words.
- You must use one of the connector words in the box below (both, like, but, or however).
- You must keep all underlined portions of the third sentence. (Phase Three only)
- Connect the third sentence with the word in parentheses. (Phase Three only)

The researcher used a post-organizer at the end of each intervention session, stating the goals for the day and summarizing the activities that were completed. The researcher then stated the goals for the next session. The post-organizer served as a review of the information and as direction for future learning (Schumaker, 1989).

Mnemonic device

The use of mnemonics in explicit instruction facilitates the memorization of the interventions steps and explanation (Deshler & Lenz, 1989). The mnemonic created for this study, WRITEr. , incorporated all the recommendations prescribed by Deshler and Lenz (1989), in that each step should (a) be succinct with familiar language, (b) start with a verb that is directly related to the mental or physical action that the step will cue, (c) be fewer than seven steps, and (d) be related to the overall process that will be addressed. Following these recommendations, the researcher created the mnemonic ♕WRITEr. (see Appendix M). The mnemonic was used as a means to recall the step-by-step procedure of the MSC protocol. The mnemonic was not used with the intention of teaching a strategy, but as an aid to develop skills. However, the use of a mnemonic can lay the groundwork for students to recall and access the step-by-step procedure as a strategy in the future.

The “W” stands for WORD, “Word your message to achieve your goal.” During this step, the students were encouraged to think about the words they needed to achieve their goal. Since the goal was to compare and contrast, the conjunction chosen was related to the kernel sentences as similarities or differences and the key words or phrases that signaled the similarities and differences. The researcher also called attention to how these word choices were dictated by the discipline of science. Within a disciplinary literacy framework, instruction should explicitly teach the differences in semantic-syntactic content when reading and writing in that discipline, so that students will learn these differences and differentiate their reading and writing across disciplines (Shanahan & Shanahan, 2012). Students who struggle with literacy are unable to organize text conceptually due to their inability to find the words that signal concepts (Scott, 2009a; Scott & Koonce, 2014). Overall, this step helped the students to organize the text conceptually within the parameters of comparison and contrast.

The letter “R” is for REMOVE, “Remove words you do not need. This step referred to the decisions the students had to make to determine which words and phrases were needed and not needed to combine the kernel sentences into one sentence. Part of the sentence-combining process involves eliminating sentential redundancies and replacing them with succinct complex syntax structures (Scott & Nelson, 2009). Furthermore, the students learned to tease out relevant sentence content, which may have improved their ability to parse information when reading (Scott, 2004, 2009a).

The letter “I” is for INTEGRATE, “Integrate the words into a new sentence.” This step referred to the process of the students putting all the necessary portions of the kernel sentences back together into one sentence. By manipulating the information into the final sentence combination, students developed the metalinguistic awareness (Scott & Nelson, 2009) and skill needed to write complex information with which they would struggle independently (Nelson, 2013b).

The letter “T” is for TEST, “Test your sentences.” This step had two parts. For the first part, the sentence combination was read aloud, so that the students could “hear” if the sentence sounded correct. The researcher called attention to the aspects of the sentence that needed to be changed to insure that the sentence combination was grammatically correct. For the second part, the researcher called attention to the main subject and predicate to check for number and tense agreement, as well as in any embedded or dependent clauses during Phase Three (The phases are discussed later in this chapter). One of the most common errors for students who struggle with literacy when writing is not maintaining subject-verb agreement and verb tense (Farrall, 2013; Scott & Balthazar, 2013).

The last part of the mnemonic is "Er.", which calls for students to ErASE, "Erase your mechanical errors." The "E" which is an uppercase letter signaled the use of a capital letter at the beginning of a sentence. The period at the end of "Er." signaled the use of a period at the end of the sentence. During this step, the students were encouraged to check their sentence from left to right to check for capitalization (i.e. capital letter at the beginning and correct and incorrect use of capitals needed for proper nouns), and punctuation (commas, semicolons, periods). Students who struggle with literacy have difficulty coordinating the complex process of writing thereby compromising writing conventions (Harris & Graham, 2013; Singer & Bashir, 2004). Researchers have proven that writing instruction that incorporates writing conventions reduces convention errors (Datchuk & Kubina, 2013).

Metalinguistic Script

The researcher created a metalinguistic script to ask specific questions and cue specific actions (see Appendix N). This procedure targeted the necessary cognitive processes, or metacognition, and the physical actions needed to follow the intervention. Research has suggested that students who struggle with literacy learning do not use the appropriate cognitive behaviors or actions to perform literacy tasks. Effective instruction for struggling students incorporates the use of explicit models and overt physical acts to help struggling students develop these skills (Schumaker, 1989).

The metalinguistic script was embedded within the steps of the mnemonic. The questions required the students to think about the phonologic, semantic, morphologic, syntactic, and pragmatic underpinnings of their writing, as well as the implications on the reader. All the cues referred back to the how the writing would affect the reader's comprehension of the message. The phonologic cues referred to how the spelling of a word would affect its meaning. The semantic cues were related to the science vocabulary the impact of word choice on the meaning of the sentence. The morphologic cues specified morphemes and called attention to how the morphemes changed the meaning of the word or sentence. The syntactic cues targeted the sentence order. For instance, the researcher discussed how the location of a conjunction or clause varied the meaning of the sentence. The pragmatic cues were related to the intent of the message and the reader's perception of the message. The use of the metalinguistic cues empowered the students to think as writers who have the power to influence their readers through conventions, word choice, and word order (Troia, 2011). The use of these specific metalinguistic cues was a unique aspect of this intervention, which is notably different from past sentence-combining studies (e.g., Saddler et al., 2008; Saddler & Graham, 2005,)

Just prior to following the steps of the mnemonic, the researcher reviewed the directions at the top of the MSC worksheet and read the sentences aloud. For the first step, WORD, the researcher asked the students to determine if the kernel sentences were expressing a similarity or a difference, and then how they made this determination (e.g., "This is a similarity because the only the noun differs between the two sentences and the rest of the sentence is identical"). Next, the researcher and students circled any key words that helped to answer this question (e.g., the word "not" signals a difference), then circled the top of the word bank to remember to choose a similarity of difference conjunction. For Phases Two and Three, the researcher and students marked next to the kernel sentences either an "S" to mark a similarity and a "D" to mark a difference. Then the researcher either stated or asked which conjunction to use and asked why (e.g., "We will use the word like to signal to the reader that this information is similar"). After that, the researcher and students circled the conjunction they used for the sentence combination. In Phase Three, an additional step was added. The researcher stated or asked the students about the third kernel sentence and its related part of speech. The researcher stated the purpose of the clause and its part of speech instead of introducing new terms such as relative or adverbial clause. Specifically, if it was a relative clause, the researcher discussed what the relative clause was modifying, which was the noun in the sentence (e.g., "This sentence is telling us more about what?...The scientist, which is the noun in the sentence. It's telling us more about the noun, like an adjective would"). If it was an adverbial clause, then the researcher discussed how the clause was modifying the verb by telling when or where (e.g. "This sentence is telling us more about the verb, it is telling about <u>when</u> the particles evaporated. It's telling us more about the verb, like an adverb would").

For the second step, REMOVE, the researcher asked the students to determine which words should or should not be removed. Then the students were asked "Why?" For kernel sentences that were similarities, the response was related to redundant information such as the same verb or prepositional phrase in both the kernel sentences that would not need to be repeated twice. For example the kernel sentences, "Elements cannot be altered by physical changes"; "Compounds cannot be altered by physical changes"; can be combined into "Both elements and compounds cannot be altered by physical changes." For kernel sentences that expressed differences, the response was related to information to signal the difference such as “can” versus “cannot.” For example, the kernel sentences, “Mixtures can be separated”; “Elements cannot be separated”, can be combined into, “Mixtures can be separated, but elements cannot be separated.” Typically, a negative word or difference in the verb signaled the difference between the kernel sentences. After determining similarity or difference, the researcher and students underlined the words that needed to remain in the new sentence combination and crossed out any words or phrases to be eliminated. The researcher asked why these choices were made, so the students had to state aloud their reasoning (e.g., “We removed those words because they are the same and not needed to be stated twice” or “We had to keep the word *not* in the sentence to signal the difference between the two properties”).

For the third step, INTEGRATE, the researcher stated or asked how to word the new sentence or the sentence combination aloud. The researcher and students crossed out the kernel sentence information as it was used in the newly combined sentence. The researcher finished this step by asking the students if all the science information had been placed into the newly combined sentence.

For the fourth step, TEST, the researcher read or asked for the sentence to be read aloud exactly as it was written. If there was a punctuation error or misspelling, the researcher read the sentence as written so the students could hear the reader’s (i.e., researchers) interpretation. For example, if a period was missing, the researcher did not drop her pitch at the end to signal that the author did not tell her the sentence was ending. If the sentence was missing a comma, the researcher read the sentence aloud without taking the necessary pauses to signal that something was missing. The researcher also asked the students "why" questions related to any changes that needed to be made and related the answers to the message that the writer needed to convey to the reader. For example, when asked, "Why was the comma needed?" the students answered with a statement such as "We needed to tell the reader to take a pause here or to emphasize that what is coming after the comma was different." Another example would be when the students were asked, "Why was it important to spell the word identity correctly?" the students answered with a statement such as "The mistake of the word *identify* is actually a verb and *identity* is a noun. This is incorrect and would confuse the reader." The second part to this step was to check for noun and verb agreement. The researcher pointed to each noun/verb pair, stated the pair in isolation, and then asked the students if the noun and verb matched (i.e., singular or plural form). After that, the students stated why the noun and verb matched and or did not match. For example, if the noun "elements" was paired with the verb "mixes" the researcher or students noted that this was incorrect, and then made an explicit statements such as "The word elements is plural, or more than one, so the verb needs to be mix. The word mixes would be paired with the word element which is singular or means one element."

For the last step, ERASE, the researcher checked the sentence starting from left to right and encouraged the students to do the same. The researchers then stated or asked the purpose of elements of capitalization and punctuation. The students would respond with answers such as the following examples: "The capital letter shows the reader that this is the beginning of my thought," "The comma tells the reader where to take a pause in this sentence," or "The period is there to tell the reader this is the end of my thought." Researchers have noted that developing conscious awareness of the semantic, syntactic, and pragmatic purposes of writing conventions can aid in the development of improved written conventions (Saddler, 2013; Scott, 2006; Troia, 2006, 2012).

Assessment Procedures

Pretesting and Posttesting

All students across participating classes that were present during pretesting dates, completed the pretesting assessment measures. The pretesting and posttesting testing dates were agreed upon with the teachers, principal, and assistant principal, prior to the start of the study. In order to prevent threats to internal validity, all the testing was administered by a team of four trained research assistants. The research assistants all had their bachelor's degree in Communication Sciences and Disorders. The researcher was present at the school, but not in the classrooms during the assessments. The first day of pretesting and posttesting included administration of the science compare and contrast double bubble map and the TOWL-4 Sentence Combining Subtest. The students were permitted to use 10 minutes to complete the science compare and contrast double bubble map and 25 minutes to complete the TOWL-4 Sentence Combining Subtest. The second day of pretesting and posttesting included administration of the science compare and contrast writing prompt and the science expository essay. The students were permitted to use 10 minutes to complete the science compare and contrast writing prompt and 25 minutes for the science expository essay. Pretesting days were on a Thursday and Friday in February, respectively. The posttesting days were on a Wednesday and Thursday in April, respectively. There was little flexibility with the testing and intervention schedule due to time constraints with spring break week, the Florida State Assessment in writing and the impending Florida Comprehensive Assessment (FCAT) in Science and Florida State Assessments in Reading and Math in mid-April. Although attempts were made to make up missed testing, there was no time available for students before or after school.

Intervention Procedures

Interventionist

To control for intervener effects, the researcher conducted all the intervention sessions. The researcher is a state licensed speech-language pathologist (SLP) and is certified by the American Speech-Language Hearing Association (ASHA).

Teachers

At the beginning of the study, all teachers attended an orientation session to review the purpose and procedure of the study and to answer any questions. The teachers in the comparison group were only given general information about the study , as well as more specific information about the procedures for both the pre/posttesting and the videotaping for intervention fidelity checks. The teachers in the experimental group were given information about the specifics of the study, the intervention calendar and schedule, as well as the procedures for both the pre/posttesting and the videotaping for intervention fidelity checks. Upon study completion, all participating teachers would reconvene for a culminating session to discuss results of the study. The possibility of maintenance and generalization of the intervention would be discussed with teachers and administrators at that time.

Intervention Dosage

All intervention sessions for the experimental group classes were conducted during their regularly scheduled science classes. No intervention sessions were completed with the comparison classes. The researcher conducted the intervention sessions at the beginning of the science class for 20 minutes. The teachers and administration felt that the beginning of the class was the time when the students would have the best attention and it would be easier for the science teacher to segue into their instruction for the remainder of the class (15-25 minutes). The experimental group science teachers were present during the intervention and were encouraged to circulate around the room to maintain student attention and motivation. The researcher consulted the science teachers as the "science experts" during the intervention, such as when the researcher and students were determining correct word choices for the sentence combinations.

The intervention spanned a total of seven weeks to allow for completion of the 20-day protocol, with equitable intervention for both the experimental classes. Both experimental classes received the same intervention protocol and exercises, on the same days, at the beginning of their regularly scheduled science class. The researcher worked with all the teachers to determine a schedule that would allow them to cover all the science content needed prior to the state science exam in mid-April. The first week had a Monday holiday, so the researcher conducted four subsequent sessions for the remainder of the week. For the following six weeks, the researcher worked with the students three days per week, on Monday, Tuesday, and Wednesday. There were only two weeks when the researcher held only two intervention sessions, the third week due to the Florida State Assessment in Writing and the last week when only two sessions were needed to complete the 20-day protocol. There was a week of spring break in between session 15 and 16. The total treatment hours completed was the same as originally planned, which was 20 sessions, for 20 minutes, for a total of 400 minutes or 6.6 hours.

Intervention Phases

The intervention consisted of three phases. During each intervention phase, each session provided metalinguistic sentence-combining practice with a certain level of instructional support. The instructional levels of support were in a sequential order that is described later in this chapter. For each intervention session, the students were expected to complete two sentences, which were their final sentence combinations.

The goal of Phase One was for the students to be able to combine two sentences into one sentence to compare or contrast science information. This phase was for the students to become oriented with the sentence-combining process and the concepts of comparing and contrasting. The students were given two different combinations or sets of two kernel sentences to combine into one sentence. This phase was a total of eight sessions. The first two sessions only targeted comparing. Session three targeted contrasting with teacher modeling and session four targeted contrasting with student enlistment model. Starting on the fourth day of intervention in this phase, the researcher presented the editing procedure. For the remainder of the intervention, the students were given the first two minutes of the session to edit their work. Sessions five through eight were for the students to compare or contrast.

The goal of Phase Two was for the students to take four sentences and combine them into two sentences. Two of the kernels sentences expressed a similarity and two of the kernel sentences expressed a difference. This phase was to continue the students' practice of sentence combining for comparing and contrasting, but increase the amount of semantic content that needed to be comprehended and expressed. Students who struggle with literacy have difficulty parsing dense sentences with multiple clauses for comprehension and expression of content (Fang, 2008; Scott & Balthazar, 2013). Due to the increase in information and cognitive demand, the students only completed one set of four sentences that they had to combine into two sentences. This phase was a total of six sessions.

The goal of Phase Three was for the students to combine three sentences into one sentence. This phase was to continue to practice comparing and contrasting, but to add an extra clause into the sentence. The extra clause could be embedded within the sentence or expanded by placing the clause at the beginning before the main clause, or at the end of the sentence.

Students who struggle with literacy have deficits in comprehending and writing sentences with more than one main clause or several elaborative phrases (Dockrell, 2014; Scott, 2009a; Wallach & Weis-Liebergott, 1984). This phase was also a total of six sessions. See Appendix O for the Intervention Outline and Appendix P for the Scope and Sequence Plan.

Editing

In session four, the researcher explained the editing process to the students. The use of editing has had a large to moderate impact on decreasing errors for writers who struggle in grades 8-12 (Rogers & Graham, 2008). The editing process used an "uncoded" feedback system. Uncoded feedback is information given by the teacher by underlining or circling an error (Bitchener, Young, & Cameron, 2005). The students were left to then diagnose and correct the error themselves. The researcher explained to the students that all writers needed to edit their work to check for errors or to make changes. The researcher explained to the students that they would be given the opportunity to edit the two sentences they had written during the prior class for the first two minutes of each session. The errors were either circled or marked with underscored line to signal the type of edit that was required. If a letter, word, or punctuation mark was circled, then it meant that it needed to be changed. These marks indicated if a word was incorrectly capitalized, a word that was on the worksheet was spelled (or copied) incorrectly or the verb tense or verb choice did not match the noun. If there was an underscored line placed after a word or placed in between words, that meant that something was missing. These marks indicated if a period or comma was missing, a verb ending was missing, or words were missing. Each student was given one session to make the appropriate edits. The researcher encouraged the students to ask questions about their edits during the editing time period. If the student did not edit correctly the first time, the researcher allowed the student to edit one more time, with a written note giving a clue what is needed such as “Check your spelling”, or “Test this noun and verb aloud.”

Instructional Levels of Support

The researcher implemented six different levels of support as a method of gradual release of instruction, within each phase of the intervention. The level of support is referring to the amount of support or modeling provided by the researcher. The researcher provided verbal cues (i.e., question prompts) and completed overt actions (e.g., crossing out words, writing the sentence) during the intervention. The level of support ranged from the researcher modeling all aspects of the intervention (Teacher Model) to a steady reduction of the researchers level of modeling (a Student Enlistment Model, Guided Practice, Faded Guided Practice), to support from peers (Peer-Supported Practice), to the students completing the intervention independently (Independent Practice). However, the students did have the mnemonic WRITEr. to use as an aid to help recall the steps of the intervention if needed; therefore, the students were never completely independent of any support. At each phase of the intervention, the researcher provided a level of support in the same order: (1) a Teacher Model, (2) a Student Enlistment Model, (3) Guided Practice, (4) Faded Guided Practice, (5) Peer-Supported Practice, and (6) Independent Practice.

The first level was a Teacher Model. At this level, the teacher modeled all aspects of the intervention and demonstrated the overt actions. While the researcher modeled and demonstrated the intervention, she used “think-alouds” to reveal her cognitive processes and engaged in “self-talk” to provide answers (Schumaker, 1989). Students who struggle with literacy are lacking these essential cognitive processes and this explicit part of instruction is fundamental (Deshler & Lenz, 1989). The researcher instructed the students to watch and listen to what the researcher said, did, and then asked them to imitate it. According to Schumaker (1989) successful learners imitate models without being told through watching and listening; however, learners who struggle do not attend to models or lack the ability to imitate a model without further repetition or assistance. Therefore, the researcher monitored the students carefully to make sure they were following each of the steps.

The second level was a Student Enlistment Model. At this level, the researcher initiated a process where the student would become an active participant in the intervention and its guided practice (Schumaker, 1989). The teacher enlisted the students to start to tell or describe the steps of the intervention as a means of gauging their understanding of the steps of the intervention. The goal of this level was for the students to become more confident with the steps of the intervention while still having the support of the teacher feedback (Deshler & Lenz, 1989). The researcher asked questions at a level that the students could be successful as a means of “errorless teaching.” The researcher asked the students yes/no questions related to the steps or actions of the intervention that had a visual cue either in the mnemonic or on the MSC intervention sheet. For example, the researcher asked, “Is the word ‘both’ a word we use for a similarity?" If the students struggled with the answer, the researcher could point to the section of the word bank that listed the similarity connective words to help the student retrieve the answer. The students imitated the actions performed by the researcher throughout this level.

The third level was Guided Practice. Instead of the researcher asking the students yes/no questions, during this level the researcher asked wh-questions such as "what", "which", and "why." The questions guided the sequence of the intervention steps, but the students were expected to perform the actions with more independence. The researcher then asked the students to tell the steps of the intervention as specified in the mnemonic, specifically, what the researcher should be thinking or asking at each step, and then what the researcher should be doing at each step. The students should have started to perform the actions as a result of these question prompts, but could rely on imitation if needed, because the researcher performed the actions simultaneously. During fading, if a student had been struggling, the researcher would go back and provide scaffolding with a yes/no question and visual supports as needed. Fading also provided the researcher the ability to being to check for students' understanding in order to provide corrective feedback. Another element added at this level was that the researcher asked questions that would allow students to make syntactic choices, such as conjunctions, other wording, or clause placement during Phase Three. These options are encouraged in sentence-combining instruction to teach students the variety of syntactic options when writing (Saddler, 2012). The researcher explicitly discussed the thought processes that should occur as a result of these choices, the actions taken by the writer, and the implications on the reader. For example, the researcher asked the students whether they want to use "but" or "however" for a sentence combination that was expressing a difference. Then, the researcher explicitly stated to remember the need for a comma before the word "but", or the need for a semicolon before and comma after for the word "however", and why this punctuation was needed. The answer to why the punctuation was needed related to the implications for the reader such as a statement such as, "It signals to the reader to take a pause because the next part of the sentence is going to contrast the first part."

The fourth level was Faded Guided Practice. During this level, the researcher monitored the level of support the students needed and then gradually faded the support such asking wh-questions. The researcher instead requested that the student tell the researcher the steps, cognitive processes, and actions to perform, while doing them on their own. The researcher could further check students' understanding and provide corrective feedback as they performed the tasks more independently. The researcher continued to encourage the students to make choices and to state their thought processes, actions, and possible implications of their choices.

The fifth level was Peer-Supported Practice. During this level, the students were paired with a partner. The researcher instructed the students to work with their partners to complete the MSC exercises. The researcher circulated the room to provide support and corrective feedback (Schumaker et al., 2006). At the end of these sessions, the researcher used two examples the students had created to demonstrate the correct answers.

The sixth level was Independent Practice. The researcher instructed the students to work independently. The researcher circulated the room and provided minimal prompts. At the end of these sessions, the researcher collected the student's work. Then, the researcher provided two different examples of correct answers.

Motivational Techniques

The researcher arranged scheduled and unscheduled tangible reinforcements to keep the students motivated. The researcher collaborated with the two experimental teachers to determine what type of tangibles reinforcements would be highly motivating and what was permitted by the school administration. First, the students received a sticker after each session if both the sentences were correct. If the students had edits, then they received the sticker after editing the sentences. Second, after a period of thirteen sessions, the researcher gave out a coupon for the school store to the students with the most stickers. Another coupon was also distributed at session 18. Third, all the students received other tangible reinforcers on randomly selected days if they completed the two sentences that had to be written for the day.

Data Analysis

This section begins with the power analysis conducted and the baseline equivalency testing used for data analysis. Then, the assumptions that were met prior to conducting a hierarchical repeated measures analysis of variance (ANOVA) or a hierarchical repeated measures multivariate analysis of variance (MANOVA) are discussed. After that, the alpha level and effect size measurements used are listed, as well as the specific statistical analysis conducted to test each research hypothesis.

Power

Power, the probability of correctly rejecting a false null hypothesis, was conducted a priori and post hoc. Prior to identifying potential participants, G*power 3.13 software (Faul, Erdfelder, Lang, & Buchner, 2007) was used to determine appropriate sample sizes. This software cannot make estimations for hierarchical repeated measures ANOVA or hierarchical repeated measures MANOVA. Therefore, the first estimation was made based on the closest proxy, a repeated measures ANOVA within-between measures interaction, comparing two groups, with two measurements, at a significance level of .05, and a power level of .80. The results revealed that a sample size of 200 is needed to detect a small effect, a sample size of 34 was needed to detect a medium effect, and a sample size of 16 to detect a large effect. The second estimation was made based on the closest proxy, a repeated measures MANOVA within-between measures interaction, comparing two groups, with two measurements, at a significance level of .05, and a power level of .80. The results revealed a sample size of 199 was needed for a small effect, a sample size of 34 was needed for a medium effect, a sample size of 15 for a large effect. In addition, post hoc observed power was provided by the Statistical Package for Social Sciences (SPSS) version 22, and is listed for each research question separately.

Baseline Equivalence

Equivalence testing is the process of determining the extent that two groups are equivalent to each other, which cannot be assumed with a nonrandomized sample (Rogers, Howard, & Vessey, 1993). Equivalence testing was conducted to determine the baseline distance between the pretest mean scores of the experimental and comparison groups. First, assumptions testing for an independent *t*-test was completed. Then the independent t-test was generated to compare the outcomes for the experimental and comparison groups. If there is no statistical significance ($p > .05$), as well as small or near zero effect ($d < .20$; Cohen, 1988), between the experimental and comparison groups pretest scores, then baseline equivalence was assumed. Baseline equivalence is reported for each research question.

Conditions and Assumptions

The following conditions of hierarchical repeated measures ANOVA and MANOVA were met: (a) the independent variable or between-subjects factor with two or more levels fixed by the researcher; (b) the subjects are only exposed to one level of a nominal independent variable; (c) the dependent variable is measured at interval or ratio level; and (d) there is control of experimentwise error rate through an omnibus test (Hahs-Vaughn, in progress; Lomax & Hahs-Vaughn, 2012).

The use of hierarchical ANOVA requires the testing of the assumptions of (a) independence, (b) normality, and (c) homogeneity of variances (Lomax & Hahs-Vaughn, 2012). In addition to these assumptions, mixed ANOVA (i.e., includes both between and within factors) has the assumption of sphericity. However, sphericity is applicable only to analyses that include at least three categories for the between-subjects factor. Thus, sphericity was not applicable in this study, and was not considered further. The use of hierarchical MANOVA requires the same assumptions of ANOVA, as well as multivariate normality, linearity, and homogeneity of covariances (Hahs-Vaughn, in progress).

Independence

The use of repeated measures design reduces variability due to individual differences (Stevens, 2008). However, the assumption of independence is sensitive to Type I and or Type II errors that occur when this assumption is violated (Lomax, 2007). Independence is typically met with the use of randomization. This research study did not employ randomization; therefore, other measures were used. For hierarchical repeated measures ANOVA, scatterplots of standardized residuals of the within-subjects factors by the between-subjects factor (i.e., group) were reviewed. For hierarchical repeated measures MANOVA, examination of scatterplots of standardized residuals by group are reviewed. Results of independence testing are reported and discussed for each research question.

Normality

The assumption of normality assumes the populations follow the normal distribution. Violations of normality are typically due to outliers (Lomax, 2007), and both ANOVA and MANOVA models are sensitive to outliers. Data were examined for potential outliers and considered for removal to help achieve normality or a normal distribution. Univariate normality assumptions were tested with graphs (e.g., histograms, boxplots), normal probability (Q-Q) plots, skewness, and kurtosis statistics, as well as formal normality tests such as the Shapiro-Wilk test (Lomax & Hahs-Vaughn, 2012). The *F*-test is robust to moderate violations of the normality assumption for ANOVA and less severe with large or nearly equal *n*'s (Lomax, 2007).

For hierarchical repeated measures MANOVA, non-zero skewness (larger than 1.5 or 2.0) will have an impact on parameter estimates, but non-zero kurtosis will have a minimal effect (Hahs-Vaughn, in progress). Skewness and kurtosis are considered when analyzing data with hierarchical repeated measures MANOVA. For multivariate normality, the use of the De Carlo (1997) SPSS macro was used to determine multivariate kurtosis, skewness, and omnibus test. MANOVA tests are also robust to violations of normality when equality in group size is maintained (Olson, 1974). Results of normality testing are reported and discussed for each research question.

Linearity

Linearity is an assumption of hierarchical repeated measures MANOVA. The assumption of linearity is met when regression of the dependent variables are linear (Hahs-Vaughn, in progress). For hierarchical repeated measures MANOVA, linearity of the dependent variables can be examined by scatterplots of all pairs of dependent variables. Therefore, the matrix scatterplots of dependent variables are reviewed to test this assumption. Linearity testing is reported and discussed for research questions two and three that utilized hierarchical repeated measures MANOVA.

Homogeneity of Variance and Covariance

Violations of the homogeneity of variance, or assuming the variances of each population are the same, may result in Type I and Type II errors (Lomax, 2007). For hierarchical repeated measures ANOVA, the violations of this assumption are trivial if the sample sizes are similar across groups (Lomax & Hahs-Vaughn, 2012). This assumption is tested with Levene's test and Box's test for hierarchical repeated measures ANOVA. MANOVA models are not robust for violations of this assumption and are worse when the number of dependent variables increases (Seltman, 2012). Box's test is highly sensitive to non-normality and cells with larger variance-covariance matrices (Olson, 1974). Therefore, for hierarchical repeated measures MANOVA, Levene's test for univariate analysis is used to determine the dependent variables(s) with heterogeneous variables (Hahs-Vaughn, in progress). A large number of dependent variables can contribute to unequal variance-covariance matrices and may reduce power (i.e., failing to find statistical significance when in fact there is significance). Thus, there is less reason for concern of violation of this assumption if statistical significance is found (Hahs-Vaughn, in progress). Homogeneity of variance and covariance is reported and discussed for each research question.

Statistical Analyses

The data were analyzed with statistical software, SPSS version 22.0. All questions were examined with an alpha level of .05. Effect size, or practical significance, was determined by Cohen's (1988) standards for partial eta squared: small effect, multivariate $\eta^2 = .01$; medium effect, multivariate η^2 = .06; and large effect, multivariate η^2 = .14. The following analyses were conducted on each research hypothesis:

Hypothesis-Research Question One: Eighth-grade students struggling with literacy who participate in MSC instruction in science will demonstrate a greater increase in score on the Test of Written Language-4 Sentence Combining Subtest form A/B than eighth-grade students struggling with literacy who participate in typical science instruction alone.

Statistical Analysis -Hypothesis One: A hierarchical repeated measures ANOVA was conducted with the independent variable of treatment group (experimental and comparison), the dependent variable was the pretest and posttest scores, and the hierarchical factor was the classroom (i.e., teacher).

Hypothesis-Research Question Two: Eighth-grade students struggling with literacy who participate in MSC instruction in science will demonstrate greater increases in measures of sentence complexity in response to a science compare/contrast prompt than eighth-grade students struggling with literacy who participate in typical science instruction alone. Sentence complexity will be measured using the following seven measures: (1) sentence length, (2) sentence connectives, (3) agentless passive voice, (4) number of correct word sequences, (5) number of targeted sentence connectives, (6) correct versus incorrect sequences count (CIWS), and (7) number and type of morpho-syntactical errors.

Statistical Analysis-Hypothesis Two: A hierarchical repeated measures MANOVA was conducted with the independent variable of treatment group (experimental and comparison), the dependent variable was the pretest and posttest scores, and the hierarchical factor was the classroom (i.e., teacher).

Hypothesis-Research Question Three: Eighth-grade students struggling with literacy, who participate in MSC instruction in science will demonstrate greater increases in sentence complexity when writing a science expository essay than eighth-grade students struggling with literacy who participate in typical science instruction alone. Sentence complexity will be measured using the following eight measures: (1) sentence connectives, (2) words before the main clause, (3) agentless passive voice, (4) noun phrase density (5) verb phrase density, (6) prepositional phrase density, (7) correct versus incorrect sequences count (CIWS), and (8) number and type of morpho-syntactical errors.

Statistical Analysis-Hypothesis Three: A hierarchical repeated measures ANOVA was conducted with the independent variable of treatment group (experimental and comparison), the dependent variable was the pretest and posttest scores, and the hierarchical factor was the classroom (i.e., teacher).

Hypothesis-Research Question Four: Eighth-grade students struggling with literacy who participate in MSC instruction in science will demonstrate a greater increase in their ability to determine similarities and differences (compare/contrast) related to science content as measured by an increase in score on the compare and contrast double bubble map by Thinking Maps ®

than eighth-grade students struggling with literacy who participate in typical science instruction alone.

Statistical Analysis-Hypothesis Four: A hierarchical repeated measures ANOVA was conducted with the independent variable of treatment group (experimental and comparison), the dependent variable was the pretest and posttest scores, and the hierarchical factor was the classroom (i.e., teacher).

Fidelity of Implementation

The researcher provided all the intervention to the experimental classes to control for the external threat of treatment variation (Edmonds & Kennedy, 2013). The researcher developed an Innovation Configuration (IC) Map (Hall & Hord, 2013) as the basis of the fidelity checklists (see Appendix Q). Each checklist listed the steps completed during the intervention. Three checklists were for sessions with teacher support and separated by phase (i.e., one, two, thee), as well as two separate checklists for sessions with minimal teacher support regardless of phase (i.e., peer-supported practice and independent practice (see Appendix R). The researcher completed the fidelity checklists at the end of each session. Throughout the study, the two research assistants watched the video-recorded sessions and filled out the corresponding fidelity checklist. Each checklist tallied the total number of steps completed correctly. A final percentage was determined by dividing the total number of correct steps by the total number of incorrect steps and multiplying by 100%.

In addition, all intervention sessions were video recorded to be viewed by the researcher to insure that the teachers were not utilizing sentence-combining instruction or any aspect of the intervention during their science instruction. The researcher observed 20% of the total possible instructional time that occurred over the seven weeks to document whether or not there was evidence of any instruction related to the intervention protocol or any type of sentence-combining instruction.

Inter-Rater Reliability

The point-by-point method formula of agreements/agreements + disagreements x 100% (Gast, 2010) was also used to determine inter-rater agreement for all assessment measures. The researcher created inter-rater reliability forms to tally the agreements and disagreements (see Appendix S). Prior to the study, a team of four research proctors was trained to administer and score the Test of Written Language Fourth Edition (TOWL-4) Sentence Combining Subtest assessment. For the science compare and contrast written prompt and the science expository essay, two other research assistants were trained. One was trained to analyze the participants' handwritten data samples from the science compare and contrast written prompt and the science expository essay. The research assistant typed the samples into the Coh-Metrix 3.0 computer software and then had to select and run the appropriate analyses. The other research assistant was trained to complete the Correct Versus Incorrect Word Sequences (CIWS) Count and language sample analysis to determine the number and type of targeted sentence connectives (i.e., both, like, but, however) and the morpho-syntactical errors on both the science compare and contrast written prompt and science expository essay.

For the science compare and contrast double bubble maps, a sample of the bubble maps was randomized, identifying information removed (i.e. student name), then distributed to the four science teachers to be graded. The answers (graded as correct or incorrect) were then pooled by the researcher, and then a fifth science teacher from a different school helped to reach consensus as to whether an answer would be counted as correct or incorrect. Last, the researcher created a formal answer key that was used by the two trained research assistants who scored the double bubble maps.

Chapter Summary

This chapter presented and reviewed the methodology for the current study. The study employed a quasi-experimental design to answer the noted four research questions. The setting, participants, and sampling procedures were discussed. The instrumentation and procedures for assessment and intervention were outlined. Finally, the data analytic procedures for each of the research hypotheses and fidelity of implementation were documented.

CHAPTER FOUR: RESULTS

This study investigated the effects of a metalinguistic approach to the writing intervention of sentence combining to improve written sentence complexity in science, as well as the written expression and determination of comparison and contrast of science content, for eighth-grade students who struggle with literacy. This chapter begins with information about participant attendance for both the intervention and assessments. The chapter then presents the statistical data analysis and results of each research question analyzed with either a hierarchical repeated measures analysis of variance (ANOVA) or a hierarchical repeated measures multivariate analysis of variance (MANOVA). The chapter continues with the data results of the editing process, teacher perceptions of writing survey, social validity of the intervention based on teacher and student surveys, and ends with a chapter summary.

Participant Attendance

Intervention

Students were encouraged to attend all science classes over the duration of the study; however, not all students were present each day. The researcher provided a total of 400 minutes (20 sessions for 20 minutes) of intervention to the students in the experimental classes during the study. This did not include the pretesting and posttesting. The average number of treatment sessions for the experimental group was 18 out of 20 or 90%. Eleven students (30%) were present for each session. The range of absences for students enrolled in the duration of the study was a zero (missing 0% of total treatment) to a maximum of nine sessions (missing 45% of total treatment). The average number of classes missed was 1.46 (7.3%). The comparison group participants did not receive any treatment, but attendance data were collected for their attendance in science class. A total of 25 students (53%) were present for each science class that occurred on the same days as the experimental classes who received treatment (20 classes). The range of absences was zero to a maximum of six classes. The average number of classes missed was less than one (less than 5%). One student from experimental class two participated in pre-testing and only 11 treatment sessions (55%), and then relocated to a different school. Similarly, one other student relocated into the school district. The student was placed in experimental class one during the treatment period, only participating in 14 sessions (70%) and posttesting. None of the students moved across treatment groups during the study duration. There was no minimum of treatment sessions set by the researcher for a student to be included in the data analysis; however, all the students had participants had attended 50% or more of the sessions. Table 7 presents the treatment hour data for the treatment group. Figure 4 is a histogram that depicts the total treatment hours for the experimental group. The shape of the histogram suggested a relatively normal distribution.

Table 7: *Total Treatment Hours*

Experimental Group	*M*	*SD*
Treatment Hours	18.14	2.188
	Minimum	220
	Maximum	400
	Range	180

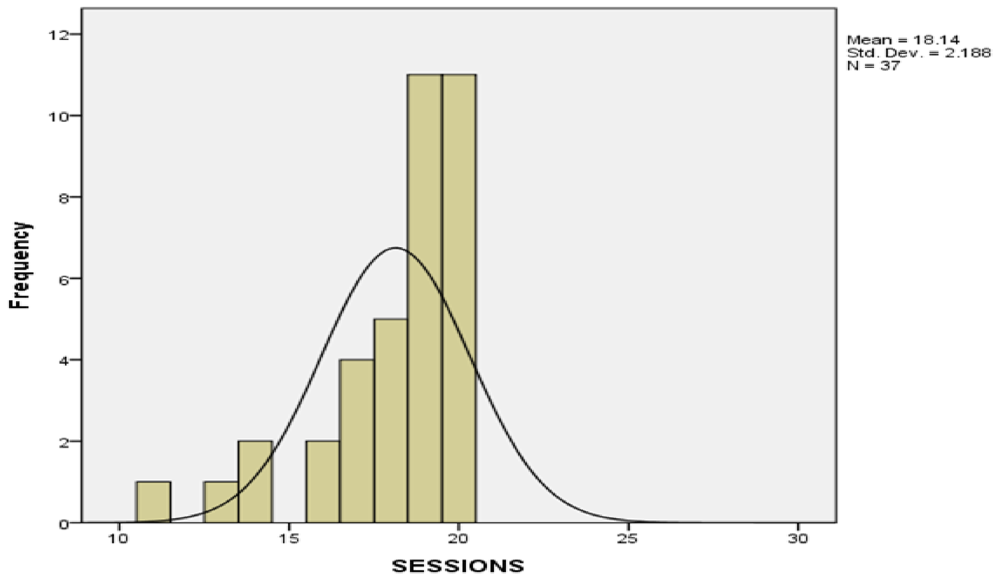


Figure 4: *Histogram for Total Treatment Hours-Experimental Group*

Assessment

Missing Data

At pretest, there were 84 possible participants (36 experimental, 48 comparison). One student in the experimental group (experimental class two) relocated to another school after pretest and one student relocated into the participating school (experimental class one) prior to the posttest, keeping the number of participants consistent for the experimental group (n = 36) for the duration of the study. Similarly, one student moved out of the comparison group after pretest, and one student relocated into the participating school prior to posttest in comparison class one; keeping the number of participants consistent for the comparison group (n = 48) for the duration of the study.

The researchers scheduled the pretesting and posttesting, which was agreed upon by the school administration and teachers. Despite these arrangements, there were students absent. The Test of Written Language Sentence Combining Subtest-Fourth Edition (TOWL-4) and the science compare/contrast double bubble maps were administered on day one of pretesting. Out of 84 participants, 72 participants (34 experimental, 38 comparison) or 85% completed the pretest. Thirteen participants (three experimental, 10 comparison), or 15%, were absent. Of the 72 participants who completed the pretesting, 68 participants (31 experimental, 37 comparison) completed the posttesting and four participants (three experimental, one comparison) or 5%, were not present. Two of the four students who were missing had relocated to another school (one experimental, one comparison); the other two students were absent that day. Thus, the final sample size for the TOWL-4 Sentence Combining Subtest and the science compare/contrast double bubble map (day one of pretest and posttest) was 68 or 81% of possible participants (86% of the experimental group, 77% of the comparison group). Appendix T presents a flow chart depicting the collected and missing data for pretesting and posttesting for day one.

The science compare and contrast writing prompt and the science expository essay were administered on day two of pretesting. Out of 84 possible participants, 60 participants (26 experimental, 34 comparison) or 71% completed pretesting. Twenty-four participants (10 experimental, 14 comparison) or 29% were absent. Of the 60 participants who completed pretesting, 49 participants (21 experimental, 28 comparison) completed the posttesting, and 11 students (five experimental, six comparison) or 18% were not present. Two of the 11 students were students who had relocated to another school (one experimental, one comparison), and the other nine were absent that day. Thus, the final sample size for the science compare and contrast writing prompt and the science expository essay (day two of pretest and posttest) was 49, or only 58% of possible participants (58% of experimental group, 58% of comparison group). In order to include the students who completed the pretesting on day two, but not posttesting, an intent-to-treat analysis (ITT) was used. Appendix U presents a flow chart depicting the collected and missing data for day two of pretesting and posttesting.

Intent-To-Treat (ITT) Analysis

An intent-to-treat analysis (ITT) was used as a second statistical analysis for research question two (the science compare and contrast writing prompt) and research question three (science expository essay). The attrition of students who completed the writing pretest measures could be analyzed to allow for an increased sample size and probable increased statistical power (Gall et al., 2006). ITT analysis uses data for all participants, using their pretest scores as their posttest score (Torgerson & Torgerson, 2008). A negative aspect of ITT analysis is that it can weaken any difference in treatment effect, which can result in loss of information on the efficacy of the treatment (Armijo-Olivo, Warren, & Magee, 2009). However, a positive effect of ITT analysis is that it is a cautious approach that can give an unbiased estimate of treatment effect, while preserving sample size, minimizing type I error while allowing for the greatest generalizability (Gupta, 2011). In addition, researchers recommend conducting data analysis in more than one method to test and compare the validity of conclusions (Armijo-Olivo et al., 2009). For research questions two and three, which are related to the science writing prompt and expository essay, these questions were first examined as treatment-on-the-treated analysis (i.e., using only data from the students who had both pretest and posttest measures), then again with ITT and comparisons were made. Twenty-four participants, 10 from experimental group and 14 from the comparison group, were not present for pretesting and were excluded from the sample entirely. A total of 49 students (21 experimental, 28 comparison) completed both the pretesting and the posttesting. There were 11 other students who completed only the pretesting and were counted in the ITT analysis, yielding a total sample size of 60 (26 experimental students or 72% of possible experimental students; 34 comparison students or 71% of possible experimental students).

Data Analysis Results

Hierarchical repeated measures ANOVA statistics were used to test research questions one and four. Hierarchical repeated measures MANOVA statistics were used to test research questions two and three. The hierarchical factor for both analyses was the nesting of students within the classroom (i.e., teacher). The experimental treatment condition and classroom were fixed factors. Each research question had different dependent variables that are listed in each research question.

Research Question One

Do eighth-grade students struggling with literacy who participate in metalinguistic sentence combining (MSC) instruction in science demonstrate an increase in their sentence-combining ability as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

A hierarchical repeated measures ANOVA model was generated to answer this question. The independent variable was treatment group (experimental or comparison), the dependent variable was the pre-test and posttest mean scaled scores, and the hierarchical factor was the classroom (i.e., teacher). The hierarchical repeated measures ANOVA was conducted to determine if the within-subjects factor, the pretest (form A) to posttest (form B) of the TOWL-4 Sentence Combining Subtest was significant, and if there was a significant interaction with the between-subjects factors of treatment group (experimental or comparison), while accounting for the nesting of students within the classroom (i.e., teacher).

Baseline Equivalency Testing Results

Equivalency testing was conducted on the baseline pretest scores of the TOWL-4 Sentence Combining Subtest (dependent variable) between the two treatment groups (independent variable). Prior to running the independent *t*-test, the assumptions of the normality of the distribution was tested with the Shapiro-Wilk test (experimental group $SW = .995$, $df = 31$, $p = .114$, comparison group $SW = .994$, $df = 37$, $p = .089$) and met with *p*-values greater than alpha level of .05. Skewness and kurtosis statistics (experimental skewness= .043, kurtosis= .529; comparison skewness = -.008, kurtosis = .846) were met, which are within an absolute value of 2.0. Visual examination of Q-Q plots, histograms (shape), and boxplots (presence of outliers) suggested a normal distribution. Levene's test was completed to test the assumption of homogeneity of variances and was met with $p = .960$ which is greater than alpha level of .05. Last, an independent *t*-test was completed to determine significance (met with $p < .05$, no significant differences in scores), and small to minimal effect by Cohen's *d* (Cohen, 1988). The results of $t(66) = .213$, $p = .832$, was not significant, and $d = .05$ indicated a small effect, which suggested the two treatment groups were relatively equivalent at baseline on the TOWL-4 Sentence Combining Subtest.

Assumptions Testing Results

The assumptions of independence, normality, and homogeneity of variances and covariances were tested before running the hierarchical repeated measures ANOVA model. The assumption of independence was determined by reviewing the scatterplots of the standardized residuals of the pretest and posttest scores by treatment group. There was a relatively random display of difference scores above and below zero, with no cyclical pattern. This generally suggested evidence of independence.

The assumption of univariate normality was examined through several indices using residuals. The original sample size was n = 68 (experimental n = 31, comparison n = 37). Visual examination of the histograms, boxplots, and Q-Q plots revealed evidence of two outliers that were greater than two standard deviations above or below the mean. The outliers were removed, and the sample size was reduced to n = 66, with one case removed from each group (experimental n = 30, comparison n = 36). The Shapiro-Wilk test revealed normality for the pretest residual (experimental p = .250, comparison p = .336), but not for the posttest residual (experimental p = .024, comparison p = .017). The posttest residuals are less than alpha level of .05, which suggested non-normality. All skewness and kurtosis values were within an absolute value of 2.0, which suggested normality. More specifically, the skewness and kurtosis statistics included the following: pretest residual (skewness = .217, kurtosis = -.142) and posttest residual (skewness = -.810, kurtosis = .582) for the experimental group; and pretest residual (skewness = .014, kurtosis = -.718) and posttest residual (skewness = -.254, kurtosis = -1.096) for the comparison group. Violations of normality are robust to moderate violations for ANOVA models with nearly equal n's (Lomax, 2007). The groups sizes were experimental (n = 30), and comparison (n = 36); therefore, it was reasonable to assume that normality was met.

The assumption of homogeneity of variance was met with Levene's test (pretest, p = .673, posttest, p = .805), which was above alpha level of .05. Box's test (p = .842) was also above the alpha level of .05; therefore, the assumption of homogeneity of covariances matrices was also met. Table 8 presents the data testing for repeated measures ANOVA assumptions.

Table 8: *Results of Assumptions Testing for the TOWL-4 Sentence Combining Subtest Pretest, Posttest*

Assumption	Test		Evidence	Assumption Satisfied?
Independence	Scatterplots		No observable trends	Yes
Normality	Shapiro-Wilk	Experimental Pretest residual	*SW* = .956, *df* = 30, *p* = .250	Yes
		Posttest residual	*SW* = .918, *df = 30, p = .024*	No
		Comparison Pretest residual	*SW = .999, df = 36, p = .336*	Yes
		Posttest residual	*SW* = .925, *df* = 36, *p* = .017	No
	Boxplot/Histogram Q-Q Plot		Relatively normal distribution shape	Yes
	Skewness	Experimental Pretest residual	.217	Yes
		Posttest residual	-.810	Yes
		Comparison Pretest residual	.014	No
		Posttest residual	-.254	Yes
	Kurtosis	Experimental Pretest residual	-.142	Yes
		Posttest residual	-.582	Yes
		Comparison Pretest residual	-.718	Yes
		Posttest residual	-1.096	Yes
Homogeneity of variance/ covariance matrices	Levene's Test	Pretest scores	*F* (4, 63) = .587, *p* = .673	Yes
		Posttest scores	*F* (4, 63) = .404, *p* = .805	Yes
	Box's Test		*p* = .842	Yes

Hierarchical Repeated Measures ANOVA Results

The results for the hierarchical repeated measures ANOVA suggested that there was a statistically significant main effect for the within-subjects factor of the difference between pretest and posttest (F_{test} = 14.281, *df* = 1, *p* = .000). Multivariate partial eta squared for the main effect indicated a large effect and strong power (partial η^2_{test} = .185, observed power = .961), as determined by Cohen (1988). Approximately 19% of the variance on the TOWL-4 Sentence Combining Subtest can be accounted for by the within-subjects factor (i.e. time from pretest to posttest).

There was no statistically significant interaction between time (i.e., within-subjects factor) and the between-subjects factor of treatment group ($F_{treatment}$ = .011, *df* = 1, *p* = .830). Multivariate partial eta squared for the interaction of time and treatment group indicated a small effect and low power (partial $\eta^2_{treatment}$ = .001, observed power = .005), as determined by Cohen (1988). Less than 1% of the variance on the TOWL-4 Sentence Combining Subtest can be accounted for by treatment group (experimental or comparison).

In addition, there was not a statistically significant effect between time and nesting within classroom ($F_{classroom=}$ 1.582, *df* = 3, *p* = .203). Multivariate partial eta squared indicated medium effect and low power (partial $\eta^2_{classroom}$ = .070, observed power = .397), as determined by Cohen (1988). Approximately 7% of the variance on the TOWL-4 Sentence Combining Subtest can be accounted for by the hierarchical structure of the data (i.e., students nested within teacher).

The students in the experimental group did achieve a slightly higher marginal mean as compared to students in the comparison group (experimental *M* = 8.632, *SE* = .523, *CI* = 7.586 to 9.678; comparison *M* = 8.364, *SE* = .485, *CI* = 7.394 to 9.334) on the TOWL-4 Sentence Combining Subtest. In addition, the experimental pretest mean score was 8.03, *SD* = .523, which increased 1.23 points to a posttest mean score of 9.26, *SD* = .556. The comparison pretest mean score was 9.00, *SD* = .523 which increased less than one point (.12) to a posttest mean score of 9.12, *SD* = .485. Table 9 presents the pretest, posttest, and marginal means, standard deviation/error, and confidence intervals for both the experimental and comparison groups.

Table 9: *Means, Standard Error, and Confidence Intervals by Treatment Group for the TOWL-4 Sentence Combining Subtest*

Treatment Group		Score	Standard Deviation	Standard Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Experimental	Pretest Mean	8.03	.585			
	Posttest Mean	9.26	.556			
	Marginal Mean	8.632[a]		.523	7.586	9.678
Comparison	Pretest Mean	9.00	.542			
	Posttest Mean	9.12	.516			
	Marginal Mean	8.364[a]		.485	7.394	9.334

[a] *Based on modified population marginal mean*

Research Question Two

Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase in specific aspects of sentence complexity in response to a science compare/contrast writing prompt as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

A hierarchical repeated measures multivariate analysis of variance (MANOVA) was computed to answer this question, first by applying treatment-on-the-treated analysis (i.e., analyzing only data from students who had no missing outcome data) with a sample size of n = 49, and then utilizing ITT analysis with a sample size of n = 60. The independent variable was treatment group (experimental or comparison), the dependent variables were: (1) sentence length, (2) sentence connectives incidence, (3) mean words before the main clause, (4) agentless passive voice density incidence, (4) targeted connectives, and (5) correct versus incorrect word sequences, and the hierarchical factor was the classroom (i.e., teacher). The hierarchical repeated measures MANOVA was conducted to determine if the within-subjects factor, the pretest and post science compare and contrast writing prompt measures, were significant; and if there was a significant interaction with the between-subjects factors of treatment group (experimental or comparison), and nesting for classroom (i.e., teacher).

Treatment-on-the-Treated-Analysis

Baseline equivalency testing results

Equivalency testing was conducted on the baseline pretest scores of each of the six dependent variables between the two treatment groups (independent variable). Prior to running each of the independent *t*-tests, the assumptions of the normality of the distributions were tested with the Shapiro-Wilk test and were met for the dependent variables of sentence length, comparison group of the connectives, experimental group for words before the main clause, and the comparison group of the CIWS count, with *p*-values ranging from .213 to .846 (greater than alpha level of .05). The assumptions of normality were not met with the Shapiro-Wilk test for the dependent measures of experimental group connectives, comparison group of words before the main clause, the experimental group of the CIWS count, and for both groups for agentless passive voice and targeted connectives with *p*-values ranging from .000 to .034 (less than alpha level of .05). Skewness and kurtosis statistics were met for all dependent variables ranging from -1.592 to .556 (met if within an absolute value of 2.0), except for words before the main clause and agentless passive voice. The dependent variables of words before the main clause and agentless passive voice ranged in skewness from -2.733 to 5.022 and ranged in kurtosis from 4.802 to 25.799 (not within an absolute value of 2.0). Visual examination of Q-Q plots, histograms (shape), and boxplots (presence of outliers) were also used to determine normality. These revealed a normal distribution for connectives and CIWS count, but potential outliers for sentence length, words before the main clause, and a negatively skewed distribution and outliers for agentless passive voice and targeted connectives. Levene's test was completed to test the assumption of homogeneity of variances and was met for all dependent measures with a range of .173 to .642 ($p > .05$), except for sentence length (p =.039) and CIWS count (p =.012). Last, independent *t*-tests were completed to determine significance (met with $p < .05$, no significant differences in scores), and small to minimal effect by Cohen's *d* (Cohen, 1988). These results suggested the two treatment groups were relatively equivalent at baseline for all the dependent measures, with no statistical significance ranging from .121 to .834, and a small effect range of .06 to .29, except for CIWS count. The *t*-test for the CIWS count revealed $t(47) = 1.903$, p = .065, which was not significant; however, d =.56 indicated a medium effect or a medium proportion of difference between the two groups. The experimental group had a higher pretest score (M = 32.9048, SD =25.46155) than the comparison group (M =20.4643, SD = 18.24419). Therefore, the results of the CIWS count after statistical analysis should be interpreted with caution.

Assumptions testing results

The assumptions of independence, multivariate normality of dependent variables, linearity, and homogeneity of variances-covariances were tested before running the hierarchical repeated measures MANOVA model. The initial sample size was $n = 49$ (experimental group $n = 21$, comparison group $n = 28$). The assumption of independence was determined by plotting standardized residuals against levels of the independent variables in a scatterplot. The scatterplots generally suggested evidence of independence. There was a relatively random display of residuals above and below the horizontal line at zero for each category of the independent variables used to split the file. This suggested evidence that independence is a reasonable assumption. Standardized residuals were examined to determine univariate normality, a necessary condition for multivariate normality. The skewness and kurtosis statistics of the standardized residuals were not all within a range of an absolute value of 2.0. Review of the histograms, boxplots, and Q-Q plots revealed that there were eight potential outliers. All the outliers were greater than two standard deviations above or below the mean. The outliers were removed to achieve normal distribution. The resulting samples sizes were experimental group $n = 18$ (cases removed = 3), and comparison group $n = 21$ (cases removed = 5).

After removal of the outliers, visual examination of histograms, boxplots, and Q-Q plots suggested normality for the dependent variables. Review of the Shapiro-Wilk test for the experimental group and comparison group met assumptions of normality for most of the dependent variables, which were above alpha level of .05. The dependent variables that did not meet this assumption were pretest and posttest residuals for agentless passive voice, and pretest and posttest residuals for targeted connectives. The De Carlo (1997) SPSS macro for univariate normality also revealed that skewness for all variables met the assumption of normality ($p > .05$) except for the pretest ($p = .0000$) and posttest ($p = .0139$) of agentless passive voice standardized residuals. For kurtosis, De Carlo's (1997) SPSS macro for univariate normality revealed that all variables met the assumption of normality ($p > .05$) except for pretest for agentless passive voice ($p = .0005$). Influential points were examined by plotting Cook's distance against unleveraged values, and this revealed a relatively normal shape for all variables except the pretest and posttest for agentless passive voice. However, Cook's distance was less than one for all dependent variable residuals which indicated that no influence of individual cases were a major concern. In general, most forms of evidence suggested normality was a reasonable assumption. Although testing for univariate normality does not guarantee multivariate normality, departures from multivariate normality are usually negligible when univariate normality is met for each variable (Hahs-Vaughn, in progress).

Multivariate normality was examined using De Carlo's (1997) SPSS macro for multivariate normality and revealed that according to Small's test, skewness violated normality with p-value less than alpha level of .05 (χ^2 = 38.4372, *df* = 12, *p* = .0001), but Srivastava's test did not violate normality (χ = 16.3277, *df* = 12, *p* = .1767). For kurtosis, all the measures suggested normality (Small's variant χ^2 = 18.8510, *df* = 12, *p* = .0922; Srivastava's test = 3.1973, N(b2p) = .8932, *p* = .3718; and Mardia's test= 164.0297, N(b2p) = -.6934, *p* = .4880). The omnibus test of multivariate normality, Small's test variant, also suggested violations of normality (χ^2 = 57.2882, *df* = 24, *p* = .0002). Visual examination of the box and whisker plots revealed several pairs between the independent variables that were different suggesting lack of homogeneity of variances and co-variances. Although most of the dependent variables met the assumption of normality, some did not, which suggested violations to this assumption. However, violations of multivariate normality have minimal effect on Type I errors or rejecting the null hypothesis when it is true (Hahs-Vaughn, in progress).

The linearity of the dependent variables was examined by reviewing the matrix scatterplots of all pairs of dependent variables. All scatterplots revealed straight positive shapes, which suggested that the assumption of linearity was met.

The assumption for homogeneity of variances of the independent variables was met for with Levene's test, except for the pretest for sentence length (p = .028), and the pretest (p = .013) and posttest (p = .000) of agentless passive voice. Examination of boxplots of the dependent variables were all examined as a visual means to determine the extent to which equal variances can be assumed. Most of the boxplots for the dependent variables by group had varying box and whisker lengths, which suggested violation of homogeneity of variances. Because this assumption cannot be made with certainty, the omnibus test of Pillai's trace was used as it is more robust in MANOVA designs where homogeneity of variance-covariance is violated and less balanced (Hahs-Vaughn, in progress). Table 10 presents the data testing for the hierarchical repeated measures ANOVA assumptions. Table 11 presents the data testing for the measures of univariate normality. Table 12 presents the data testing for the measures of homogeneity of variances for the dependent variables.

Table 10: *Results of Assumptions Testing for the Pretest, Posttest Science Compare and Contrast Writing Prompt for Treatment-on-the-Treated Data*

Assumption	Test		Evidence	Assumption Satisfied?
Independence	Scatterplots	Standardized residuals	No observable trends	Yes
Univariate Normality	Shapiro-Wilk		Not met for all variables (see Table 11)	No
	Boxplot/ Histogram/ Q-Q Plot		Relatively normal shape for all variables	Yes
		Standardized residuals	Met for all variables except for pretest agentless passive voice comparison group (4.193)	No
		De Carlo (1997)	All were p <.05 except for pretest (*p* = .0000) and posttest (*p* = .0139) agentless passive voice	No
		Standardized residuals	Met for all variables except pretest agentless passive voice for the experimental (2.354) and comparison groups (19.154)	No
		DeCarlo (1997)	Met for all except pretest agentless passive voice (*p* = .0005)	No
	Scatterplot-Cook's vs. Unleveraged		Relatively normal shape for all variables except pretest and posttest for agentless passive voice	No
	Cook's Distance		All < 1.00	Yes
Multivariate Normality DeCarlo (1997)	Skewness	Small's test	χ^2= 38.4372, *df*= 12, *p* = .0001	No
		Srivastava's test	χ =16.3277, *df* 12, *p* = .1767	Yes
	Kurtosis	Small's variant	χ^2=18.8510, *df*=12, *p* = .0922	Yes
		Srivastava's test	χ = 3.1973, N(b2p) = .8932, *p* = .3718	Yes
		Mardia's test	b2p = 164.0297, N(b2p) = -.6934, *p* = .4880	Yes
	Omnibus	Small's variant	χ^2 = 57.2882, *df* = 24, *p* = .0002	No
Linearity	Scatterplots		Straight positive linear shapes	Yes
Homogeneity of variances-covariances	Levene's Test		Not met for all variables (see Table 12)	No
	Spread-vs-level plots		Most boxplots have varying box and whisker lengths	No

Table 11: *Shapiro-Wilk Test of Univariate Normality for Treatment Groups for the Science Compare and Contrast Writing Prompt for Treatment-on-the-Treated Data*

Dependent Variable Standardized Residual	Group	S-W	df	Significance	Assumption Satisfied?
PreSL	Experimental	.913	18	.097	Yes
	Comparison	.964	23	.555	Yes
PostSL	Experimental	.950	18	.418	Yes
	Comparison	.981	23	.918	Yes
PreCONN	Experimental	.943	18	.330	Yes
	Comparison	.919	23	.063	Yes
PostCONN	Experimental	.913	18	.096	Yes
	Comparison	.957	23	.401	Yes
PreWBMC	Experimental	.905	18	.070	Yes
	Comparison	.977	23	.853	Yes
PostWBMC	Experimental	.957	18	.542	Yes
	Comparison	.973	23	.756	Yes
PreAPV	Experimental	.506	18	.000*	No
	Comparison	.420	23	.000*	No
PostAPV	Experimental	.759	18	.000*	No
	Comparison	.872	23	.007*	No
PreTCONN	Experimental	.759	18	.000*	No
	Comparison	.872	23	.007*	No
PostTCONN	Experimental	.934	18	.230	Yes
	Comparison	.888	23	.014*	No
PreCIWS	Experimental	.953	18	.468	Yes
	Comparison	.981	23	.927	Yes
PostCIWS	Experimental	.947	18	.382	Yes
	Comparison	.949	23	.275	Yes

Note. Pre=pretest, Post=posttest, SL=sentence length, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, TCONN=targeted connectives, CIWS=correct versus incorrect word sequences

Table 12: *Levene's Test of Homogeneity of Variances for Dependent Variables for the Science Compare and Contrast Writing Prompt for Treatment-on-the-Treated Data*

Variable		df1	df2	Sig.
PreSL	3.078	4	36	.028*
PostSL	1.537	4	36	.212
PreCONN	.740	4	36	.571
PostCONN	1.403	4	36	.252
PreWBMC	1.363	4	36	.266
PostWBMC	2.458	4	36	.063
PreAPV	3.681	4	36	.013*
PostAPV	10.145	4	36	.000*
PreTCONN	.832	4	36	.514
PostTCONN	2.641	4	36	.050
PreCIWS	1.106	4	36	.369
PostCIWS	1.989	4	36	.117

Note. Pre=pretest, Post=posttest, SL=sentence length, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, TCONN=targeted connectives, CIWS=correct versus incorrect word sequences

Hierarchical repeated measures MANOVA results

The results for the hierarchical repeated measures MANOVA suggested that there was a statistically significant main effect ($F_{test} = 2.473$, $df = 6, 31$, $p = .045$) indicating that the combined dependent variables differed from pretest to posttest. Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a large effect and moderate power (partial $\eta^2_{test} = .324$, observed power = .746), as determined by Cohen (1988). Thirty-two percent of the total variance of the combined dependent variables on the science compare and contrast writing prompt can be accounted for by the within-subjects factor (i.e. time from pretest to posttest).

There was not a statistically significant interaction of treatment group and testing ($F_{treatment} = .306$, $df = 6, 31$, $p = .929$). Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a small effect and low power (partial $\eta^2_{treatment} = .056$, observed power = .120), as determined by Cohen (1988). Approximately 6% of the total variance of the combined dependent variables on the science compare and contrast writing prompt measures can be accounted for by treatment group.

In addition, there was not a statistically significant interaction for testing and the between-subjects factor of the nested classrooms by teacher within treatment group ($F_{classroom} = .954$, $df = 18, 88.167$, $p = .519$). Multivariate partial eta squared for the effect of the between-subjects factor of the nested classrooms by teacher within treatment group indicated a large effect and moderate power (partial $\eta^2_{test} = .152$, observed power = .654), as determined by Cohen (1988). Approximately 15% of the total variance of the combined dependent variables on the science compare and contrast writing prompt measures can be accounted for by the nested classrooms by teacher within treatment group. The large effect indicated that there was a large proportion of difference among classrooms. Review of pairwise comparisons revealed that there was a statistically significant difference ($p = .048$) between only two of the classes on the sentence length. Comparison class two had marginal mean of 14.261, which was much higher than experimental class two who had a marginal mean of 7.279.

The within-subjects contrasts indicated that for the main effect of difference between pretest and posttest was only statistically significant for the linear differential of the dependent variable of agentless passive voice from pretest to the posttest (F = 6.844, p =.013, partial η^2=.160, observed power = .721), as well as indicating large effect and moderate power (Cohen, 1988). The dependent variable of targeted connective was close to achieving significance (p = .054). There were no dependent variables that were statistically significant for the linear differential of the interaction for testing by treatment group (all p-values > .05). The dependent variable of agentless passive voice was also statistically significant for linear differential growth of the interaction of testing and the between-subjects factor of the nested classrooms by teacher within treatment group (F= 3.085 p = .039, partial η^2= .205, observed power = .672), indicating large effect and moderate power (Cohen, 1988). The second comparison class scored lower on the dependent variable of agentless passive voice (M = .001) as opposed to the other classes (all M > 2.50).

The experimental group achieved a higher marginal mean on the dependent variables of agentless passive voice (experimental M =5.55, SE = 2.046, CI = 1.406 to 9.704; comparison M = 3.926, SE = 1.659, CI= .562 to 7.289) and targeted connectives (experimental M = 1.546, SE =.225, CI = 1.091 to 2.002; comparison M =1.182, SE =.182, CI = .812 to 1.551). It should be noted that the experimental group did have a higher marginal mean by eight points on the CIWS count (M = 29.554, SE = 4.947, CI = 19.521 to 39.587) than the comparison group (M = 21.061, SE = 4.011, CI= 12.927 to 29.195). However, the baseline equivalency testing had indicated that there was a medium effect in the difference between the experimental and the comparison groups at pretest. The experimental group pretest score (*M* = 32.9048) was twelve points higher than the comparison group pretest score (*M* =20.4643). There was only a marginal mean difference of eight in favor of the experimental group after posttest; however, any gains for the comparison group would be conservative given the baseline difference. Therefore, the CIWS scores should be interpreted with caution. Table 13 presents the marginal means, standard deviation/error, and confidence intervals for the dependent variables by treatment group.

Table 13: *Marginal Means for Dependent Variables by Treatment Group for Science Compare and Contrast Writing Prompt for Treatment-on-the-Treated Data*

				95% Confidence Interval	
Variable	Treatment Group	Mean	Standard Error	Lower Bound	Upper Bound
SL	Experimental	10.280[a]	1.110	8.029	12.531
	Comparison	11.663[a]	.900	9.838	13.488
CONN	Experimental	71.854[a]	11.627	48.275	95.434
	Comparison	90.032[a]	9.426	70.915	109.148
WBMC	Experimental	3.121[a]	.378	2.355	3.886
	Comparison	3.100[a]	.306	2.479	3.721
APV	Experimental	5.555[a]	2.046	1.406	9.704
	Comparison	3.926[a]	1.659	.562	7.289
TCONN	Experimental	1.546[a]	.225	1.091	2.002
	Comparison	1.182[a]	.182	.812	1.551
CIWS	Experimental	29.554[a]	4.947	19.521	39.587
	Comparison	21.061[a]	4.011	12.927	29.195

[a] *Based on modified population marginal mean.*

Intent-to-Treat-Analysis

A hierarchical repeated measures MANOVA was then conducted using ITT analysis. As indicated previously, pretest scores were used to replace missing outcome data. This increased the sample size from $n = 49$ (experimental $n = 21$, comparison $n = 28$) to a larger sample size of $n = 60$ (experimental $n = 26$, comparison $n = 34$).

Baseline equivalency testing results

Equivalency testing was conducted on the baseline pretest scores of each of the six dependent variables between the two treatment groups (independent variable). Prior to running each of the independent *t*-tests, the assumptions of the normality of the distributions were tested with the Shapiro-Wilk test and were met for the dependent variables of sentence length, CIWS count, the comparison group of the connectives, experimental group for words before the main clause, with *p*-values ranging from .324 to .801 (greater than alpha level of .05). The assumptions of normality were not met with the Shapiro-Wilk test for the dependent measures of experimental group connectives, comparison group of words before the main clause, and for both groups for agentless passive voice, and targeted connectives with *p*-values ranging from .000 to .003 (less than alpha level of .05). Skewness and kurtosis statistics were met for all dependent variables with a range of -1.588 to .457 (met if within an absolute value of 2.0), except for the dependent variables of comparison group of words before the main clause, and both groups of agentless passive voice, which were all within a range of skewness of 2.144 to 5.538, and range of kurtosis of -7.546 to 31.388 (not within an absolute value of 2.0). Visual examination of Q-Q plots, histograms (shape), and boxplots (presence of outliers) were also used to determine normality. These revealed a normal distribution for all dependent measures except words before the main clause agentless passive voice, which were negatively skewed with potential outliers. Levene's test was completed to test the assumption of homogeneity of variances and was met for all dependent measures with a range of .087 to .714 (p > .05) except for CIWS count (p = .001). Last, the independent t-tests were completed to determine significance (met with p < .05, no significant differences in scores), and small to minimal effect by Cohen's d (Cohen, 1988). These results suggested the two treatment groups were relatively equivalent at baseline for all the dependent measures with no statistical significance ranging from .602 to .923, and a small effect sizes ranging from .02 to .21, except for the CIWS count. The independent t-test for the CIWS count revealed t(41.712) = 1.937, p = .059, which was not significant; however, d = .51 indicated a medium effect or a medium proportion of difference in scores that occurred between the two groups. The experimental group had a higher pretest score (M = 29.5769, SD =25.88617) than the comparison group (M =18.1471, SD = 17.52424). Therefore, the results of the CIWS count after statistical analysis should be interpreted with caution.

Assumptions testing results

The assumptions of independence, univariate normality, multivariate normality, linearity, and homogeneity of variances and covariances were tested before running this second hierarchical repeated measures MANOVA model. The assumption of independence was determined by plotting standardized residuals against levels of the independent variables in a scatterplot. The scatterplots generally suggested evidence of independence. There was a relatively random display of residuals above and below the horizontal line at zero for each category of the independent variables used to split the file. This suggested evidence that independence is a reasonable assumption.

Standardized residuals were used to determine univariate normality, a necessary condition for multivariate normality. The skewness and kurtosis statistics of the standardized residuals were not all within a range of an absolute value of 2.0. Review of the histograms, boxplots, and Q-Q plots revealed that there were eight potential outliers. The outliers were above and below two standard deviations from the mean. The outliers were removed to achieve normal distribution. The resulting samples sizes were experimental group $n = 24$ (cases removed = 2), and comparison group $n = 28$ (cases removed = 6).

After removal of the outliers, visual examination of histograms, boxplots, and Q-Q plots suggested normality for the dependent variables. Review of the Shapiro-Wilk test for the experimental group and comparison group met assumptions of normality for most of the dependent variables, which were above alpha level of .05. The dependent variables that did not meet the assumption with Shapiro Wilk test were pretest residual ($p = .039$) and posttest residual ($p = .000$) for connectives for the comparison group, and pretest and posttest of both groups for both agentless passive voice (all $p = .000$), and targeted connectives (experimental $p < .05$). Most variables met the assumption of normality with skewness and kurtosis (within an absolute value of 2.0) except for kurtosis for the pretest for agentless passive voice for th**e e**xperimental (2.186) and comparison (2.323) groups. The De Carlo (1997) SPSS macro for univariate normality also revealed that skewness of all variables met the assumption of normality (p > .05) except for the standardized residuals for pretest (p = .0000) and posttest (p = .0001) agentless passive voice and kurtosis for agentless passive voice. For kurtosis, De Carlo's (1997) SPSS macro for univariate normality revealed that all variables met the assumption of normality (p > .05) except for pre-test connectives (p =.0062), and pretest (p = .0001) and posttest (p = .0054) of agentless passive voice. Influential points were examined by plotting Cook's distance against unleveraged values and revealed a relatively normal shape for all variables except the pretest and posttest for agentless passive voice. However, Cook's distance was less than one for all dependent variable residuals indicating that no influence of individual cases that are a major concern. In general, most forms of evidence suggest normality is a reasonable assumption. Although testing for univariate normality does not guarantee multivariate normality, departures from multivariate normality are usually negligible when univariate normality are met for each variable (Hahs-Vaughn, in progress).

Multivariate normality was examined using DeCarlo's (1997) SPSS macro for multivariate normality and revealed that skewness violated normality according the Small's test (χ^2 = 56.3541, df= 12, p = .0000) and Srivastava's test (χ= 28.4348, df =12, p = .0048). Kurtosis tests revealed that kurtosis violated normality for Small's variant (χ^2 = 35.0787, df=12, p = .0005), Srivastava's test (χ = 3.8134, N(b2p) = 4.1477, p = .0000), but met normality assumptions for Mardia's test = 177.4990, N(b2p) = - 1.8684, p = .0617. The omnibus test of normality, Small's test variant, also suggested violations of normality (χ^2= 91.4328, df = 24, p = .0000). Visual examination of the box and whisker plots revealed several pairs between the independent variables that were different suggesting lack of homogeneity of variances and co-variances. Although most of the dependent variables met the assumption of normality, some did not suggesting violations to this assumption. However, violations of multivariate normality have minimal effect on Type I errors or rejecting the null hypothesis when it is true (Hahs-Vaughn, in progress).

The linearity of the dependent variables were examined by matrix scatterplots of all pairs of dependent variables. All scatterplots revealed straight positive shapes, which suggest that the assumption of linearity was met.

The assumption for homogeneity of variances of the independent variables was met with Levene's test except for the pretest for sentence length (p = .016), and the pretest (p = .005) and posttest (p = .000) of agentless passive voice, and posttest for CIWS (.021). Examination of boxplots of the dependent variables were all examined as a visual means to determine the extent to which equal variances can be assumed. Most of the boxplots for the dependent variables by group had varying box and whisker lengths, which also suggested violation of homogeneity of variances. Because this assumption can be not made with certainty, the omnibus test of Pillai's trace will be used as it is more robust in MANOVA designs where homogeneity of variance-covariance is violated and less balanced (Hahs-Vaughn, in progress). Table 14 presents the data testing for hierarchical repeated measures MANOVA assumptions. Table 15 presents the data testing for the measures of univariate normality. Table 16 presents the data testing for the measures of homogeneity of variances for the dependent variables.

Table 14: *Results of Assumptions Testing for the Pretest, Posttest Science Compare and Contrast Writing Prompt for Data Utilizing ITT Analysis*

Assumption	Test		Evidence	Assumption Satisfied?
Independence	Scatterplots	Standardized residuals	No observable trends	Yes
Univariate Normality	Shapiro-Wilk	Standardized residuals	Not met for all variables (see Table 15)	No
	Boxplot/ Histogram/ Q-Q Plot		Relatively normal shape for all variables	Yes
	Skewness	Standardized Residuals	Met for all variables except for all variables except for pretest agentless passive voice experimental group (2.186)	No
		DeCarlo (1997)	All were p <.05 except for pretest (*p* = .0000) and posttest (*p* = .0001) agentless passive voice	No
	Kurtosis	Standardized Residuals	Met for all variables except for posttest agentless passive voice comparison group (2.686)	No
		DeCarlo (1997)	All were p <.05 except for pre-test connectives (*p* =.0062), pretest (*p* = .0001) and posttest (*p* = .0054) agentless passive voice	No
	Scatterplot Cook's vs. Unleveraged		Relatively normal shape for all variables	Yes
	Cook's Distance		All <1	Yes
Multivariate Normality	Skewness	Small's test	χ^2= 56.3541, *df*= 12, *p* = .0000	No
		Srivastava's test	χ(b1p)= 28.4348, *df* =12, *p* = .0048	No
DeCarlo (1997)	Kurtosis	Small's variant	χ^2= 35.0787, *df* =12, *p* = .0005	No
		Srivastava's test	χ = 3.8134, N(b2p) = 4.1477, *p* = .0000	No
		Mardia's test	b2p = 177.4990, N(b2p) = - 1.8684 *p* = .0617	Yes
	Omnibus	Small's variant	χ^2= 91.4328, *df* = 24, *p* = .0000	No
Linearity	Matrix Scatterplots		Straight positive linear shapes for most variables	Yes
Homogeneity of variances-covariances	Levene's Test		Not met for all (see Table 16)	No
	Spread-vs-level plots		Most boxplots have varying box and whisker lengths	No

Table 15: *Shapiro-Wilk Test of Univariate Normality for Treatment Groups for the Science Compare and Contrast Writing Prompt for Data Utilizing ITT Analysis*

Dependent Variable Standardized Residual	Group	S-W	df	Significance	Assumption Satisfied?
PreSL	Experimental	.953	24	.307	Yes
	Comparison	.963	28	.407	Yes
PostSL	Experimental	.975	24	.782	Yes
	Comparison	.959	28	.335	Yes
PreCONN	Experimental	.921	24	.061	Yes
	Comparison	.922	28	.039*	No
PostCONN	Experimental	.949	24	.254	Yes
	Comparison	.628	28	.000*	No
PreWBMC	Experimental	.947	24	.234	Yes
	Comparison	.963	28	.410	Yes
PostWBMC	Experimental	.971	24	.689	Yes
	Comparison	.946	28	.160	Yes
PreAPV	Experimental	.557	24	.000*	No
	Comparison	.724	28	.000*	No
PostAPV	Experimental	.798	24	.000*	No
	Comparison	.830	28	.000*	No
PreTCONN	Experimental	.917	24	.049*	No
	Comparison	.911	28	.021*	No
PostTCONN	Experimental	.887	24	.012*	No
	Comparison	.918	28	.030*	No
PreCIWS	Experimental	.942	24	.178	Yes
	Comparison	.982	28	.903	Yes
PostCIWS	Experimental	.942	24	.179	Yes
	Comparison	.907	28	.017*	No

Note. Pre=pretest, Post=posttest, SL=sentence length, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, TCONN=targeted connectives, CIWS=correct versus incorrect word sequences

Table 16: *Levene's Test of Homogeneity of Variances for Dependent Variables for the Science Compare and Contrast Writing Prompt for Data Utilizing ITT Analysis*

Variable	F	df1	df2	Sig.
PreSL	3.399	4	47	.016*
PostSL	1.502	4	47	.217
PreCONN	1.315	4	47	.278
PostCONN	1.190	4	47	.328
PreWBMC	.567	4	47	.688
PostWBMC	1.997	4	47	.110
PreAPV	4.266	4	47	.005*
PostAPV	8.219	4	47	.000*
PreTCONN	.540	4	47	.707
PostTCONN	4.695	4	47	.003*
PreCIWS	1.633	4	47	.182
PostCIWS	3.190	4	47	.021*

Note. Pre=pretest, Post=posttest, SL=sentence length, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, TCONN=targeted connectives, CIWS=correct versus incorrect word sequences

Hierarchical repeated measures MANOVA results

The results for the hierarchical repeated measures MANOVA from the data utilizing ITT analysis suggest that there was a statistically significant main effect ($F_{test} = 3.149$, $df = 6, 42$, $p = .012$), which indicated that the combined dependent variables differed between pretest and posttest . Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a large effect and moderate power (partial $\eta^2_{test} = .310$, observed power = .876), as determined by Cohen (1988). Thirty-one percent of the total variance of the combined dependent variables on the science compare and contrast writing prompt can be accounted for by the within-subjects factor (i.e., time from pretest to posttest).

There was not statistically significant interaction of treatment group and testing ($F_{treatment} = .313$, $df = 6, 42$, $p = .927$). Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a small effect and low power (partial $\eta^2_{treatment} = .043$, observed power = .120), as determined by Cohen (1988). Four percent of the total variance of the combined dependent variables on the science compare and contrast writing prompt measures can be accounted for by treatment group.

In addition, there was no significant interaction for testing and the between subjects factor of the nested classrooms by teacher within treatment group ($F_{classroom} = .483$, $df = 18, 119.279$, $p = .516$). Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a large effect and moderate power (partial $\eta^2_{classroom} = .118$, observed power = .671), as determined by Cohen (1988). Approximately 12% of the total variance of the combined dependent variables on the science compare and contrast writing prompt can be accounted for by the nested classrooms by teacher within treatment group. The large effect indicated that there was a larger proportion of difference in scores among classes. Pairwise comparisons revealed two dependent variables that were statistically significant between the same two classes. The first was sentence length (p = .036), in which experimental class one had a higher marginal mean (M = 13.859) than the comparison class one (M = 8.698). The CIWS count was also significant (p = .032). Experimental class one had a higher marginal mean (M = 37.00) than comparison class one (M= 14.607). At baseline, there was a medium effect detected between pretest scores. The experimental group had a higher pretest score (M = 29.5769) than the comparison group (M =18.1471) by almost 11 points. There was a considerably larger marginal mean difference of 23 between experimental class one and comparison class one after posttest, which may reflect gains in favor of experimental class one. However, any gains for the comparison group would be conservative given the baseline difference; therefore, the CIWS scores should again be interpreted with caution. There were no other significant differences among any of the other classes.

The within-subjects contrasts indicated that for the main effect of difference between pretest and posttest, was only statistically significant for the linear differential of the dependent variable of agentless passive voice from pretest to the posttest (F = 11.576, p =.001, partial η^2=.198, observed power = .915), as well as indicating large effect and large power. The dependent variable of words before the main clause was close to achieving significance (p = .054). There were no dependent variables that were statistically significant for the linear differential of the interaction for testing by treatment group (all *p*-values >.05). The dependent variable of agentless passive voice was also statistically significant for linear differential growth of the interaction of testing and the between subjects factor of the nested classrooms by teacher within treatment group ($F = 3.029$, $p = .039$, partial $\eta^2 = .205$, observed power = .676), which indicated a large effect and moderate power. The second comparison class scored lower on the dependent variable of agentless passive voice ($M < .001$) as opposed to the other classes (all $M >$ 1.80).

The experimental group achieved higher marginal means on the dependent variables of sentence length (experimental *M* = 11.283 *SE* = .956, *CI* = 9.361 to 13.206; comparison *M* = 10.975, *SE* =.911, *CI* = 9.143 to 12.807), words before the main clause (experimental *M* = 3.211, *SE* = .305, *CI* = 2.598 to 3.824; comparison *M* = 2.975, *SE* = .290, *CI* = 2.391 to 3.559), agentless passive voice (experimental *M* = 6.139, *SE* = 1.763, *CI* = 2.593 to 9.685; comparison *M* =3.679, *SE* = 1.679, *CI* = .301 to 7.058), targeted connectives (experimental *M* = 1.400, *SE* = .205, *CI* =.987 to 1.813; comparison *M* = 1.184, *SE* = .196, *CI* = = .790 to 1.577), and CIWS count (experimental *M* =28.639, *SE* = 4.087, *CI* = 20.416 to 36.862; comparison, *M* = 19.376, *SE* = 3.894, *CI* = 11.542 to 27.210). Table 17 presents the marginal means, standard error, and confidence intervals for the dependent variables by treatment group.

Table 17: *Marginal Means for Dependent Variables by Treatment Group for the Science Compare and Contrast Writing Prompt for Dependent Variables for Sample Utilizing ITT Analysis*

				95% Confidence Interval	
Variable	Treatment Group	Mean	Std. Error	Lower Bound	Upper Bound
SL	Experimental	11.283[a]	.956	9.361	13.206
	Comparison	10.975[a]	.911	9.143	12.807
CONN	Experimental	79.850[a]	9.748	60.240	99.461
	Comparison	85.684[a]	9.287	67.000	104.367
WBMC	Experimental	3.211[a]	.305	2.598	3.824
	Comparison	2.975[a]	.290	2.391	3.559
APV	Experimental	6.139[a]	1.763	2.593	9.685
	Comparison	3.679[a]	1.679	.301	7.058
TCONN	Experimental	1.400[a]	.205	.987	1.813
	Comparison	1.184[a]	.196	.790	1.577
CIWS	Experimental	28.639[a]	4.087	20.416	36.862
	Comparison	19.376[a]	3.894	11.542	27.210

[a] *Based on modified population marginal mean. Note. SL=sentence length, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, TCONN=targeted connectives, CIWS=correct versus incorrect word sequences*

Morph-Syntactical Analysis for Both Samples

Morpho-syntactical errors were calculated in addition to writing conventions and semantic errors as part of the dependent variable measure of the CIWS count. However, additional descriptive statistics were calculated via language sample analysis to determine the number and type of morpho-syntactical errors that students exhibited on the science compare and contrast writing prompt. The treatment-on-the-treated-data sample was 41 students (18 experimental, 23 comparison). There were 19 out of 41 (46%) students who exhibited no errors on both the pretest and posttest. There were 22 out 41 (54%) students who exhibited morpho-syntactical errors. Out of the 22 student with errors, there were 9 experimental students (50% of this group) and 13 comparison students (56% of this group).

The ITT data sample was 52 students (24 experimental, 28 comparison). There was a total of 34 out of 52 (65%) students who exhibited no errors on both the pretest and posttest. There were 18 out 52 (35%) of students who exhibited morpho-syntactical errors. Out of the 18 student with errors, there were 8 experimental students (25% of this group) and 10 comparison students (35% of this group) who exhibited morpho-syntactical errors.

According to Table 18, the most common morpho-syntactical errors were similar across both samples. The morpho-syntactical errors that occurred in descending order were (1) subject-verb agreement plural form, (2) subject-verb agreement singular form, (3) regular past tense verbs, (4)plurals, (5) irregular past tense verbs, (6) particles, (7) prepositions, and (8) present perfect tense.

Table 18: *Descriptive Statistics of the Type of Morpho-Syntactical Errors on the Science Compare and Contrast Writing Prompt for the Treatment-on-the-Treated and ITT Analysis Data*

Morpho-Syntactical Error	Total Errors for the Treatment-on-the Treated Data Sample	Total Errors for the ITT Data Sample
Subject verb agreement plural errors	50%	48%
Subject verb agreement singular errors	23%	18%
Regular Past Tense Verb errors	10%	9%
Plural errors	3.4%	6%
Irregular Past Tense Verb errors	3.4%	6%
Article errors	3.4%	4%
Preposition errors	3.4%	9%
Present perfect errors	3.4%	0
Possessive errors	0	0
Inconsistent verb tense errors	0	0
Infinitive errors	0	0
Present tense verb error	0	0

Summary of the Analyses

In summary, both the treatment-on-the-treated sample and the sample with ITT analysis yielded similar results. Both analyses indicated statistical significance ($p < .05$) of the difference of the combined dependent variables between the pretest and posttest scores on the science compare and contrast writing prompt, with large effect and moderate power. Both analyses indicated no statistically significant interaction of treatment group and test of the combined dependent variables on the science compare and contrast writing prompt, with small effect and low power. Both analyses indicated no statistically significant interaction of the nested classrooms by teacher within treatment group of the combined dependent variables on the science compare and contrast writing prompt, with large effect and moderate power. The large effect and moderate power indicated a large proportion of difference in scores among classes. For both samples, pairwise comparisons revealed statistically significant differences between only two classes in each sample. For the treatment-on-the treated sample there was a difference in favor of comparison class two over experimental class two in sentence length. On the ITT analysis, experimental class one had higher marginal means on both sentence length and CIWS count than comparison class one. The within-subjects contrasts indicated that for both analyses, there was statistical significance ($p < .05$) for the dependent variable of agentless passive voice for both the pretest, posttest and for interaction of the nested classrooms by teacher. For the treatment-on-the-treated sample, the marginal means for the experimental group were higher than the comparison group on agentless passive voice, targeted connectives, and correct word sequences. The sample with ITT analysis indicated higher marginal means for the experimental group than the comparison group on those same dependent variables, as well as sentence length and words before the main clause. Analysis of morpho-syntactical errors revealed similar error patterns on both the treatment-on-treated and ITT analysis data samples.

Research Question Three

Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase in specific aspects of sentence complexity when writing a science expository essay as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

First, a hierarchical repeated measures multivariate analysis of variance (MANOVA) was computed to answer this question first applying treatment-on-the-treated analysis (i.e., analyzing only data from students who had no missing outcome data), $n = 49$; and then utilizing ITT analysis, $n = 60$. For the hierarchical repeated measures MANOVA, the independent variable was treatment group (experimental or comparison), the dependent variables were (1) sentence connectives, (2) words before the main clause, (3) agentless passive voice, (4) noun phrase density, (5) verb phrase density, (6) prepositional phrase density, and the (7) number of correct versus incorrect word sequences, and the hierarchical factor was the classroom (teacher). The hierarchical repeated measures MANOVA was conducted to determine if the within-subjects factor, the pretest and post science expository essay measures were significant, and if there was a significant interaction with the between-subjects factors of treatment group (experimental or comparison), and nesting for the classroom (i.e., teacher).

Treatment-on-the-Treated Analysis

Baseline equivalency testing results

Equivalency testing was conducted on the baseline pretest scores of each of the six dependent variables between the two treatment groups (independent variable). Prior to running each of the independent *t*-tests, the assumptions of the normality of the distributions were tested with the Shapiro-Wilk test and was met for the experimental group of words before the main clause and agentless passive voice, and the comparison group for verb phrase density, with all *p*-values ranging from .145 to .479 (greater than alpha level of .05). The Shapiro-Wilk test did not meet the assumption of normality for the dependent variables for both treatment groups for connectives, noun phrase density, prepositional phrase density, and CIWS count, as well as the experimental group for verb phrase density, and the comparison group for words before the main clause and agentless passive voice. The variables *p*-values ranged from .000 to .046 (less than alpha level of .05). Skewness and kurtosis statistics were met for all dependent variables with a range of -1.336 to .638 (met if within an absolute value of 2.0). Skewness and kurtosis statistics were not met for the dependent variables of the experimental group for verb phrase density, comparison group for words before the main clause and prepositional phrase density, and both treatment groups for agentless passive voice, which were all in a range for skewness from -2.303 to 3.068 and all in a range for kurtosis from -2.858 to 10.481 (not within an absolute value of 2.0). Visual examination of Q-Q plots, histograms (shape), and boxplots (presence of outliers) were also used to determine normality. These revealed a negatively skewed distribution and potential outliers for all the dependent measures. Levene's test was completed to test the assumption of homogeneity of variances and was met for all dependent measures with a range of .117 to .543 ($p > .05$). Last, the independent *t*-tests were completed to determine significance (met with $p < .05$, no significant differences in scores), and small to minimal effect by Cohen's *d* (Cohen, 1988). These results suggested the two treatment groups were relatively equivalent at baseline for all the dependent measures with no statistical significance ranging from .283 to .968, and a minimal to small effect range of .01 to .32.

Assumptions testing results

The assumption of independence was determined by reviewing scatterplots of each pretest and posttest standardized residuals, which generally suggested evidence of independence, with a relatively random display of difference scores above and below zero, with no cyclical pattern.

Visual examination of histograms, boxplots, and Q-Q plots suggested non-normality for a few dependent variables. There was the presence of 13 outliers; however, removal of all these cases would significantly reduce the sample size, as well as possible power and effect. The researcher proceeded with analysis with all cases with a sample size of $n = 49$ (experimental $n = 21$, and comparison $n = 28$). The full data analysis of treatment-on-the-treated data with removal of the outliers is listed in Appendix V. The sample with the outliers removed is compared to this sample in the summary addressed later in this chapter.

Review of the Shapiro-Wilk test for experimental group and comparison groups did not meet the assumptions of normality (*p* >.05) of about half of the variables (see Table 24). All variables did not meet the assumption of normality for skewness and kurtosis (within a value of absolute 2.0 or a liberal value of absolute 3.0). The variables that did not meet this assumption were pretest words before the main clause comparison group (skewness = 4.120, kurtosis = 19.794); posttest words before the main clause experimental (skewness = 2.817 , kurtosis = 11.332) and comparison (skewness = 1.955, kurtosis = 7.412); pretest agentless passive voice experimental (skewness = 2.833, kurtosis =9.654) and comparison (skewness = 2.629, kurtosis = 6.751); and posttest agentless passive voice experimental (skewness = 2.725, kurtosis = 8.815) and comparison (skewness = 2.348, kurtosis = 5.256); and pretest preposition phrase density comparison (skewness = 1.890, kurtosis = 7.080). According to West, Finch, and Curran (1996), kurtosis only above 7.0 is a threat to normality, and according Hahs-Vaughn (in press) departures from normality are not particularly concerning if confined to non-zero kurtosis. Therefore, the violations of normality observed were not concerning given these parameters. The De Carlo (1997) SPSS macro for univariate normality also revealed that skewness for most variables did not meet the assumption of normality (*p* > .05) including pretest and posttest for words before the main clause, pretest and posttest for agentless passive voice, posttest for verb phrase density, pretest and posttest noun phrase density, pretest for preposition phrase density, and pretest for correct versus incorrect word sequences (CIWS). For kurtosis, De Carlo’s (1997) SPSS macro for univariate normality revealed that some variables did not meet the assumption of normality (*p* > .05) including pretest and posttest for agentless passive voice, pretest and posttest for words before the main clause, and pretest for prepositional phrase density. Influential points were examined by plotting Cook's distance against unleveraged values, and revealed a relatively normal shape for all variables. However, Cook's distance was less than one for all dependent variable residuals indicating that no influence of individual cases were a major concern. In general, most forms of evidence suggest normality was a reasonable assumption. Although testing for univariate normality does not guarantee multivariate normality, departures from multivariate normality are usually negligible when univariate normality is met for each variable (Hahs-Vaughn, in progress).

Multivariate normality was examined using De Carlo's (1997) SPSS macro for multivariate normality and revealed that according to Small's test, skewness violated normality with an alpha level above .05 ($\chi^2 = 193.0385$, $df = 14$, $p = .0000$), as well as Srivastava's test of normality (χ (b1p)= 85.2394, df =14, $p = .0000$). For kurtosis, all the measures did not suggest normality (Small's variant χ^2=112.6694, $df = 14$, $p = .0000$, Srivastava's test $\chi = 4.5094$ N(b2p) = 8.0697, $p = .0000$, and Mardia's test = 271.2555 , N(b2p) = 7.8142, $p = .0000$). The omnibus test of multivariate normality, Small's test variant, also suggested violations of normality (χ^2= 305.7080, $df = 28$, $p = .0000$). Visual examination of the box and whisker plots revealed several pairs between the independent variables that were different suggesting lack of homogeneity of variances and co-variances. Most of the dependent variables did not meet the assumption of normality. Violations of multivariate normality have minimal effect on Type I errors or rejecting the null hypothesis when it is true (Hahs-Vaughn, in progress).

The linearity of the dependent variables were examined by matrix scatterplots of all pairs of dependent variables, All scatterplots revealed straight positive shapes, which suggested that the assumption of linearity was met.

The assumption for homogeneity of variances of the independent variables was met for most variables with Levene's test except for the posttest for words before the main clause (p = .015) , posttest for verb phrase density (p = .005), pretest and posttest for noun phrase density (p < .022), and posttest for prepositional phrase density (p < .016); with a levels below alpha level of .05. Examination of boxplots of the dependent variables were all examined as a visual means to determine the extent to which equal variances can be assumed. Most of the boxplots for the dependent variables by group had varying box and whisker lengths, which also suggested violation of homogeneity of variances. Because this assumption can be not made with certainty, the omnibus test of Pillai's trace was used as it is more robust in MANOVA designs where homogeneity of variance-covariance is violated and less balanced (Hahs-Vaughn, in progress). Table 19 presents the data testing for hierarchical repeated measures MANOVA assumptions. Table 20 presents the data testing for the measures of univariate normality. Table 21 presents the data testing for the measures of homogeneity of variances for the dependent variables.

Table 19: *Results of Assumptions Testing for the Pretest, Posttest Science Expository Essay for Treatment-on-the-Treated Data (All Cases)*

Assumption	Test		Evidence	Assumptions Satisfied?
Independence	Matrix Scatterplots		No observable trends	Yes
Univariate Normality	Shapiro-Wilk		Not met for all (see Table 20)	No
	Boxplot/ Histogram/ Q-Q Plot		Suggested nonnormality for some variables with outliers present	No
	Skewness	Standardized Residuals	Not met for all variables	No
		DeCarlo (1997)	Not met for all variables at p >.05	No
	Kurtosis	Standardized Residuals	Not met for all variables	No
		DeCarlo (1997)	Not met for all variables at p >.05	No
	Scatterplot (Cook's vs. Unleveraged values)		Relatively normal shape for all variables	Yes
	Cook's Distance		All < 1.00	Yes
Multivariate Normality	Skewness	Small's test	χ^2 = 193.085, *df* = 14.0000, *p* = .0000	No
		Srivastava's test	χ (b1p)=85.2394, *df* =14.000, *p* = .0000	No
DeCarlo (1997)	Kurtosis	Small's variant	χ^2=112.6694, *df*=14.0000 , *p* = .0000	No
		Srivastava's test	χ = 4.5094, N(b2p) = 8.0697, *p* = .000	No
		Mardia's test	b2p = 271.2555 , N(b2p) = 7.8142, *p* = .00	No
	Omnibus	Small's variant	χ^2= 305.7080, *df* = 28.0000 *p* = .0000	No
Linearity	Matrix Scatterplots		Straight positive linear shapes	Yes
Homogeneity of variances-covariances	Levene's Test		Not met for all (see Table 21)	No
	Spread-vs-level plots		Most boxplots have varying box and whisker lengths	No

Table 20: *Shapiro-Wilk Test of Univariate Normality for Treatment Groups for the Science Expository Essay for Treatment-on-the-Treated-Data (All Cases)*

Dependent Variable	Group	S-W	df	Significance	Assumption Satisfied?
PreCONN	Experimental	.878	21	.013*	No
	Comparison	.904	28	.014*	No
PostCONN	Experimental	.962	21	.558	Yes
	Comparison	.942	28	.127	Yes
PreWBMC	Experimental	.969	21	.713	Yes
	Comparison	.536	28	.000*	No
PostWBMC	Experimental	.681	21	.000*	No
	Comparison	.827	28	.000*	No
PreAPV	Experimental	.636	21	.000*	No
	Comparison	.585	28	.000*	No
PostAPV	Experimental	.654	21	.000*	No
	Comparison	.569	28	.000*	No
PreVPD	Experimental	.875	21	.012*	No
	Comparison	.944	28	.141	Yes
PostVPD	Experimental	.901	21	.036*	No
	Comparison	.936	28	.086	Yes
PreNPD	Experimental	.861	21	.007	No
	Comparison	.835	28	.000*	No
PostNPD	Experimental	.837	21	.003*	No
	Comparison	.937	28	.095	Yes
PrePPD	Experimental	.927	21	.121	Yes
	Comparison	.834	28	.000*	No
PostPPD	Experimental	.963	21	.580	Yes
	Comparison	.960	28	.353	Yes
PreCIWS	Experimental	.918	21	.078	Yes
	Comparison	.882	28	.005*	No
PostCIWS	Experimental	.973	21	.799	Yes
	Comparison	.954	28	.247	Yes

Note. Pre=pretest, Post=post-test, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences

Table 21: *Levene's Test of Homogeneity of Variances for Dependent Variables for the Science Expository Essay for Treatment-on-the Treated Data (All Cases)*

Variable	F	df1	df2	Sig.
PreCONN	2.151	4	31	.098
PostCONN	1.917	4	31	.133
PreWBMC	.711	4	31	.590
PostWBMC	3.161	4	31	.027*
PreAPV	3.842	4	31	.012*
PostAPV	3.535	4	31	.017*
PreVPD	1.703	4	31	.174
PostVPD	6.436	4	31	.001*
PreNPD	11.258	4	31	.000*
PostNPD	4.186	4	31	.008*
PrePPD	3.375	4	31	.021*
PostPPD	3.279	4	31	.024*
PreCIWS	2.338	4	31	.077
PostCIWS	2.286	4	31	.082

Note. Pre=pretest, Post=post-test, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences

Hierarchical repeated measure MANOVA results

The results for the hierarchical repeated measures ANOVA suggest that there was not a statistically significant main effect ($F_{test} = .300$, $df = 7, 38$, $p = .949$) indicating that the combined dependent variables did not differ from pretest to posttest. Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a small effect and low power (partial $\eta^2_{test} = .052$, observed power = .126), as determined by Cohen (1988). Approximately 5% of the total variance of the combined dependent variables on the science expository can be accounted for by the within-subjects factor (i.e. time from pretest to posttest).

There was not a statistically significant interaction of treatment group and testing ($F_{treatment} = .981$, $df = 7, 38$, $p = .459$). Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a large effect and low power (partial $\eta^2_{treatment} = .153$, observed power = .363), as determined by Cohen (1988). Approximately 15% of the total variance of the combined dependent variables can be accounted for by treatment group. There was a large effect, which indicated a large proportion of difference in scores between groups. The experimental group scored a higher marginal mean on the CIWS count (experimental *M* =59.714, *SE* = 9.471, *CI* = 40.626 to 78.803), than the comparison group (*M* = 45.644, *SE* =7.922, *CI* = 29.678 to 61.609). The experimental group increased four points from a pretest (*M* = 61.9524, *SD* = 54.79642) to posttest (*M*= 65.2857, *SD* = 43.88296). In contrast, the comparison group decreased more than five points from pretest (*M* = 49.1071, *SD* = 43.75973) to posttest (*M* = 37.5357, *SD* = 32.95281). See Table 22 for pretest means, posttest means, marginal means, standard deviation/error, and confidence intervals by treatment group for the CIWS count.

Table 22: *Means, Standard Deviation/Error, and Confidence Intervals by Treatment Group for the Science Expository Essay for Correct Versus Incorrect Word Sequences Count for the Treatment-on-the-Treated Data (All Cases)*

Treatment Group		Score	Standard. Deviation	Standard Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Experimental	Pretest Mean	61.9524	54.79642			
	Posttest Mean	65.2857	43.88296			
	Marginal Mean	59.714		9.471	40.626	78.803
Comparison	Pretest Mean	49.1071	43.75973			
	Posttest Mean	37.5357	32.95281			
	Marginal Mean	45.644[a]		7.922	29.678	61.609

[a] *Based on modified population marginal mean.*

In addition, there was no statistically significant interaction for testing and the between-subjects factor of the nested classrooms by teacher within treatment group ($F_{classroom}$= .765, *df* = 21, 120, *p* = .756). Multivariate partial eta squared for the effect of the between-subjects factor of the nested classrooms by teacher within treatment group indicated a medium effect and moderate power (partial η^2_{test} =.118, observed power = .566), as determined by Cohen (1988). Approximately 12% of the total variance of the combined dependent variables on the science compare and contrast writing prompt measures can be accounted for by the nested classrooms by teacher within treatment group. There was medium effect, which indicated a medium proportion of difference in scores between classes. The experimental class one's marginal mean was in a range of 8-36 points higher than the other classrooms on the CIWS count (*M* =71.429, *SE*

=10.937, *CI* = 49.387 to 93.470). In addition, experimental class one was the only class to increase of 4.98 in mean score from pretest (*M*= 68.9286, *SD* = 51.18) to posttest (*M* = 73.9286, *SD* = 36.50809). Experimental class two maintained the same mean score of 48.000 from pretest to posttest. The three comparison classes decreased in mean score from pretest to posttest in a range of 6-15 points. See Table 23 for the CIWS count mean scores by classroom for the science expository essay for treatment-on-the-treated data.

Table 23: *Means, Standard Deviation/Error, and Confidence Intervals by Classroom for the Science Expository Essay for Correct Versus Incorrect Word Sequences Count for the Treatment-on-the-Treated Data (All Cases)*

					95% Confidence Interval	
Treatment Group		Score	Standard. Deviation	Standard Error	Lower Bound	Upper Bound
Experimental Class 1	Pretest Mean	68.9286	51.18062			
	Posttest Mean	73.9286	36.50809			
	Marginal Mean	71.429[a]		10.937	49.387	93.470
Experimental Class 2	Pretest Mean	48.0000	63.18755			
	Posttest Mean	48.0000	54.84828			
	Marginal Mean	48.000[a]		15.467	16.829	79.171
Comparison Class 1	Pretest Mean	43.4167	38.30608			
	Posttest Mean	28.3333	24.10331			
	Marginal Mean	35.875[a]		11.813	12.068	59.682
Comparison Class 2	Pretest Mean	43.2222	56.06197			
	Posttest Mean	31.8889	33.29957			
	Marginal Mean	37.556[a]		13.640	10.065	65.046
Comparison Class 3	Pretest Mean	66.4286	35.63171			
	Posttest Mean	60.5714	38.81519			
	Marginal Mean	63.500[a]		15.467	32.329	94.671

[a] *Based on modified population marginal mean.*

Intent-to-Treat-Analysis

Baseline equivalency testing results

Equivalency testing was conducted on the baseline pretest scores of each of the six dependent variables between the two treatment groups (independent variable). Prior to running each of the independent *t*-tests, the assumptions of the normality of the distributions were tested with the Shapiro-Wilk test, and was met only for the experimental group for prepositional phrase density $p = .246$ (greater than alpha level of .05). The Shapiro-Wilk test did not meet the assumption of normality for the rest of the dependent variables with all *p*-values ranging from .000 to .040 (less than alpha level of .05). Skewness and kurtosis statistics were met for all dependent variables with a range of -.153 to 1.572 (met if within an absolute value of 2.0), except the dependent variables of experimental group of connectives, comparison group of words before the main clause, and prepositional phrase density, and both treatment groups for agentless passive voice and noun phrase density. These variables were in a range for skewness from -2.034 to 4.873; and in a range for kurtosis from -2.858 to 12.426 (not within an absolute value of 2.0). Visual examination of Q-Q plots, histograms (shape), and boxplots (presence of outliers) were also used to determine normality. These revealed a negatively skewed distribution and potential outliers for all dependent measures. Levene's test was completed to test the assumption of homogeneity of variances and was met for all dependent measures with a range of .094 to .482 ($p > .05$). Last, the independent *t*-tests were completed to determine significance (met with $p < .05$, no significant differences in scores), and small to minimal effect by Cohen's *d* (Cohen, 1988). These results suggested the two treatment groups were relatively equivalent at baseline for the all dependent measures with no statistical significance ranging from .319 to .727, and a minimal to small effect range of .011 to .32, except for the dependent variable of words before the main clause. The independent *t*-test for words before the main clause revealed $t(58) = -.695$, $p = .490$, which was not significant; however, $d = .62$ indicated a medium effect or medium proportion of difference in scores between the two groups. The experimental group had a lower pretest mean ($M = 3.3719$, $SD = 1.84913$) than the comparison group ($M = 4.3344$, $SD = 6.86298$). Therefore, the results of the dependent variables words before the main clause after statistical analysis should be interpreted with caution.

Assumptions testing results

The assumption of independence was determined by reviewing scatterplots of each pretest and posttest standardized residuals, which generally suggested evidence of independence, with a relatively random display of difference scores above and below zero, with no cyclical pattern.

Visual examination of histograms, boxplots, and Q-Q plots suggested non-normality for a few dependent variables. There was the presence of sixteen outliers; however, removal of all those cases would significantly reduce the sample size, as well as possible power and effect. The researcher proceeded with analysis with all cases with a sample size of n = 60 (experimental *n* = 24, and comparison *n* = 36). The full data analysis for the ITT analysis data with removal of the outliers is listed in Appendix W. The sample with the outliers removed is compared to this sample in the summary addressed later in this chapter.

Review of the Shapiro-Wilk test for experimental group and comparison group did not meet the assumptions of normality (p >.05) for most of the variables (see Table 26). All variables did not meet the assumption of normality with skewness and kurtosis (within a value of absolute 2.0) or a liberal value of absolute 3.0. The variables that did not meet this assumption were the pretest for words before the main clause for the comparison group (skewness = 4.652, kurtosis = 24.785); posttest for words before the main clause for the experimental group (skewness = 2.962 , kurtosis = 12.882) and comparison group (skewness = 1.940, kurtosis = 7.642); and the pretest for agentless passive voice for experimental (skewness = 3.039, kurtosis =

11.610) and comparison groups (skewness = 2.961, kurtosis = 8.879); and the posttest for agentless passive voice for the experimental (skewness = 3.063, kurtosis = 11.322) and comparison groups (skewness = 2.663, kurtosis = 6.950); and the pretest for noun phrase density of the experimental group (skewness =-.914, kurtosis = 3.385). According to West et al. (1996), kurtosis only above 7.0 is a threat to normality, and according Hahs-Vaughn (in press) departures from normality are not particularly concerning if confined to non-zero kurtosis. Therefore, violations of normality observed were not concerning given these parameters. The De Carlo (1997) SPSS macro for univariate normality also revealed that skewness for most variables did not meet the assumption of normality ($p > .05$) including pretest and posttest for words before the main clause, pretest and posttest for agentless passive voice, posttest verb phrase density, pretest and posttest for noun phrase density; pretest for preposition phrase density, and pretest for correct versus incorrect word sequences (CIWS). For kurtosis, De Carlo's (1997) SPSS macro for univariate normality revealed that most variables did not meet the assumption of normality ($p > .05$) including pretest and posttest for agentless passive voice, pretest and posttest for words before the main clause, posttest for noun phrase density, pretest and posttest for noun phrase density, and pretest for prepositional phrase density. Influential points were examined by plotting Cook's distance against unleveraged values, and revealed a relatively normal shape for all variables. However, Cook's distance was less than one for all dependent variable residuals indicating that no influence of individual cases were a major concern. In general, most forms of evidence suggest normality was a reasonable assumption. Although testing for univariate normality does not guarantee multivariate normality, departures from multivariate normality are usually negligible when univariate normality is met for each variable (Hahs-Vaughn, in progress).

Multivariate normality was examined using De Carlo's (1997) SPSS macro for multivariate normality and revealed that according to Small's test, skewness violated normality with an alpha level above .05 (χ^2 = 239.7435, *df* = 14, *p* = .0000), as well as Srivastava's test of normality (χ (b1p)=61.7129, *df* =14, *p* = .0000). For kurtosis, all the measures did not suggest normality (Small's variant χ^2=132.8015, *df*=14, *p* = .0000, Srivastava's test χ = 4.5024 N(b2p) = 8.8884, *p* = .0000, and Mardia's test = 299.0686 , N(b2p) = 13.7361, *p* = .0000). The omnibus test of multivariate normality, Small's test variant, also suggested violations of normality (χ^2= 372.5451, *df* = 28, *p* = .0000). Visual examination of the box and whisker plots revealed several pairs between the independent variables that were different suggesting lack of homogeneity of variances and co-variances. Most of the dependent variables did not meet the assumption of normality. Violations of multivariate normality have minimal effect on Type I errors or rejecting the null hypothesis when it is true (Hahs-Vaughn, in progress).

The linearity of the dependent variables were examined by matrix scatterplots of all pairs of dependent variables, All scatterplots revealed straight positive shapes, which suggested that the assumption of linearity was met.

The assumption for homogeneity of variances of the independent variables was met for most variables with Levene's test, except for the posttest for words before the main clause (*p* = .039), posttest for verb phrase density (*p* = .003), pretest and posttest for noun phrase density (*p*

= .006), and pretest and posttest for prepositional phrase density (*p* = .020), with all levels below alpha level of .05 Examination of boxplots of the dependent variables were all examined as a visual means to determine the extent to which equal variances can be assumed. Most of the boxplots for the dependent variables by group had varying box and whisker lengths, which also suggested violation of homogeneity of variances. Because this assumption can be not made with certainty, the omnibus test of Pillai's trace was used as it is more robust in MANOVA designs where homogeneity of variance-covariance is violated and less balanced (Hahs-Vaughn, in progress). Table 24 presents the data testing for hierarchical repeated measures MANOVA assumptions. Table 25 presents the data testing for the measures of univariate normality. Table 26 presents the data testing for the measures of homogeneity of variances for the dependent variables.

Table 24: *Results of Assumptions Testing for the Pretest, Posttest Science Expository Essay for Data Utilizing ITT Analysis (All Cases)*

Assumption	Test		Evidence	Assumptions Satisfied?
Independence	Matrix Scatterplots		No observable trends	Yes
Univariate Normality	Shapiro-Wilk		Not met for most variables(see Table 25)	No
	Boxplot/ Histogram/Q-Q Plot		Suggested nonnormality for some variables with outliers present	No
	Skewness	Standardized Residuals	Not met for all variables (all within value of absolute 2.0)	No
				No
		DeCarlo (1997)	Not met for all variables at p >.05	
	Kurtosis	Standardized Residuals	Not met for all variables	No
		DeCarlo (1997)	Not met for all variables at p >.05	No
	Scatterplot (Cook's vs. Unleveraged values)		Relatively normal shape for all variables	Yes
	Cook's Distance		All < 1.00	Yes
Multivariate	Skewness	Small's test	χ^2 = 239.7435, df = 14.0000, p = .0000	No
Normality		Srivastava's test	χ (b1p)=61.7129, df =14.000, p = .0000	No
DeCarlo (1997)	Kurtosis	Small's variant	χ^2=132.8015, df=14.0000 , p = .0000	No
		Srivastava's test	χ = 4.5024, N(b2p) = 13. 7361, p = .000	No
		Mardia's test	b2p = 299.0686 , N(b2p) = 13.7361, *p* = .00	No
	Omnibus	Small's variant	χ^2= 372.5451, df = 28.0000 p = .0000	No
Linearity	Matrix Scatterplots		Straight positive linear shapes	Yes
Homogeneity of variances-covariances	Levene's Test		Not met for all (see Table 26)	No
	Spread-vs-level plots		Most boxplots have varying box and whisker lengths	No

Table 25: *Shapiro-Wilk Test of Univariate Normality for Treatment Groups for the Science Expository Essay for Data Utilizing ITT Analysis (All Cases)*

Dependent Variable	Group	S-W	df	Significance	Assumption Satisfied?
PreCONN	Experimental	.889	26	.009*	No
	Comparison	.891	34	.003*	No
PostCONN	Experimental	.956	26	.319	Yes
	Comparison	.929	34	.029*	No
PreWBMC	Experimental	.969	26	.604	Yes
	Comparison	.490	34	.000*	No
PostWBMC	Experimental	.694	26	.000*	No
	Comparison	.834	34	.000*	No
PreAPV	Experimental	.637	26	.000*	No
	Comparison	.555	34	.000*	No
PostAPV	Experimental	.622	26	.000*	No
	Comparison	.532	34	.000*	No
PreVPD	Experimental	.839	26	.001*	No
	Comparison	.945	34	.085	Yes
PostVPD	Experimental	.855	26	.002*	No
	Comparison	.934	34	.042*	No
PreNPD	Experimental	.845	26	.001*	No
	Comparison	.851	34	.000*	No
PostNPD	Experimental	.786	26	.000*	No
	Comparison	.931	34	.033*	No
PrePPD	Experimental	.931	26	.082	Yes
	Comparison	.852	34	.000*	No
PostPPD	Experimental	.958	26	.363	Yes
	Comparison	.956	34	.187	Yes
PreCIWS	Experimental	.922	26	.049*	No
	Comparison	.879	34	.001*	No
PostCIWS	Experimental	.969	26	.596	Yes
	Comparison	.923	34	.020	Yes

Note. *Pre=pretest, Post=post-test, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences*

Table 26: *Levene's Test for Homogeneity of Variance for Dependent Variables for the Science Expository Essay for Data Utilizing ITT Analysis (All Cases)*

Variable	F	df1	df2	Sig.
PreCONN	2.413	4	55	.060
PostCONN	2.219	4	55	.079
PreWBMC	1.678	4	55	.168
PostWBMC	2.712	4	55	.039*
PreAPV	1.922	4	55	.120
PostAPV	1.970	4	55	.112
PreVPD	1.998	4	55	.108
PostVPD	4.655	4	55	.003*
PreNPD	8.414	4	55	.000*
PostNPD	4.235	4	55	.005*
PrePPD	3.215	4	55	.019*
PostPPD	4.559	4	55	.003*
PreCIWS	.653	4	55	.627
PostCIWS	.965	4	55	.434

Note. Pre=pretest, Post=post-test, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences

Hierarchical repeated measure MANOVA results

The results for the hierarchical repeated measures ANOVA suggest that there was not a statistically significant main effect (F_{test} = .245, *df* = 7, 49, *p* = .972) indicating that the combined dependent variables did not differ from pretest to posttest. Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a small effect and low power (partial η^2_{test} = .034, observed power = .113, as determined by Cohen (1988). Approximately 3% of the total variance of the combined dependent variables on the science expository can be accounted for by the within-subjects factor (i.e. time from pretest to posttest).

There was not a statistically significant interaction of treatment group and testing ($F_{treatment}$ = .998, *df* = 7, 49, *p* = .444). Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a medium effect and low power (partial $\eta^2_{treatment}$ =.125, observed power = .384), as determined by Cohen (1988). Approximately 13% of the total variance of the combined dependent variables on the science expository can be accounted for by treatment group. There was a medium effect, which indicated a medium proportion of difference in scores that occurred between the treatment groups. The experimental group marginal mean was higher than the comparison group for words before the main clause (experimental *M* =3.731, *SE*= .685, *CI* = 2.359 to 5.104; comparison *M* =3,525 *SE* =.622, *CI* =2.278 to 4.772); verb phrase density (experimental *M* =215.313, *SE* = 17.135, *CI* = 180.973 to 249.653); comparison *M* =210.501, *SE* = 15.565, *CI* = 179.308 to 241.684); and CIWS count (experimental *M*=52.794, *SE* = 8.346, *CI* = 36.069 to 69.519; comparison *M* =44.071, *SE* =7.581, *CI* =28.878

to 59.263) . Table 27 below shows marginal means, standard error, and confidence intervals for the dependent variables on the science expository essay for the intent-to-treat data.

Table 27: *Marginal Means for Dependent Variables by Treatment Group for the Science Expository Essay for Data Utilizing ITT Analysis (All Cases)*

				95% Confidence Interval	
Variable	Treatment group	Mean	Standard Error	Lower Bound	Upper Bound
CONN	Experimental	94.723[a]	8.207	78.276	111.169
	Comparison	100.114[a]	7.455	85.174	115.053
WBMC	Experimental	3.731[a]	.685	2.359	5.104
	Comparison	3.525[a]	.622	2.278	4.772
APV	Experimental	3.650[a]	1.735	.173	7.127
	Comparison	5.426[a]	1.576	2.267	8.584
VPD	Experimental	215.313[a]	17.135	180.973	249.653
	Comparison	210.501[a]	15.565	179.308	241.694
NPD	Experimental	315.775[a]	25.619	264.434	367.117
	Comparison	316.860[a]	23.271	270.224	363.497
PPD	Experimental	61.927[a]	8.588	44.715	79.138
	Comparison	69.574[a]	7.801	53.940	85.208
CIWS	Experimental	52.794[a]	8.346	36.069	69.519
	Comparison	44.071[a]	7.581	28.878	59.263

[a] *Based on modified population marginal mean. Note. CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences*

In addition, there was no statistically significant interaction for testing and the between-subjects factor of the nested classrooms by teacher within treatment group ($F_{classroom}$= .743 *df* = 21, 153, *p* = .783). Multivariate partial eta squared for the effect of the between-subjects factor of the nested classrooms by teacher within treatment group indicated a medium effect and moderate power (partial η^2_{test} =.092, observed power = .564), as determined by Cohen (1988).

Approximately 9% of the total variance of the combined dependent variables on the science compare and contrast writing prompt measures can be accounted for by the nested classrooms by teacher within treatment group. There was a medium effect, which indicated there was a medium proportion of difference in scores that occurred among the classes. The experimental class one (*M* =69.687, *SE* =10.352, *CI* = 48.943 to 90.432) had a higher marginal mean in a range 6-36 higher than the other classrooms on the correct versus incorrect word sequences (CIWS) count. In addition, experimental class one was the only class to increase of 4.375 points in mean score from pretest (*M* = 68.9286, *SD* = 51.18062) to posttest (*M* =73.9286, *SD* = 36.50809). Experimental class two maintained the same mean score of 35.9 from pretest to posttest. The three comparison classes decreased in score from pretest to posttest in a range of 6-11 points. See Table 28 for CIWS count pretest, posttest and marginal mean scores by classroom for the science expository essay for data utilizing ITT analysis.

Table 28: *Means, Standard Deviation/Error, and Confidence Intervals by Treatment Group for the Science Expository Essay for Correct Versus Incorrect Word Sequences Count for the Data Utilizing ITT Analysis (All Cases)*

					95% Confidence Interval	
Treatment Group		Score	Standard. Deviation	Standard Error	Lower Bound	Upper Bound
Experimental Class 1	Pretest Mean	67.5000	51.64365			
	Posttest Mean	71.8750	39.60114			
	Marginal Mean	69.687[a]		10.352	48.943	90.432
Experimental Class 2	Pretest Mean	35.9000	55.54268			
	Posttest Mean	35.9000	49.28252			
	Marginal Mean	35.900[a]		13.094	9.660	62.140
Comparison Class 1	Pretest Mean	40.2353	40.30281			
	Posttest Mean	29.5882	31.51400			
	Marginal Mean	33.800[a]		13.094	7.560	60.040
Comparison Class 2	Pretest Mean	38.9000	54.59436			
	Posttest Mean	28.7000	32.97491			
	Marginal Mean	37.556[a]		13.640	10.065	65.046
Comparison Class 3	Pretest Mean	66.4286	35.63171			
	Posttest Mean	60.5714	38.81519			
	Marginal Mean	63.500[a]		15.650	32.137	94.863

[a] *Based on modified population marginal mean.*

Morpho-syntactical analysis for both samples

Morpho-syntactical errors were calculated in addition to writing conventions and semantic errors as part of the dependent variable measure of the CIWS count. However, additional descriptive statistics were calculated via language sample analysis, to determine the type and incidence of morpho-syntactical errors that students exhibited on the science expository essay. The treatment-on-the treated-data sample was 49 students (21 experimental, 28 comparison). There were a total of 18 out of 49 students (37%) who exhibited no errors on both the pretest and posttest. There were 31 out 41 (63%) of students who exhibited morpho-syntactical errors. Out of the 31 students with errors, there were 15 experimental students (71% of this group) and 16 comparison students (57% of this group) who exhibited morpho-syntactical errors.

The ITT data sample was 60 students (26 experimental, 34 comparison). There were a total of 22 out of 60 (37%) of students who exhibited no errors on both the pretest and posttest. There were 38 out 52 (63%) of students who exhibited morpho-syntactical errors. Out of the 38 student with errors, there were 18 experimental students (69% of this group) and 20 comparison students (59% of this group) who exhibited morpho-syntactical errors.

According to Table 29, the most common errors were similar across both samples. The morpho-syntactical errors that occurred in descending order were (1) inconsistent verb tense, (2) regular past tense verb errors, (3) irregular past tense verb errors, (4) plural errors, (5) subject verb agreement singular form, (6) subject verb agreement plural form, past perfect tense verbs,

(7) articles, (8) infinitives, (9) prepositions, (10) omitting copula verbs, (11) present perfect tense verbs, and (12) possessives

Table 29: *Descriptive Statistics of the Type of Morpho-Syntactical Errors on the Science Expository Essay for the Treatment-on-the-Treated and ITT Analysis Data*

Morpho-Syntactical Error	Total Errors for the Treatment-on-the Treated Data Sample	Total Errors for the ITT Data Sample
Inconsistent verb tense errors	30%	26%
Regular Past Tense Verb errors	20%	22%
Irregular Past Tense Verb errors	13%	16%
Plural errors	13%	9%
Subject verb agreement singular errors	8%	7%
Subject verb agreement plural errors	6%	7%
Past perfect tense errors	4%	4%
Article errors	2%	3%
Infinitive errors	2%	3%
Preposition errors	1%	1%
Omitting copula errors	1%	1%
Present perfect tense errors	0	1%
Possessive errors	0	0

Summary of the Data for Research Question Three

Data analyses with all cases

In summary, both treatment-on-the treated data sample with data with ITT analysis yielded similar results. Both analyses did not indicate statistical significance ($p < .05$) of the difference of the combined dependent variables between the pretest and post-test scores on the science expository essay, with low to medium effect and low power. Both analyses indicated no statistically significant interaction of treatment group and test of the combined dependent variables on the science expository essay, with moderate to large effect and low power. There was a large effect and that the experimental group's marginal mean was higher than the comparison group for words before the main clause, verb phrase density, and CIWS count (experimental M =52.794 SE = 8.346, CI = 36.069 to 69.519; comparison M =44.071, SE =7.581, CI =28.878 to 59.263) . Both analyses indicated no statistically significant interaction of the nested classrooms by teacher within treatment group of the combined dependent variables on the science expository essay, with medium effect and moderate power. Both analyses also revealed medium effect and that the experimental class one had a higher marginal mean and increase from pretest to posttest on the CIWS count when compared to the four other classrooms. Analysis of morpho-syntactical errors revealed similar error patterns on both the treatment-on-treated and ITT analysis data samples.

One difference between the two analyses was noted. The treatment-on-the treated sample revealed higher marginal scores for the experimental group on the CIWS count than the comparison group. This was the same for the ITT analysis sample; however, on the ITT sample analysis, the experimental group also had higher marginal means on words before the main clause, and verb phrase density, than the comparison group. As the sample size increased with the ITT analysis, this may have increased the sample enough to reveal differences between the groups.

Data analyses for samples with outlier cases removed

Both the treatment-on-the-treated data sample and the ITT analysis sample had the presence of a significant number of outliers. Both samples revealed average means scores of zero on several dependent measures. The dependent variables for this question were measuring the presence of complex syntactic structures. If a student did not use the syntactic structure, their score was a zero. The sample population were students who struggle with literacy; therefore, it is logical that a distribution with presence of outliers would be evident. All samples failed to reject the null hypothesis, which was similar across all four analyses, thus the chance that Type I error (false positive) had occurred due to the distribution of the sample without outliers removed was unlikely. Two additional analyses were completed with removal of the outliers to make comparisons to the two analyses conducted without the outliers removed. Again, the results should be interpreted with caution, as these samples had much smaller samples sizes (treatment-on-the-treated n = 36; ITT analysis n = 44), which may have reduced power and effect. Refer to Appendix V for the assumptions testing and hierarchical repeated measures MANOVA results for the treatment-on-the-treated sample with outliers removed. Refer to Appendix W for the assumptions testing and hierarchical repeated measures MANOVA results for the sample data utilizing ITT analysis with outliers removed. Both the samples with outliers removed (treatment-on-the-treated and ITT analysis) yielded similar results. Both analyses did not indicate statistical significance ($p < .05$) of the difference of the combined dependent variables between the pretest and post-test scores on the science expository essay, with medium effect and low power. Both analyses indicated no statistically significant interaction of treatment group and test of the combined dependent variables on the science expository essay, with large effect and low power. Both analyses indicated no statistically significant interaction of the nested classrooms by teacher within treatment group of the combined dependent variables on the science expository essay, with medium effect and low to moderate power. Both analyses also revealed that the experimental class one had a higher score on the CIWS count when compared to the four other classrooms. Overall, these analyses were similar to the two analyses conducted with all cases.

One difference between the two analyses with outliers removed was noted. The treatment-on-the-treated sample revealed higher marginal means for the experimental group on agentless passive voice, and CIWS than the comparison group. This was the same for the ITT analysis sample; however, on the ITT sample analysis the experimental group also had higher marginal means on sentence connectives, words before the main clause, verb phrase density, and noun phrases density, than the comparison group. The trend that emerged across all four analyses was that as the sample became more normal when outliers were removed, as well as increased in size with ITT analysis, there were more dependent variables with higher marginal means in favor of the experimental group. For example, the treatment-on-the-treated data had one dependent variable in favor of the experimental group, which increased to two with the outliers removed, then increased to three variables with an increase in sample size with ITT data analysis, and then increased to six variables when the outliers were removed from the ITT analysis data.

In sum, as the sample increased and became more normally distributed, more areas were in favor of the experimental group. This same trend occurred with the CIWS count for the experimental class one when compared to other classes. The differences in marginal means increased for both treatment=on-the-treated data and ITT data after the outliers were removed and the sample was a more normal distribution. Table 30 summarizes the data trends in favor of the experimental group for all four analyses conducted on the data for the science expository essay.

Table 30: *Data Trends in Favor of the Experimental Group for Research Question Three*

Analysis		Sig.	Effect (partial η^2)	Power	Marginal Mean Experimental Group Scores > Marginal Mean Comparison Group Scores
Treatment - on-the Treated [All Cases] *n* =49 **E**=21 **C** =28	Pretest, Posttest	No	Small	Low	
	Treatment by group	No	Large	Low	CIWS higher for experimental group. Experimental increased, Comparison decreased.
	Treatment by class	No	Medium	Moderate	For CIWS, experimental class one 8-36 points higher than all other classes.
Treatment -on-the-Treated [Outliers Removed] *n* = 36 **E** =16 **C** =20	Pretest, Posttest	No	Medium	Low	
	Treatment by group	No	Large	Low	APV & CIWS higher for experimental group.
	Treatment by class	No	Small	Low	For CIWS, experimental class one 17-58 points higher than all other classes.
ITT [All Cases] *n* =60 **E** =26 **C** =34	Pretest, Posttest	No	Small	Low	
	Treatment by group	No	Medium	Low	WBMC, VPD, & CIWS higher for experimental group.
	Treatment by class	No	Medium	Moderate	For CIWS, experimental class one 16-36 points higher than all other classes.
ITT [Outliers Removed] *n* =44 **E** =17 **C** =27	Pretest, Posttest	No	Medium	Low	
	Treatment by group	No	Large	Low	CONN, WBMC, APV, VPD, NPD, & CIWS all higher for experimental group
	Treatment by class	No	Medium	Moderate	For CIWS, experimental class one 14-52 points higher than all other classes.

Note. ITT-Intent-to-Treat Analysis data, Sig.=Statistical significance, ***E****=experimental group,* ***C****=Comparison group; Dependent variables-CONN=sentence connectives APV=agentless passive voice, VPD=verb phrase density, NPD=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences*

Research Question Four

Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase their ability to determine similarities and differences (compare/contrast structure) related to science content as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

A hierarchical repeated measures ANOVA model was generated to answer this question. The independent variable was treatment group (experimental or comparison), the dependent variable was the pre-test and posttest scores, and the hierarchical factor was the classroom (i.e., teacher). The hierarchical repeated measures ANOVA was conducted to determine if the within-subjects factor, the pretest to posttest of the science compare and contrast double bubble map was significant; and if there was a significant interaction with the between subjects factors of treatment group (experimental or comparison), while accounting for the nesting of students within the classroom (i.e., teacher).

Baseline Equivalency Testing

Equivalency testing was conducted on the baseline pretest scores of the science compare and contrast double bubble map (dependent variable) between the two treatment groups (independent variable). Prior to running the independent *t*-test, the assumptions of the normality of the distribution were tested with the Shapiro-Wilk test and were met for the experimental group ($SW = .940$, $df = 31$, $p = .083$) with *p*-value greater than alpha level of .05, but not for the comparison group ($SW = .971$, $df = 31$, $p = .011$) with *p*-value less than .05. Skewness and kurtosis statistics (experimental skewness = .544, kurtosis = .311; comparison skewness = .172, kurtosis = -.622) were met, which were within an absolute value of 2.0. Visual examination of Q-Q plots, histograms (shape), and boxplots (presence of outliers) suggested normal distributions. Levene's test was completed to test the assumption of homogeneity of variances and was met with p = .703 which is greater than alpha level of .05. Last, the independent *t*-test was completed to determine significance (met with p < .05, no significant differences in scores), and small to minimal effect by Cohen's *d* (Cohen, 1988). The results of the independent *t*-test was $t(66)$ = .827, p = .411, which was not significant, and d = .20, which indicated a small effect. The results suggested the two treatment groups were relatively equivalent at baseline on the compare and contrast double bubble map.

Assumptions Testing Results

The assumptions of independence, normality, and homogeneity of variances and covariances were tested before running the hierarchical repeated measures ANOVA model. The assumption of independence was determined by reviewing the scatterplots of the standardized residuals of the pretest and posttest scores by treatment group. There was a relatively random display of difference scores above and below zero, with no cyclical pattern. This generally suggested evidence of independence.

The assumption of univariate normality was examined through several indices using residuals. The Shapiro-Wilk's test revealed normality for the pretest residual (experimental p = .087, comparison p = .288), and the posttest residual (experimental p = .534, comparison, p = .310) were both greater than an alpha level of .05, which suggested normality. All skewness and kurtosis values were within an absolute value of 2.0, which also suggested normality. More specifically, the skewness and kurtosis statistics included the following: pretest residual (skewness = .546, kurtosis = .319) and posttest residual (skewness = -.107, kurtosis =-.836) for the experimental group; and pretest residual (skewness = .222, kurtosis = -.457) and posttest residual (skewness = .305, kurtosis = -.496) for the comparison group. Visual examination of the histograms, boxplots, and Q-Q plots revealed evidence of a normal distribution. The sample size was n = 68 (experimental n = 31, comparison n = 37).

The assumption of homogeneity of variance was met with Levene's Test (pretest, p = .916; posttest, p = .194), which was above alpha level of .05. Box's Test (p = .812) was also above the alpha level of .05; therefore, the assumption of homogeneity of covariances matrices was also met. Table 31 presents the data testing for repeated measures ANOVA assumptions.

Table 31: *Results of Assumptions Testing for the Science Compare and Contrast Double Bubble Map*

Assumption	Test		Evidence	Assumption Satisfied?
Independence	Scatterplots		No observable trends	Yes
Normality	Shapiro-Wilk	Experimental		
		Pretest residual	*SW* = .941, *df* = 31, *p* =.087	Yes
		Posttest residual	*SW* = .971, *df* = 31, *p* = 534	No
		Comparison		
		Pretest residual	*SW* =.965, *df* = 37, *p* = .288	Yes
		Posttest residual	*SW* = .966, *df* = 3, *p* = .310	Yes
	Boxplot/Histogram /Q-Q Plot		Relatively normal distribution shape	Yes
	Skewness	Experimental		
		Pretest residual	.546	Yes
		Posttest residual	-.107	Yes
		Comparison		
		Pretest residual	.222	Yes
		Posttest residual	-.305	Yes
	Kurtosis	Experimental		
		Pretest residual	.319	Yes
		Posttest residual	-.836	Yes
		Comparison	-.457	Yes
		Pretest residual	-.496	Yes
		Posttest residual		
Homogeneity of variance and	Levene's Test	Pretest scores	*F* (4, 63) = .238, *p* = .916	Yes
		Posttest scores	*F* (4, 63) = 1.566, *p* = .194	Yes
covariance matrices	Box's Test		*p* = .821	Yes

Hierarchical Repeated Measures ANOVA results

The results for the hierarchical repeated measures ANOVA suggest that there was a significant main effect for the within-subjects factor of the difference between pretest and posttest (F_{test} = 20.114, *df* = 1, *p* = .000). Multivariate partial eta squared for the main effect indicated a large effect and strong power (partial η^2_{test} = .242, observed power = .993), as determined by Cohen (1988). Approximately 24% of the variance on the science compare and contrast double bubble map can be accounted for by the within-subjects factor (i.e., time from pretest to posttest).

There was also a significant interaction between time (i.e., within subjects factor) and the between-subjects factor of treatment group ($F_{treatment}$ = .4.057, *df* = 1, *p* = .048). Multivariate partial eta squared for the interaction of time and treatment group indicated a medium effect and moderate power (partial $\eta^2_{treatment}$ = .060, observed power = .509), as determined by Cohen (1988). 6% of the variance on the science compare and contrast double bubble map can be accounted for by treatment group (experimental or comparison). The marginal mean for the experimental group (*M* = 3.668, *SE* = .313, *CI* = 3.042 to 4.293) was higher than the comparison group (*M* = 3.508, *SE* = .290, *CI* = 2.928 to 4.088) on the science compare and contrast double bubble map. The experimental group increased .41 points from pretest (*M* = 2.81, *SD* = 1.905) to posttest (*M* = 3.22, *SD* =2.136). The comparison group increased .37 points from a pretest (*M* = 3.33, *SD* = 2.289), to posttest score (*M* = 3.70, *SD* =2.093). Table 32 presents the scores, standard error, and confidence intervals for both the experimental and comparison groups.

Table 32: *Means, Standard Deviation/Error, and Confidence Intervals by Treatment Group for the Science Compare and Contrast Double Bubble Map*

Treatment Group		Score	Standard. Deviation	Standard Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Experimental	Pretest Mean	2.81	1.905			
	Posttest Mean	4.55	2.234			
	Marginal Mean	3.668[a]		.313	3.042	4.293
Comparison	Pretest Mean	3.22	2.136			
	Posttest Mean	3.70	2.093			
	Marginal Mean	3.508[a]		.290	2.928	4.088

[a] *Based on modified population marginal mean*

In addition, there was a significant effect between time and nesting within classroom ($F_{classroom=}$ 4.397, *df* = 3, *p* = .007). Multivariate partial eta squared indicated large effect and strong power (partial $\eta^2_{classroom}$ = .173, observed power = .853), as determined by Cohen (1988). Approximately 17% of the variance on the science compare and contrast double bubble map can be accounted for by the hierarchical structure of the data (i.e., students nested within classrooms). There was a large effect; however, the univariate pairwise comparisons between classrooms by treatment group did not reveal statistically significant differences p >.05 between the pretest and posttest scores between any of the five classrooms. Therefore, a univariate ANOVA was generated for only the posttest scores on the science compare and contrast double bubble map, to determine any classes that were statistically significant different from the other. Pairwise comparisons revealed that there was a statistically significant difference between experimental class one and comparison class one on the posttest of the science compare and contrast double bubble map. The effect size was calculated by Cohen's *d* and was found to be *d* = 5.17, indicating a very large effect. Experimental class one had a pretest (*M* = 2.81, *SD* =1.682) and increased 2.31 points to a posttest score (*M* = 5.12, *SD* =1.821). In contrast comparison class one pretest score (*M* = 3.33, *SD* =2.289) had decreased .86 points to a posttest score (*M* = 2.47, *SD* =1.506). All other pairwise comparisons revealed that there were no other statistically significant (p >.05) differences on the posttest scores of the science compare and contrast double bubble map between any of the other classrooms. The other three classes made steady increases. Experimental class two had a 1.09 increase (pretest *M* = 3.58, *SD* = 1.929 to posttest M= 4.67, SD 2.188). Comparison class two had a 1.08 increase (pretest *M* = 2.60, *SD* = 2.221) to posttest *M* = 4.40, *SD* = 1.955). Comparison class three had a 1.09 increase (*M* = 3.58, *SD* = 1.929 to posttest *M* = 4.67, *SD* = 2.188). Table 33 presents the pretest means, posttest means, standard deviation/error, and confidence intervals by classroom (teacher) by treatment group for the science compare and contrast double bubble map. Table 34 outlines the univariate pairwise differences found between classes on the posttest of the science compare and contrast double bubble map.

Table 33: *Means, Standard Deviation/Error, and Confidence Intervals by Classroom for the Science Compare and Contrast Double Bubble Map*

Treatment Group		Score	Standard. Deviation	Standard Error	95% Confidence Interval Lower Bound	Upper Bound
Experimental Class 1	Pretest Mean	2.81	1.682			
	Posttest Mean	5.12	1.821			
	Marginal Mean	3.969[a]		.435	3.099	4.839
Experimental Class 2	Pretest Mean	2.80	2.178			
	Posttest Mean	3.93	2.520			
	Marginal Mean	3.367[a]		.450	2.468	4.265
Comparison Class 1	Pretest Mean	3.33	2.289			
	Posttest Mean	2.47	1.506			
	Marginal Mean	2.900[a]		.450	2.001	2.900[a]
Comparison Class 2	Pretest Mean	2.60	2.221			
	Posttest Mean	4.40	1.955			
	Marginal Mean	3.500[a]		.551	2.399	4.601
Comparison Class 3	Pretest Mean	3.58	1.929			
	Posttest Mean	4.67	2.188			
	Marginal Mean	4.125[a]		.503	3.120	5.130

[a] *Based on modified population marginal mean*

Table 34: *Univariate Pairwise Differences Found Between Classes on the Posttest of the Science Compare and Contrast Double Bubble Map*

(I) Classroom	(J) Classroom	Mean Difference (I-J)	Standard Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Experimental 1	Comparison 1	2.658[a]	.727	.005*	.545	4.772
	Comparison 2	.725[a]	.815	1.000	-1.646	3.096
	Experimental 2	1.192[a]	.727	1.000	-.922	3.305
	Comparison 3	.458[a]	.772	1.000	-1.788	2.704
Experimental 2	Comparison 1	1.467[a]	.738	.513	-.681	3.614
	Experimental 1	-1.192[a]	.727	1.000	-3.305	.922
	Comparison 2	-.467[a]	.825	1.000	-2.868	1.934
	Comparison 3	-.733[a]	.783	1.000	-3.011	1.545
Comparison 1	Experimental 1	-2.658[a]	.727	.005*	-4.772	-.545
	Comparison 2	-1.933[a]	.825	.223	-4.334	.468
	Experimental 2	-1.467[a]	.738	.513	-3.614	.681
	Comparison 3	-2.200[a]	.783	.066	-4.478	.078
Comparison 2	Comparison 1	1.933[a]	.825	.223	-.468	4.334
	Experimental 1	-.725[a]	.815	1.000	-3.096	1.646
	Experimental 2	.467[a]	.825	1.000	-1.934	2.868
	Comparison 3	-.267[a]	.866	1.000	-2.785	2.252
Comparison 3	Comparison 1	2.200[a]	.783	.066	-.078	4.478
	Experimental 1	-.458[a]	.772	1.000	-2.704	1.788
	Comparison 2	.267[a]	.866	1.000	-2.252	2.785
	Experimental 2	.733[a]	.783	1.000	-1.545	3.011

[a] *Based on estimated marginal means*

Editing

Starting the fourth day of the treatment, the experimental students were given time at the beginning of each class to edit their writing from the previous session. The editing was considered complete if the student had made the corrections specified by the researcher (i.e. circled or underlined portions or "uncoded feedback") by the next class period, or the following class period if additional time was needed to make the edits. According to Table 35, the students in first experimental classroom (*n* = 20), 19 students edited their work 100% of the time, and one student edited 50-75% of the time. The students in the second experimental classroom (*n* = 17), utilized the opportunity to edit as follows: eight students edited their work 100% of the time, two students edited their work 75-99% of the time, two students edited their work 50-74% of the time, three students edited their work 25% or less of the time, and two students edited their work 0% of the time.

Table 35: *Experimental Classes Use of Editing*

	Experimental Class 1 (*n* = 20)		Experimental Class 2 (*n* = 17)	
Percentage of Time Editing	*n*	%	*n*	%
100% of the time	19	95	8	47
75%-99% of the time	1	5	2	12
50-74% of the time	---	__	2	12
49-25% of the time	---	---	---	---
Less than 25% of the time	---	---	3	17
0% of the time	---	---	2	12

Teacher Writing Survey

The Experimental and Comparison Pre and Post-Surveys were conducted to gather demographic information related to the science teachers, student writing in science, and the science teacher's perception of writing in their eighth-grade science class. Results indicated that all the teachers reported that their students write about 10-15 minutes per day. The content of the student's writing was reported as data analysis, note taking, and creating a hypothesis or conclusion. The teachers reported that they also model this type of writing for the students. The teachers noted that when students are tested the writing requirement may be short answer question, answering a multiple choice question, or creating and labeling diagrams. The teachers each named different aspects of the writing process that are challenging for their students, such as conventions, vocabulary, spelling, or listening to directions. All the teachers expressed the importance of writing and its use in science; however, they all noted that the time devoted to work on writing in their science classes is limited. With regard to their background and training, the teachers have had minimal to no formal training on how to teach students to write. It should be noted that there were no differences in responses to all the questions across the pre and post-survey for all the teachers. Table 36 presents the results of the survey.

Table 36: *Pre and Post-Intervention Teacher Survey Results Related to Writing*

Question	Response	Percentage of teachers
How often do students write in your class?	10-15 minutes per day	100%
What type activities do students complete in your class that require writing?	data analysis	75%
	note taking	50%
	hypothesis and conclusions	50%
What type of writing do you model for your students?	labeling, note taking, and summarizing	50%
	writing hypothesis and drawing conclusions	50%
What kind of writing is required for test taking?	50% short answers	50%
	25% multiple choice	25%
	25% creating and labeling diagrams	25%
With regard to writing, what do you feel your students struggle with the most?	Grammar, conventions, and spelling	25%
	Listening to directions	25%
	Science vocabulary	25%
	Lack of detail and organization	25%
Do you feel that writing is important in science?	Yes	100%
Why?	The state standards require writing in science in high school	25%
	They must be able to make valid conclusions that are reported supported with evidence	75%
How many courses did take in teaching writing as part of obtaining your college degree?	None	50%
	One or more courses that incorporate writing techniques	50%
What professional development have you completed in writing outside of my degree?	None	0%
	Some professional development on how to teach writing	50%
	One course on writing in content areas	25%
I think writing is science is…	important to teach, but I never have the time	75%
	something that helps them learn science content	100%
	just as important as learning content	75%
	different from writing in other academic subjects	50%
	every teacher’s responsibility at my school	50%

Social Validity

To determine the social validity of the MSC intervention, surveys were given to the experimental teachers post-intervention and the experimental students at pre and post intervention. Overall, both the teachers and the students were satisfied with the intervention. The teacher and student measures are presented separately.

Experimental Teacher Post-Intervention Survey

The two experimental teachers completed a social validity survey that was created by the researcher. Both teachers stated that they thought the MSC intervention has a positive impact on writing in science. One teacher commented that it provided additional support in English basics that students often do not get. The second teacher felt that it helped the students to learn to formulate comparison and contrast sentences that they often need in science. Both teachers remarked that they felt that the students developed a better awareness of the grammatical constructions that may be found specifically in comparing and contrasting in science. One teacher noted that the students never used punctuation when writing in her class, and they now have become aware of how punctuation changes the meaning of what they have written. One teacher also felt that the grammatical aspect of the intervention had also impacted their oral responses. Both teachers expressed the opinion that although the MSC intervention was meaningful, the time that it took made curriculum pacing difficult. For example, one teacher stated "I think it is a good intervention, but unfortunately, I don't have a lot of time to spare as we have lots of content to cover." The other teacher had stated "The time devoted to metalinguistic sentence combining, while meaningful, made maintaining curriculum pacing difficult." Both teachers' overall impression was positive; they felt it was a well-structured intervention that reminded students of details such as punctuation and capitalization. Both teachers commented that they would consider using MSC in the future and recommend it to others. One teacher did note that although she felt it was a good intervention, due to time constraints, it may be more favorable in a class such as reading or writing class. The other teacher thought that it might be helpful to review MSC instruction at the beginning of the school year as a possible intervention tool.

Student Surveys

All students in the experimental classes (*n* = 36) completed the pre and post-intervention survey that was created by the researcher. The pre survey asked about their perception of writing. The post survey asked the same questions about their perception of writing, and then some specific questions about the metalinguistic sentence combining (MSC) intervention. The survey asked students to rate their answers on a Likert scale of 5 (strongly agree) to 1 (strongly disagree). There were also two open-ended questions related to the importance of writing and MSC intervention in the future. The survey was conducted anonymously, so that the students could feel comfortable to answer honestly. Overall, from pretest to post-test, the experimental students increased in their report of liking writing from 6.25% strongly agree to 14% strongly agree and with a decrease in strongly disagree from 31.25% to 9%. Students increased in thinking that writing is important in all classes from pretest 3% strongly agree to 23% strongly agree, and a decrease in strongly disagree from 6.25% to 3%. Students also increased in thinking that writing in science is important with an increase of strongly agree from 6.25% at pretest to 14% at posttest, and a decrease of strongly disagree from 12.5% at pretest to 0% at posttest. Specific to MSC intervention, more than 50% of students agreed or strongly agreed that MSC helped them to write in science class (63%), write better in all their classes (59%), and think more about their writing (57%). Thirty-six percent of students agreed or strongly agreed that MSC helped them read and learn science concepts better. Forty-seven percent remained neutral on that question. More than half of the students when asked why writing was important had stated that writing skills would be need for their future jobs. The second most popular answer was that writing skills are needed to be successful in high school and college. Some students also mentioned the importance of writing for specific educational and communicative tasks such as studying, remembering things, writing a letter, or completing paperwork. When asked specifically about how MSC impacted their writing for the future, the majority or students again mentioned the need to write better for high school, college or a future job. Other students spoke about how MSC would help their writing quality. Students mentioned different aspects of writing quality such as using clearer, shorter sentences, correct use of commas, as well as knowing how to combine sentences without using the word “and.” One student mentioned that it would help him/her to take better notes in science. Table 37 reports the specific ratings for pre and post-survey related to writing. Table 38 reports the specific rating for questions asked on the post survey related to the MSC intervention.

Table 37: *Pre and Post-Intervention Student Survey Results Related to Writing*

Statement		Strongly Agree 5	Agree 4	Neutral 3	Disagree 2	Strongly Disagree 1
I like to write	Pretest	6.25%	28%	25.25%	9.25%	31.25%
	Posttest	14%	26%	40%	11%	9%
I think writing is important in all my classes	Pretest	3%	32%	50%	6.25%	6.25%
	Posttest	23%	35%	33%	6%	3%
I think writing is important in science	Pretest	6.25%	41%	31.25%	9%	12.5%
	Posttest	14%	31%	40%	14%	0%

Table 38: *Post-Intervention Student Survey Results Related to MSC Intervention*

Statement	Strongly Agree 5	Agree 4	Neutral 3	Disagree 2	Strongly Disagree 1
I like that learning MSC helped my writing in science class	20%	43%	29%	9%	0%
I think learning MSC will help me write better in <u>all</u> my classes	16%	43%	34%	7%	0%
I think that learning MSC helped me to think more about my writing.	26%	31%	29%	14%	0%
I think MSC helped me to read and learn science concepts better	16%	20%	47%	9%	3%

Fidelity of Implementation

Two trained research assistants chose a random sample of intervention sessions in each instructional phase, as well one peer-supported, and one independent session. Each rater observed eight sessions (four for experimental class one and four for experimental class two). This was 20% of sessions for each experimental class, which is suggested to document implementation fidelity (Borrelli, 2012). For rater one, the average of fidelity determined of the four sessions for experimental class one was 96% and for experimental class two was 100%. For rater two, the average of fidelity of the four sessions for experimental class one was 98.5% and for experimental class two was 99%. Overall fidelity of implementation was 98%. See Table 39 for the fidelity of implementation results.

Table 39: *Fidelity of Implementation Results*

	Number of Classes Observed	Rater 1-Percentage of Key Elements Observed		Rater 2-Percentage of key Elements Observed	
Protocol Phase		Class 1	Class 2	Class 1	Class 2
Phase 1	4	100%	100%	100%	100%
Phase 2	4	96%	100%	100%	96%
Phase 3	4	100%	100%	94%	100%
Peer-Supported	2	88%			100%
Independent	2		100%	100%	
Average of All	16	96%	100%	98.5%	99%

The comparison classes were also video recorded and reviewed by the researcher. This insured that one, all experimental students received equitable treatment, and two, that there was no diffusion or the inadvertent application of MSC intervention to the comparison group students (Edmonds & Kennedy, 2013). The use of video recording also prevented the possibility of observer effects that can occur when there is an observer present in the room (Gall et al., 2006). Since all participating classes and teachers were on the same campus and one teacher was assigned with a class both in the experimental and comparison groups, it was necessary to insure there were no aspects of the MSC intervention in all the participating classes. Prior to the study, the researcher asked the two participating teachers not to discuss any aspect of the study with anyone or with each other. The researcher video recorded and reviewed 20% of the total possible instructional time that occurred over the seven weeks, or 237 minutes out of 1185 minutes, for the two experimental classes and three comparison classes. There was zero evidence of any aspects of MSC intervention in any of the five participating classes.

Inter-Rater Reliability

The reliability for the pre and posttesting of the TOWL-4 Sentence Combining Subtest was 94%. The reliability for the science compare and contrast writing prompt and science expository essay using Coh-Metrix 3.0 (i.e., typing in sample and then running the correct analyses) for pretest and posttest was 99%. The reliability for the CIWS sequences count was 97.5%, 100% for type of sentence connectives, and 98% for morpho-syntactical errors. The reliability for the pre and posttest compare and contrast science double bubble maps was 100%. Overall inter-rater reliability for all measures was between 94-100%.

Chapter Summary

In this chapter, the results of the study were presented. The results revealed that the intervention had a small effect between the experimental and comparison group on sentence-combining ability, with the experimental group having a greater increase in score from pretest to posttest, and slightly larger marginal means. On the science compare and contrast written prompt there was a small effect on the combined dependent variables that measured sentence complexity between groups, with larger marginal means in favor of the experimental group on a few variables. There was a large effect between classes on the science compare and contrast writing prompt. These results were in favor of experimental class one who had a higher marginal mean in sentence length and CIWS count, based on the ITT analysis of the science compare and contrast writing prompt. On the science expository essay, there was a trend across several analyses that revealed a moderate to large effect on the combined dependent variables that measured sentence complexity between groups, with higher marginal means in favor of the experimental group. There was a large effect between classes on the science expository essay. These results were in favor of experimental class one on the CIWS count. On the science compare and contrast double bubble map, there was a statistically significant difference between the experimental and comparison group, with medium effect, in favor of the experimental group.

CHAPTER FIVE: DISCUSSION

This chapter discusses conclusions with respect to each research question, as well as social validity. Limitations, implications for practice, recommendations for future research, and overall conclusions are also addressed. The purpose of this study was to increase metalinguistic awareness of science content to improve written sentence complexity in science, as well as the written expression and determination of comparison and contrast of science content, for eighth-grade students who struggle with literacy. The results of the intervention revealed a significant difference in favor of the experimental group in determining science comparison and contrast concepts. Results were also in favor of the experimental group in sentence combining, as well as written sentence complexity in response to a science compare and contrast written prompt and science expository essay.

Statistically significant differences were not achieved in all areas; however, there were a significant number of students who could not be included in the analyses of all research questions. With regard to the experimental group, there were 36 students assigned to this group. There were 10 students missing out of this 36 (28%) for the pretest. Because this was a nonrandomized sample, there were no data analyses that could be used to count those students, even if they were present for the posttesting. For the posttesting, there were five experimental students who were not present (another decrease of 14%), but could be counted with ITT analysis. ITT analysis allows the pretest score to be transferred and counted as the posttest score

(Torgerson & Torgerson, 2008). As noted in Chapter Four, a negative aspect of ITT analysis is that it can weaken any difference in treatment effect and lessen possible information that would yield the efficacy of the treatment (Armijo-Olivo, Warren, & Magee, 2009). In other words, the five experimental students who were counted with ITT analysis, did not reflect any possible gains that these students may have made due to the intervention. In sum, 15 students out of 36 students, or 42% of the experimental group, could not have data analyzed that may have determined if gains were achieved as a result of the intervention. This is close to half of the experimental group. Therefore, the results of research questions two and three may be conservative with regard to the differences between the treatment groups, or the magnitude of the intervention effect. Samples with missing or excluded data may not allow for true representation of the study population. Fortunately, there are three benefits of ITT analysis (Gupta, 2011), ITT analysis (a) increases the sample size, minimizes type I error and allows for greater generalization; (b) reveals important differences in scores between the treatment groups; (c) is considered a conservative measure that may not reveal score differences, but notable differences between the groups can be considered plausible.

Discussion of the Findings

Research Question One

Do eighth-grade students struggling with literacy who participate in metalinguistic sentence combining (MSC) instruction in science demonstrate an increase in their sentence-combining ability as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

The results of this question revealed a statistically significant difference for all students from pretest to posttest. The results revealed a small effect between the treatment groups or nested classes by treatment group on sentence combining. These findings did not achieve statistical significance and had low power. The experimental group achieved a slightly higher marginal mean than the comparison group, with a larger gain from pretest to posttest. Prior to the study, it was anticipated that the experimental group, after 400 minutes of instruction with sentence combining, would have made statistically significant gains over the comparison group on a standardized measure of sentence combining. These results do not mirror the results of the three most current sentence-combining (SC) research studies (Saddler, Asaro, et al., 2008; Saddler, Behforooz, et al., 2008; Saddler & Graham 2005), that have reported gains on this same standardized measure, the TOWL Sentence Combining Subtest (third instead of fourth edition). However, there are important differences between these three studies and this study, which may have influenced the difference in statistical significance of the test results.

First, the three previous research studies utilized instructional time that was greater than 400 minutes, which yielded 10-20 sentence combinations per 20-minute session. This study's protocol incorporated other components such as editing, responding to metalinguistic questions discussing answers, and modeling overt actions. Consequently, each intervention session only yielded two sentence combinations per 20-minute session. Saddler and Graham's (2005) study implemented an intervention protocol of 750 minutes, which is almost double the amount of practice with sentence combining than this study (400 minutes). The other two studies were specifically 630 minutes with only six participants (Saddler, Asaro, et. al, 2008) and 450 minutes with only four participants (Saddler, Behforooz, et. al, 2008). These conditions (i.e., smaller number of students and more time) may have allowed for more sentence-combining practice. All three studies were also conducted outside the classroom, which may have had fewer distractions than those that can occur in a full classroom of students. Therefore, it is reasonable to conclude then that the students in this study may have needed a larger dosage of intervention to achieve significant gains.

Second, in all three of the most current SC research studies (Saddler, Asaro, et. al, 2008; Saddler, Behforooz, et al., 2008; Saddler & Graham, 2005) either a group of students or all the participants had pretest scores on the TOWL-3 Sentence Combining Subtest that were more than two standard deviations below the mean. For Saddler and Graham (2005), this was a prerequisite for students placed in a "less skilled writer" subgroup that was nested within the two treatment groups. In the other two studies, the study participants were all students whose average pretest scores on the TOWL-3 Sentence Combining Subtest were greater than one standard deviation below the mean (less than a standard score of 7). This study had an average pretest score on the TOWL-4 Sentence Combining Subtest for both treatment groups that was within one standard deviation from the mean (experimental group average pretest mean score = 8.03, comparison group average pretest score = 7.86). Other researchers have documented that SC instruction has yielded greater gains in scores for students whose scores are lowest prior to treatment than students with higher scores (Evans et al., 1988; Scott & Nelson, 2009).

Finally, the three most current SC research studies (Saddler, Asaro et al., 2008; Saddler, Behforooz, et al., 2008, Saddler & Graham, 2005) replicated the same intervention protocol (developed by Saddler & Graham, 2005), under different conditions. The Saddler and Graham protocol covered five different units that incorporated the use of (a) contrastive and clausal conjunctions to form complex sentences, (b) embedding of adjectives and adverbs, (c) embedding adjectival clauses, (d) embedding adverbial clauses, and (e) combining sentences with multiple embedding. The TOWL-3 or TOWL-4 Sentence Combining Subtest assesses a myriad of different syntactic structures, which may have been better aligned with the syntactic skills targeted in the Saddler and Graham (2005) intervention protocol, that was also replicated in the other two research studies (Saddler, Asaro, et al. 2008; Saddler, Behforooz, et al., 2008). This study only covered the use of compare and contrastive conjunctions and extension or embedding of relative or adverbial clauses, using only science expository information. The TOWL-4 Sentence Combining Subtest is a broader measure of sentence-combining skills. The experimental students' scores from this study may not reflect gains they made in their sentence-combining ability, which was in a narrower syntactic context and used denser expository text. In addition, SC research has documented standardized measures (such as the TOWL-4) do not always capture gains from SC intervention, which can be more evident in formative measures (Fusaro, 1993; Straw & Steiner, 1982). Moreover, Neville and Searls (1988), after conducting a meta-analysis of SC studies, documented that standardized measures may contain sentences that are written at the simpler level of syntactic complexity than in a SC intervention. Thus, student performance on the standardized measure will not reflect the effect of the SC intervention. For this study, the results of this research question further supports the speculations of previous SC researchers (Fusaro, 1993; Neville & Searls, 1988; Straw & Steiner, 1982), that a standardized measure may not capture the effectiveness of the specific type of SC skills targeted in the intervention.

Furthermore, the marginal means of the experimental students were slightly higher than the marginal means of the comparison students. This indicated that the intervention was effective in yielding small gains in general sentence-combining skills. There was not a statistically significant difference between the treatment groups. However, this may have been due to this study's (a) shorter intervention protocol used with a whole class of students; (b) study participants that were higher performing at baseline (not less than one standard deviation below the mean at pretest); and (c) an intervention protocol that was narrower in scope, but used more complex disciplinary syntactic targets. These aspects of this study may have thwarted the possibility of statistically significant gains on the TOWL-4 Sentence Combining Subtest than in prior SC research studies. In addition, the TOWL-4 Sentence Combining Subtest, a standardized general measure, may not have been able to detect the true gains of SC ability after an intervention with more complex discipline specific syntactic structures.

Research Question Two

Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase in specific aspects of sentence complexity in response to a science compare/contrast writing prompt as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

This research question was tested utilizing two analyses, one with the treatment-on-the-treated data (n = 49) and one with intent-to-treat (ITT) analysis data (n = 60). The results of this question revealed a statistically significant difference for all students from pretest to posttest. The results revealed a small effect and low power on the combined dependent measures between the treatment groups, with marginal mean scores in favor of the experimental group on a few variables. There was a large effect and moderate power detected on both analyses between nested classes. On both analyses, this difference was reflecting a difference between just two classes. The researcher had anticipated at least a moderate to large effect size between the experimental and comparison groups, as well as possible statistical significance. However, the researcher could not anticipate the significant number of students who were not present for the pretesting and posttesting of the writing assessments. A total of 15 students out of 36 students, or 42% of the experimental group's data could not be included. The excluded data may have better determined if significant gains were made on the science compare and contrast writing prompt as a the intervention. The excluded data were close to half of the experimental group; therefore, the results of this question may be conservative with regard to the differences between the treatment groups. The data analyses to test differences between the treatment groups were both low in power indicating that the smaller sample sizes may not have revealed true differences in the data.

Descriptive analysis of the experimental group's pretest and posttest results of the science compare and contrast writing prompt revealed some important differences in their written responses from the comparison group. First, there was a considerable number of students who used the four targeted conjunctions from the intervention. For similarities, the experimental students used the conjunctions "both" and "like." For differences, the experimental students used the conjunctions but and however. Second, qualitative examination of the experimental group's posttest writing prompt, when compared to pretest writing prompt, revealed that the experimental students had longer complete sentences with embedded or added clauses, as well as the use of agentless passive voice. Last, the qualitative examination of the posttest writing prompts when compared to the pretest writing prompts, revealed that the experimental students had increased their use of writing conventions such as correct capitalization, comma use, and period ending. These differences were all specific targets of the intervention.

The difference in scores between the experimental and comparison group indicated a small effect with findings that did not achieve statistical significance. Again, there was observed low power for both analyses, which indicated that all differences may not have been revealed. More importantly, the marginal mean differences in the measured areas of sentence complexity were similar to the type of differences in sentence complexity noted after qualitative examination of the pretest and posttest writing prompts. For example, the results of the treatment-on-the treated data revealed higher marginal scores for the experimental group for targeted connectives(TCONN), agentless passive voice (APV), and the correct versus incorrect word sequences count (CIWS). The targeted connectives score looked specifically at the use of the four connectives or conjunctions targeted in the intervention, which was higher for the experimental group. After the ITT analysis, these three areas (TCONN, APV, CIWS) again revealed higher marginal means, with two of these dependent measures exhibiting notable increases. The APV for the experimental group score was almost two times greater with the ITT data (experimental *M* = 6.139, comparison *M* = 3.679). The increases in CIWS count should be interpreted with caution, for the experimental group did have a larger pretest mean with a medium effect. However, the experimental group did maintain a higher marginal mean on CIWS after posttest (experimental *M* =28.639 and comparison *M* =19.376). Therefore, the higher marginal mean in CIWS count may indicate that writing conventions, as well as phonologic, semantic, and morpho-syntactic aspects of the science compare and contrast writing prompts were better for the experimental group than the comparison group, after the intervention.

Unlike the treatment-on-the-treated analysis, the ITT analysis also revealed higher marginal means for the experimental group for sentence length (SL) and words before the main clause (WBMC). Both of these measures indicate an increase in sentence productivity (Nippold, 2010), or the use of sentences that contain more information. In addition, WBMC or what is also referred to as "left embeddedness", indicated the use of more complex syntactical forms (McNamara et al., 2014). Thus, the two measures indicate increases in sentence productivity and complexity in favor of the experimental group over the comparison group.

Overall, the mean differences increased in favor of the experimental group for five out of six of the dependent measures. With an increase in sample size from the treatment-on-treatment data to ITT analysis, it is plausible to assume that if the sample size continued to increase towards its original size, then the data may have increased in a direction of even higher marginal mean differences and statistical significance between the experimental and comparison groups. The analysis between the two groups also revealed low power, which also supports the possibility that the size of the sample may not have been large enough to detect all possible differences between the groups.

The trends in marginal mean differences revealed two important things. First, the experimental group had higher marginal means that indicate they were writing longer and more complex sentence forms when writing in a different writing context, which was a writing prompt and not sentence combining. Second, the longer more complex sentences contained syntactic forms. This finding is critical. The researcher did not incorporate any probes into the intervention to practice writing comparison and contrast sentences in response to a writing prompt. The use of the targeted conjunctions, sentences with embedded or added clauses, attention to writing conventions, and passive voice were targeted and practiced only within the parameters of the MSC exercises. Although it was routinely discussed that the purpose of the intervention was to compare and contrast science information, there were no intervention exercises tailored to match the type of general writing prompt administered for the testing, which was "Tell two similarities and two differences between…" In addition, the researcher was not present during the posttesting, so her presence could not bias the results of the experimental students by associating the test writing prompt with the intervention. All the students were only reminded to use "correct and complete sentences" for both the pretest and posttest science compare and contrast written prompt, with no further hints. Therefore, it is hypothesized that the experimental students were able to generalize the written semantic, morpho-syntactic, and conventional aspects of the intervention for the intended semantic purpose, which was to compare and contrast science information in writing. Past research has noted that research studies frequently do not include measurement of writing skills across contexts (Gersten & Baker, 2001); studies that had included measurement across contexts had mixed results. The results of this study provide data that suggest positive trends toward the use of morpho-syntactic structures and writing conventions in a different writing context (i.e., prompt vs. MSC exercise), without specific instruction in generalization to other contexts or genres. Hence, this study adds further support to the research literature that has indicated positive results in writing that is instructed in one writing context that resulted in gains in another writing context.

Moreover, the generalization of these skills from the intervention, to the science writing prompt indicated generalization of syntactic structures inherent in the discipline of science. Past research has recommended the use of academic content information to improve writing in that academic area (Graham & Herbert, 2010, 2011). Writing intervention should also focus on the syntactical structures that are specific to an academic discipline (Schleppegrell, 2007; Scott & Balthazar, 2010, 2013). First, the increase in scores for the experimental group on agentless passive voice (APV) and the targeted connectives (TCONN) indicate an increase the syntactic structures that are inherent in science text. Agentless passive voice is used to establish an authoritative tone (Fang, 2005; Fang et al., 2006). Science text uses conjunctions to signal comparative relations (Schleppegrell, 2007; Troia, 2009). Second, the increase in sentence length (SL) and words before the main clause (WBMC), suggest increases in sentence productivity and complexity that are specific to the discipline of science. Science text is characteristic of longer, dense information (Schleppegrell, 2007). In science, there are longer sentences that may contain left embedded clauses or several words before the main clause (Fang, 2005, Scott & Balthazar, 2010). Overall, the results of the experimental group's performance on the science compare and contrast writing prompt supports the hypothesis that writing instruction that focuses on syntactical structures inherent of an academic discipline, can generalize to students' independent writing of that academic discipline.

A feasible explanation for the experimental group's achieving higher marginal means, but not scores that were statistically different from the comparison group, may have been related to the length of the intervention and the use of expository text. This study protocol may not have been a long enough amount of time to reveal the effectiveness of sentence-combining intervention with academic expository text. The most current sentence-combining (SC) studies (Saddler, Asaro, et al., 2008; Saddler, Behforooz, et al., 2008; Saddler & Graham 2005), yielded statistically significant gains in writing after sentence-combining instruction with time frames of 450 to 750 minutes, but were conducted utilizing narrative text. Researchers have noted that comprehending and using expository text is more challenging for students than narrative text (Nippold, 2010; Shanahan et al., 2012). Therefore, a reasonable conclusion is that the length of time for this study was not sufficient for writing gains with academic expository text.

Unfortunately, there are no SC studies published where expository academic content has been used that could further support this conclusion. One SC study, conducted by Neville and Searls (1985) used social studies text, but only assessed the students writing in a cloze (i.e., fill in the blank) format that is not comparable to the full sentences required for the science compare and contrast writing prompt for this study. A study conducted by Cervetti et al. (2012) used an expository science writing prompt as a pretest, posttest measures to assess improvement in science writing with adolescents. The intervention did not implement SC, but an integrated science and literacy approach. The results indicated statistically significant differences in favor of the experimental group over the comparison groups ($p < .05$) when writing a science written prompt of four to five sentences. Most importantly, that study was conducted over an entire school year (10 months). The results of the study conducted by Cervetti et al. (2012) suggest a considerably longer time period for students to make noticeable gains in writing a prompt using science expository text. This reinforces the conclusion that an intervention for only 400 minutes, may not have allowed enough time for the experimental students in this study to make statistically significant gains over the comparison group when writing prompts with expository science text.

In summary, the results of the science compare and contrast writing prompt indicated that the experimental group had higher marginal means on two dependent measures of sentence complexity that then increased to five areas, when the sample size increased through the use of ITT analysis. This trend of increase in marginal mean may have continued to increase with a larger sample size, which yielded statistically significant gains in these areas. These increases indicate a promising trend that justifies reasonable speculation that the intervention was effective in increasing these areas of sentence complexity. Even more promising is that this trend suggested generalization of sentence complexity and the use of syntactical structures of the discipline of science when writing in a different writing context (i.e., writing prompt) than what was targeted in the intervention (i.e. MSC exercises), as a result of the intervention. Another reasonable conclusion is that statistically significant differences would have been achieved if this study had a longer intervention protocol that allowed for enough time for gains in complex expository syntactic structures to be evident on testing measures.

Research Question Three

Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase in specific aspects of sentence complexity when writing a science expository essay as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

This research question was tested utilizing four different analyses, treatment-on-the-treated data, all cases (n = 49), treatment-on-the-treated data, outlier cases removed (n = 36), and intent-to-treat (ITT) analysis data, all cases (*n* = 60), and intent-to-treat (ITT) analysis data with outlier cases removed (*n* = 60). For all four analyses, the results revealed no statistically significant differences across all the dependent measures, from pretest to posttest, between the treatment groups, or among nested classes by treatment group. However, the analyses between groups revealed a medium to large effect between groups with low power, and a medium effect and moderate power for the nested classes. The difference in nested classes was in favor of experimental class one and was only between this class and comparison class one for the CIWS count. The researcher anticipated that the intervention would be in favor of the experimental group on at least some of the measures. The intervention did not focus on this type of writing; therefore, this hypothesis was based on prior research, which has indicated that writing practice in one writing context can yield gains when writing in a different writing context. As with the science compare and contrast written prompt, a total of 15 students out of 36 students, or 42% of the experimental group could not have data analyzed that may have determined if gains were made on the science expository essay. However, four analyses were utilized to reveal possible trends to render the most accurate conclusions regarding the data. For the treatment-on-the-treated (all cases) sample that was increased with ITT analysis (all cases), the use of ITT analysis may be more conservative, but allow for reduction on type I error from smaller samples and generalization of results (Gupta, 2011). The dependent measures used for this question were measures of complex syntactic structures. If the students do not use that complex syntactic structure, then the score is a zero. Thus, non-zero scores are an outlier. Considering the student participants were all students who struggle with literacy, this result was not surprising. This may be the legitimate data distribution that would be expected with this population. However, the presence of outliers still threatens error rates. Therefore, the two samples were also run with the outliers removed. Samples with the outlier cases removed are beneficial for they may have represented a closer to normal distribution of this population. This process reduced the size of the sample significantly, which could have reduced power and effect. It should be noted on all analyses, even with outliers removed, that there was medium to large effect sizes between groups. Effect sizes indicate differences independent of sample size. In contrast, power was affected in that observed power was low on all analyses and may not have detected all differences between the groups. This would be expected with smaller samples, as well as lack statistical significance. Overall, the trends in marginal mean data, with moderate to large effect, indicated notable differences between groups. These trends revealed crucial information about the effectiveness of the intervention on the science expository essay.

The treatment-on-the-treated analyses revealed higher marginal scores for the experimental group on correct versus incorrect word sequences (CIWS) than the comparison group. The analysis the treatment on the treated data with outliers removed revealed a higher marginal mean for CIWS, as well as agentless passive voice (APV). The larger samples revealed more areas where the experimental group had higher scores. On the ITT with all cases, the experimental group had higher marginal means on CIWS, as well as words before the main clause (WBMC), verb phrase density (VPD), than the comparison group. The ITT analysis with the outliers removed revealed the same results (i.e., experimental higher marginal mean on WBMC, VPD, CIWS), in addition to agentless passive voice (APV) and sentence connectives

(CONN). No analysis revealed higher scores for the preposition phrase density in favor of the experimental group. Overall, all the analyses revealed a higher score for the experimental group on the CIWS count than the comparison group. At least two analyses revealed that scores on the dependent measures of WBMC, VPD, APV had a higher marginal mean for the experimental group than the comparison group.

The consistent higher marginal mean for the CIWS count indicated that overall writing conventions as well as phonological, semantic, and morpho-syntactic aspects of their writing had improved more for the experimental group than the comparison group. The experimental group had participated in the intervention for 20 sessions, which called attention to the metalinguistic aspects of writing conventions, as well as subject-verb agreement. The improvement of these aspects of writing would most likely be apparent in the CIWS count. In addition, the experimental group participated in 15 sessions where they were able to edit their work. This may have helped the experimental students to focus their attention on the conventional, phonologic, and morpho-syntactic errors more than before the intervention, which consequently increased their score on the essay.

Furthermore, based on all the analyses, experimental class one achieved a marginal mean that was from eight points to 52 points higher than the other classes. As noted in Chapter Four, 95% of students in experimental class one utilized the opportunity to edit their work 100% of the time. In the experimental class two, less than half utilized the opportunity to edit 100% of the time. Moreover, 29% of experimental class two utilized the opportunity to edit only between 0-

25% of the time. It is possible that the difference between the CIWS count of the two experimental classes may be due to the fact that the students who had increased practice in editing made more gains in the overall correctness of their writing. A meta-analysis conducted by Bangert-Drowns et al. (2004) had revealed that students who wrote informational content-related text and received feedback though an editing or revising process had a small weighted effect on performance (*d* = .20) than the minimal effect for those who did not participate in this editing/revision process with feedback (weighted effect *d* = .10) . The results of the current study further support the possibility that students who participate in editing and revision processes with feedback achieve better gains in content area writing than students who do not.

In addition, like the science compare and contrast written prompt, the science expository essay data revealed higher mean scores not only on CIWS, but on agentless passive voice (APV) words before the main clause (WBMC), verb phrase density (VPD), than the comparison group on two of the analyses. The increased marginal mean of APV, WBMC and VPD suggest increased sentence writing complexity characteristic of syntactic structures in the discipline of science.

Another important point with regard to the scores on the expository essay is related to generalization. The experimental group had higher scores in relation to sentence correctness and syntactic complexity (APV, WBMC, VPD) on the science expository essay. The intervention used expository information in the MSC exercises, but the students did not write beyond two sentences per session for the purpose of comparing and contrasting. The expository essay was an open-ended topic in which the students had to write about their favorite science topic or most exciting science topic or lab, that they had completed this school year in science. The positive findings indicated that the experimental students, more than the comparison students, had generalized the use of correct and more complex syntactic structures, without any explicit instruction of writing for this purpose or length. A possible interpretation of these findings is that practice in writing in one context within an academic discipline (i.e., compare/contrast in science) can yield gains in writing in other contexts within the same academic discipline (i.e., science expository essay). This interpretation would be important to test in future research with students who struggle, because complex syntactic structures are expected to emerge during adolescence in longer essay writing (Nippold, 1993, 2010; Scott & Stokes, 1995). The generalization of these skills into a new context in the discipline with greater demands in length supports the view that any writing practice can yield gains in writing, even if practiced at the sentence level (Scott, 2009).

There was a moderate to large effect in the difference in scores on the science expository essay. However, the experimental group did not achieve statistically significant gains over the comparison group, which could have been due to the lower power of the sample. Like the science comparison and contrast writing prompt, another reasonable deduction is that the lack of statistical significance may also have been due to limitations imposed by the length of the intervention and the use of academic expository text. In other words, expository text is more complex text and may require more time to achieve a large enough difference in score to achieve statistical significance. One older SC study conducted by O'Hare 1973, utilized SC intervention with seventh graders to target specific syntactic constructions (i.e., words per clause, number of noun, adverb, adjective, clauses per terminal unit) in narrative essays. The results revealed statistically significant gains on all the measures ($p < .001$) for the experimental over the comparison group. There are two important differences between O'Hare's study and this study. First, O'Hare's study did not use expository text, but narrative text, which may not be as syntactically complex as expository text. Second, O'Hare's (1973) study was for a period of eight months, which is six months longer than the current study.

There were two comparable studies conducted with middle school students (Keys et al., 1999; Hand et al., 2004) writing longer essay length science expository text. The researchers implemented the use of the intervention "Science Writing Heuristic" and not SC. The experimental groups demonstrated gains in science essay writing. For both studies, the total intervention time was 40 sessions for 40 minutes, or 1600 minutes. This is four times the length of this study (400 minutes). The fact that statistical significance was achieved after an intervention with a considerably longer dosage, further supports the reasonable speculation that statistical significance between scores may have been evident after a longer dosage with this study's MSC intervention.

In summary, the results of research question three indicated that the experimental group had higher mean scores on CIWS on all four analyses and at least two of the analyses revealed higher marginal means for three other dependent measures (APV, WBMC, VPD). The results of the expository essay, like the results of the science compare and contrast prompt, indicated a promising trend suggesting generalization to another writing context of correct sentences with increased complexity and syntactic characteristics inherent in the discipline of science. The results for this question will guide future research in sentence combining by providing a rationale for using complex expository text and discipline-specific syntactic structure targets.

Morpho-Syntactical Errors

Morpho-syntactical error analysis was conducted on both the science compare and contrast writing prompt and science expository essay. The writing prompt revealed the most morpho-syntactical errors on subject-verb agreement plural and singular form (69.5%). This may be the result of written language of increased complexity and density that is evident in science text. Science texts utilizes forms such as embedded clauses where there is significant distance between a noun and verb or subordinate clauses in which the subject may be omitted (Scott & Balthazar, 2010) and thus may result in errors in noun and verb agreement. For the science expository essay, the results were different. First, the average number of students who exhibited errors increased. On the science compare and contrast writing prompt, the average across the two analyses was 44.5% of students had morpho-syntactical errors. On the science expository essay, there was increase in average to 57.5%. This is not a surprise, as it has been noted in past research that as writing length increases, the possibility of errors increases (Nelson, 2013b; Singer & Bashir, 2004). However, the error patterns were not the same. On the essay, the most common errors were in verb tense (average total of 63.5% for all types of tense errors). These errors specifically were inconsistent verb tense, regular past tense verb errors, and irregular past tense verb errors. As the length of content increased, the students would have to remember the tense they used initially and keep it consistent. Verb tense errors are common for students who struggle with written syntax and become fatigued during longer writing tasks such as an expository essay (Scott & Balthazar, 2013). Overall, the results of the morpho-syntactical analyses revealed that as the length of the writing sample increased, the number of errors increased and the type of errors changed. In addition, it revealed errors in subject-verb agreement and verb tensing, which are areas that have already been documented in past research as common morpho-syntactical errors for students who struggle with literacy. Thus, these data lend further evidence that subject-verb agreement and verb tensing are important focal areas to address when remediating written morpho-syntax for students who struggle with literacy.

Research Question Four

Do eighth-grade students struggling with literacy who participate in MSC instruction in science demonstrate an increase their ability to determine similarities and differences (compare/contrast structure) related to science content as compared to eighth-grade students struggling with literacy who participate in typical science instruction alone?

The results for the hierarchical repeated measures ANOVA suggest that there was a significant main effect for the within-subjects factor of the difference between pretest and posttest. There was also a significant interaction between time and treatment group (p =. 048) with a medium effect and moderate power. The experimental group achieved higher mean difference in score (M = 3.668) than the comparison group (M = 3.508). In addition, there was a significant effect between time and nested classrooms by treatment group. There was a statistical difference (p = .01) between experimental classroom one (M = 3.969) and comparison class one M = 2.900), with a large effect (d >.80) and strong observed power.

The result of statistical significance between the treatment groups and a higher score for the experimental group was the expected result. Researchers have suggested that writing in a content area can yield gains in content understanding (Pearson et al. 2010; Yore et al., 2003). Meta-analyses by Bangert-Drowns et al. (2004) indicated small effect sizes on writing and science concept knowledge (weighted d = .032), as well as Graham and Perin (2007a, 2007b) for writing and general content area learning (d = 0.23). A meta-analysis conducted by Graham and Herbert (2010) also documented the relationship between writing after reading academic text and gains in reading comprehension of the academic text. The practices Graham and Herbert recommended yielded overall effect sizes from the small to moderate range. Practices included (a) to write written summaries (weighted effect size d = .052), (b) to write written answers (weighted effect size d = .047), and (c) to write written responses (weighted effect size d =.77).

Swanson and Deshler (2003) have stated that explicit instruction over time can help students who struggle with learning to retain and enhance various types of knowledge. MSC instruction provided not only additional writing practice with science text related to comparing and contrasting, but also increased encounters with the comparison and contrast of science content through reading, listening, and speaking modes. The specific metalinguistic cues may have called attention to linguistic aspects of comparison and contrast in science text that allowed the students develop better metalinguistic awareness and skills for comparison and contrast. The double bubble map or any other graphic organizer for the concept of comparing or contrast was not used during the intervention. The gains in favor of the experimental group suggested that the metalinguistic experience with comparison and contrast of science content better enabled them to determine, organize, and use language that clearly relayed the similarities and differences.

There was statistical significance between two of the classrooms in score, specifically, experimental class one and comparison class one. All the other classes were not statistically significant from experimental class one or comparison two. Therefore, it is a reasonable assumption that the significance detected was due to a large difference in score between only those two classes. The experimental class one had increased in score and comparison class two had decreased in score. In addition, the significant difference that did occur is in favor of an experimental class, which received the intervention.

In conclusion, the experimental group had statistically significant higher scores on the compare and contrast double bubble map with medium effect and moderate power. There was a difference between experimental class one and comparison class one. No other classes had statistically significant differences among each other; more importantly, the difference was in favor of the experimental group. Therefore, it can be reasonably concluded that the MSC intervention's effectiveness transferred to performance gains on determining comparison and contrast of science content.

Fidelity of Implementation

The researcher adhered to the intervention protocol with high fidelity (98% of randomly selected sessions). Qualitative notes recorded by both research assistants noted that the 2% of the time where fidelity was not achieved the researcher had stopped the sequence of the protocol to help students maintain attention and motivation. Students who struggle with literacy often lack motivation to participate in literacy-based tasks, particularly writing (Dockrell, 2014; Graham & Harris, 2013; Singer & Bashir, 2004; Troia, 2013). Therefore, this behavior was not a surprise.

Social Validity

This area was explored through the use of both a teacher and a student survey. Both participating teachers for the experimental groups noted that the MSC intervention was valuable and targeted necessary writing skills, particularly in grammatical constructions needed for comparison and contrast exposition in science. One teacher confirmed social importance of the intervention in that it resulted in the students having a better understanding of writing conventions that they need to use in and out of the classroom. The other teacher confirmed social importance on an oral level, stating that the intervention also appeared to improve the students' oral responses. One area that was not evident on the survey concerning social validity was that the teachers did not appear to value the intervention as a tool for increasing science content knowledge. Both teachers were concerned that the time lost for the intervention, which made it difficult to cover the necessary curricular content. Despite the gains (i.e., grammar, oral language) the teachers reported, there was not a shift of perception for either teacher that maybe the intervention resulted in student gains in science content, hence a worthy substitute for curricular content instruction (i.e., lecture, labs). Their concerns about covering academic content mirror the concerns that researchers have documented through teacher survey in both middle (Graham et al., 2014a) and high school (Gillespie et al., 2014). Although teacher's report that they value the importance of writing and its benefits, the pressure to cover academic content often hinders their willingness to try evidence-based practices that could lead to further student gains. It should be noted that the surveys were administered immediately after the study and the teachers were not aware of the results. Therefore, there is a possibility that there will be a shift in perception once the teachers are informed of the results, particularly the gains in science content learning.

The student survey results were positive with regard to the students demonstrating understanding of the social importance and relevance of writing in and out of school. A large number of the experimental students indicated positive reactions to the MSC intervention and felt that the intervention directly correlated with writing in science, but also skills related to science such as reading, taking notes, studying and comprehension. It is probable that the use of a metalinguistic approach utilizing listening, speaking, reading, and writing, helped the students to gain a broader perspective of the positive effects that writing can have on overall learning. Therefore, it is reasonable to conclude that the MSC intervention was a socially valid tool for students who struggle with literacy in science, which may be valid in other content areas such as social studies or math.

Limitations

This study has the following limitations. First, the study was a quasi-experimental design and the students were not chosen randomly from the population. Moreover, the participants were defined as students who struggled with literacy, from a middle school in Central Florida. The experimental and comparison classes were assigned due to a number of parameters to allow for equitable and comparable groups. However, these parameters preclude the researcher's ability to generalize from the experimental sample to a defined population sample (Gall et al., 2006). Conversely, the study was conducted with students who struggle with literacy, for which the results will add to the research base of writing interventions for those students.

A second limitation was the school setting, and the need to maintain flexibility with the school calendar and the high stakes assessment schedule. Although all attempts were made to schedule the pre/post testing at an optimal time; attendance was poor, which may have negatively affected the results. There was no time to reschedule the testing, when the students could or would be permitted to attend. In addition, although the intervention was planned for only twenty minutes, which is based on the recommendations of previous researchers in the area of sentence combining (Saddler, 2013; Strong, 1986), it only allowed for the experimental students to complete a total of two sentences per session. This limited amount of writing practice may have also hindered the speed of progress. However, the focus on the metalinguistic script and actions of the protocol, in lieu of more writing practice, may have been the reason for the gains that occurred in such a short time period.

Another limitation was the behavior of some of the student participants in the second experimental class. Researchers have documented that students who struggle with writing will try to avoid or “shut down” during academic tasks that require writing (Singer & Bashir, 2004). This is particularly true for students who cannot read on grade level (Scott, 2004). The researcher was informed prior to the study that a select few of the students in that class had been inconsistent in their willingness to complete their work during the science class. Overall, the majority of the students in the second experimental class completed the two sentence combinations required for each session. However, a few students struggled with listening, maintaining attention, following directions, and staying on task for the twenty-minute period. The researcher noted that these students had also achieved lower scores on the state reading assessment the year prior. The academic content, paired with a writing intervention and metalinguistic approach, may have been cognitively complex and taxing on their comprehension and attention skills, even for a short time period. Despite the explicit instruction, repetition, visual supports, and positive reinforcers that were provided by the researcher, these students appeared to avoid portions of the protocol, particularly when they had to listen and then write independently. At times, these few students needed oral encouragement and individual cues, by both the researcher and science classroom teacher, to complete the tasks of the day. Although the researcher anticipated that some students would have these difficulties due to their lower literacy skills, the intervention protocol was designed to provide a variety of supports for students who struggle with literacy. Any support given beyond what was specified in the intervention protocol, would have violated the fidelity of implementation. Unfortunately, those few students may not have achieved the gains that were within their capabilities, due to the difficulties noted above.

A last limitation, which will also be discussed in the Implications section, is the fact that the researcher, who created and implemented the MSC intervention, is a speech-language pathologist (SLP). SLPs have specialized expertise in the language underpinnings of text; the reciprocal language processes of listening, speaking, reading, and writing; and knowledge of metalinguistic characteristics that would vary across disciplinary expository text (American Speech Language Hearing Association [ASHA], 2010; Ehren et al., 2012). Sentence combining has been a practice that has been used, researched, and documented across numerous professions such as special education, regular education, collegiate education, English composition, and psychology. However, this study did employ a unique metalinguistic approach. Therefore, it cannot be reasonably assumed that an instructor without the language expertise of a speech-language pathologist could easily implement the metalinguistic approach that accompanied the sentence combining. However, the high fidelity of implementation is a positive indication that this approach was systematic enough to be replicated if high quality professional learning was provided by an SLP. High quality professional learning programs should be implemented with consistent and persistent support. The goal is to bring about change in the classroom practices of teachers, in their attitudes and beliefs, and ultimately in the learning outcomes of students (Guskey, 2002).

Implications for Practice

This study has practical implications. First, sentence combining is an evidence based-writing intervention that is recommended repeatedly in current research literature related to reading, writing, and language (e.g., Andrews et al., 2006; Eberhardt, 2013; Farrall, 2013; Graham & Perin, 2007b; Hillocks, 1986; Saddler, 2012, 2013; Scott & Balthazar, 2010, 2013; Scott & Nelson, 2009; Troia & Olinghouse, 2013). It has been recommended for students of all ages and varied skill levels. Sentence combining is also an intervention that is flexible, versatile, and usable in any academic context. Although sentence combining is not specifically prescribed in the CCSS, it is an intervention that targets the syntactic structures needed for the higher level writing skills that are expected in the CCSS (Troia & Olinghouse, 2013). This study has provided evidence that this type of intervention can yield gains in sentence complexity for a targeted semantic purpose (comparison/contrast) and to a different writing context (expository essay) in a specific academic discipline. The current writing demands for secondary students require style, lexicon, detail, and command of writing conventions within content-related organization (Mo et al., 2014). Therefore, MSC intervention would be an intervention that would be effective in targeting these writing skills. More importantly, this study used a metalinguistic approach with a writing intervention. This approach targeted the linguistic aspects of comparison and contrast, which improved science content knowledge. The semantic concept of comparison and contrast is targeted across academic subjects; therefore, the intervention may yield gains applying these concepts in other academic subjects.

In relation to content area skills, researchers have documented that content area teachers often do not include writing activities into their daily teaching due to time constraints and the fear that they will not cover all the content information that their curriculum requires (Gillespie et al., 201; Graham et al., 2014a). The MSC intervention used expository information directly from the curriculum, for only 20 minutes. Both teachers commented on their survey that MSC was an intervention that could be used in other academic content classes. Ironically, although the teachers' only concern was a lack of time to cover content, the MSC intervention was the most effective in improving the student's scores on the science compare and contrast double bubble map, a content acquisition task. If the protocol was reduced to fewer than 20 minutes, over the course of a school year, this type of writing intervention may be feasible, with less of an impact on the amount of time taken away from other areas of instruction. This approach may alleviate some of the teachers concerns with loss of time, which mirror teacher concerns documented by researchers (Gillespie et al., 2014; Graham et al., 2014). However, this intervention did have statistically significant gains in the determination of comparison and contrast in science, which is an academic gain in science content. Therefore, the intervention time was beneficial and should not be viewed as taking time away from learning science content, but as time facilitating the metalinguistic awareness and skills needed for mastering science content.

As mentioned in the limitations, the researcher is a speech-language pathologist with language expertise. Although this limits the assumption that any instructor could easily use the metalinguistic approach, it does provide further evidence that supports the role of the speech language pathologist as a valuable school professional and literacy partner. ASHA (2001, 2010) has specified that it is within the scope of the profession of speech-language pathology to address listening, speaking, reading and writing for adolescents with and without disabilities. In addition, because of the language expertise of the SLP, collaboration with other professionals is expected. This expertise in language will continue to benefit secondary schools that now have higher literacy demands in academic disciplines, especially in writing. The implementation of the CCSS (2010) solidified these expectations for writing across all academic disciplines at the secondary level, and the implementation of NGSS (2013) furthered the need for advanced writing skills in science for a variety of semantic purposes. These advanced writing skills in disciplines such as science require knowledge of the linguistic structures inherent in the discipline. Research has already documented that although content area teachers are aware that they need to target writing in their discipline, many do not only have the preparation in how to teach writing or how to teach the literacy of their discipline. An SLP is an professional who could collaborate with a variety of professionals to aid in the understanding of the linguistic underpinnings of a discipline, and how these underpinnings affect content-area learning (Ehren et al., 2012). The SLP can collaborate and provide professional learning with professionals who make direct contact with students such as content area teachers, special education teachers, or teachers of English learners (Zygouris-Coe, 2012). The SLP may also provide their language expertise when collaborating with school professionals who may not make direct contact with students, but who make critical decisions related to curriculum and instruction of academic disciplines, such as literacy coachers, curriculum resource teachers, school psychologists and school administrators. It is also important to note that many students struggle with literacy at the secondary level; therefore, the SLP's role extends to all students, not only students who are only diagnosed with disabilities. Less than 30% of the students who received the intervention in this study were diagnosed with a disability and were able to benefit from the intervention.

Another implication that should be noted relates to the intervention being conducted in the science classroom, during the science instruction. By being in the science classroom, the researcher was able to work within the heart of the discipline. The researcher was able to collaborate with the teachers and ask them questions during and after the intervention to further her understanding of the science content and how it should be conveyed syntactically. In addition, the researcher was exposed to the subtle nuances of the science classroom such the organization of the room, presentation of science information, the type of instructional activities that occur in science, the typical language expectations, the use of inquiry based activities, and the academic register used in discussion of science content. As an SLP with expertise in language, the researcher was sensitive to all aspects of language (listening, speaking, reading, and writing), as well as how the classroom environment and pedagogical practices in science influenced the optimum implementation of the protocol.

Moreover, in relation to disciplinary literacy, the text of the discipline dictates the literacy processes needed to engage and master its content (Brozo et. al, 2013; Fang & Schleppegrell, 2010). It is important to note that the literacy/language processes are not just reading and writing, but include listening and speaking, If the researcher had conducted the intervention outside of the science classroom, she may not have been immersed in the listening and speaking aspects of the science discipline. For example, in relation to the orientation of the room, the students were not facing forward, which could have had a negative impact on their ability to listen, read, and write during the intervention. The researcher had to modify the environment to create a closer proximity between the students and the researcher to optimize the students' ability to listen, read, and write fluidly. In relation to the pedagogical practices, the students were used to learning information through lecture with provided notes that they could follow. The MSC intervention required the students to listen, speak, think and then write about the information, simultaneously. Without the explicit instructional components of the protocol, the students may not have been able to adapt to these different language demands in such a short time. Furthermore, often SLPs at the secondary level do not conduct their intervention in the classroom and their services are delivered outside of the classroom (Wallach, 2014). This is despite the documented benefits of an SLP providing instruction or remediation in the classroom (e.g., Ehren, 2000; Ehren & Ehren, 2001; Ehren & Whitmire, 2009) and implementing functional curricular-related tasks that target the complex disciplinary demands at the secondary level (Blosser, Roth, Paul, Ehren, Nelson, Sturm, 2012; Ehren et al., 2012). This study supports the conclusion that the instruction, by an SLP in the literacy of a discipline, may have had a greater impact on disciplinary knowledge and skills because the intervention was conducted in and during academic instruction in that discipline.

Recommendations for Future Research

As stated previously, sentence combining is an evidence-based writing intervention that has been continually recommended by researchers. Yet, there has been limited SC research conducted since the new millennium. The three most current studies (Saddler, Asaro, et al., 2008; Saddler, Behforooz, et al., 2008; Saddler & Graham 2005) all utilized the same intervention protocol. All of the studies were also with late elementary students. Two of those studies were single subject design with fourth graders (Saddler, Asaro, et al., 2008; Saddler, Behforooz, et al., 2008) with sample sizes of four to six participants. Although the results of this study do contribute data to the knowledge base of sentence combining, the factors of age and sample size limit the extent to which the results can be generalized. Therefore, the first recommendation is to conduct sentence-combining studies that have larger samples and with a variety of age groups. Even more imperative, studies conducted with adolescents in sentence combining, would add to a research base in adolescent SC that has been dormant since the mid 1990's.

Second, sentence-combining studies utilizing the metalinguistic approach should be replicated, but with some modifications. One reason this study may have yielded gains in specific areas is its unique metalinguistic approach. It would be imperative to replicate this study using the same metalinguistic approach, to see if it yields the same effects. Since the short length of the protocol was a possible limitation, if the study were replicated, it may prove beneficial to increase the dosage for one group, maintain the 400 minutes dosage for another group, and a control group that receives no intervention. This design may shed light on whether the length of the intervention was truly a detriment to greater gains.

Another factor recommended for consideration when replicating this study would be to conduct the same protocol, but with a randomized controlled design. This would then aid in the generalizability of the protocol for students who may not struggle with literacy, but may also benefit from the intervention. It would also reduce some of the possible threats to internal validity (Campbell & Stanley, 1965).

Another research recommendation relates to the academic content area. This study was conducted in a science classroom. The study should be replicated in other content areas such as social studies, math, and English language arts. Such research would provide further evidence that this intervention can be effective for any content area. Similarly, this intervention only focused on writing for the purpose of comparison and contrast. Another study might explore other discourse structures, such as cause and effect, persuasion, or argumentation. Developing metalinguistic awareness can help students derive meaning across content areas that may require different semantic purposes (Ehren et al., 2012). Thus, it would be a reasonable assumption that this intervention, which used a metalinguistic approach, could be beneficial for students across academic disciplines.

Another research recommendation would be to find alternate ways to measure aspects of writing that are not merely scored by evidence of the structure or no evidence of the structure. Specifically, variables such as agentless passive voice or targeted connectives were not evident in the writing samples of many students. Lack of evidence of a variable in the writing samples resulted in a significant amount of zero scores. This result created outliers and negatively skewed distributions. These types of measures could be analyzed separately or descriptively to avoid impacting the MANOVA analysis. They may require a much larger sample size to allow for better overall distribution of scores.

Lastly, a significant number of studies with sentence combining prior to 2000, explored the effect of sentence combining on reading comprehension. This study did not explore the impact on students' performance in that area. However, metalinguistic awareness is essential for improved reading skills; students who have strong metalinguistic awareness are typically more successful with reading and writing (Armbruster et al., 1983; Burkhalter, 1996; Ebbels & van der Lely, 2000; Ebbels et al., 2007; Finestack, 2013; Hirschman, 2000; Hodgson, 1992; Lightsey & Frey, 2004; Scholl & Ryan, 1980). Furthermore, this study used the reciprocal processes of listening, speaking, reading and writing, so improving the metalinguistic skills needed for reading, may already be inherent in the design of the MSC protocol. Metalinguistic awareness uses conscious awareness to access linguistic knowledge to manipulate language for a variety of goals (Peets, 2014). Therefore, the possibility that this protocol could improve reading comprehension is a reasonable assumption.

Conclusion

The findings of this study revealed a significant improvement in the ability to determine similarities and differences (comparison/contrast) in science, indicating mastery of science content. The experimental group exhibited a greater increase in score from pretest to posttest in sentence combining with small effect. The experimental group also had greater gains in measures of sentence complexity and use of syntactic structures that are inherent in science on both the science compare and contrast written prompt with a small effect and science expository essay with a medium to large effect. These findings are crucial, as they add empirical data supporting the use of a metalinguistic approach to sentence combining, with secondary students who struggle with the literacy. More importantly, this study used complex expository text and discipline specific syntactic structure targets which now establish a new and much needed empirical evidence to add to sentence-combining research literature.

Some of the results should be interpreted with caution. As has been previously stated, for research questions two and three, the reduced sample size due to missing data may not have adequately reflected the effectiveness of the intervention on writing skills. Further research is warranted to continue investigating the effects of a metalinguistic approach to sentence combining for other secondary aged students, using larger sample sizes and longer duration of time, employing a randomized controlled design, targeting other academic content areas and semantic purposes, controlling variables that yield a large number of outliers, and assessing the effect of this type of protocol on reading comprehension.

APPENDIX A: INSTITUTIONAL REVIEW BOARD APPROVAL LETTER

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Lynne F. Telesca**

Date: **January 16, 2015**

Dear Researcher:

On 01/16/2015, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	The Effect of Metalinguistic Sentence Combining on Writing in Eighth Grade Science for Students who Struggle with Literacy
Investigator:	Lynne F Telesca
IRB Number:	SBE-14-10841
Funding Agency:	
Grant Title:	
Research ID:	N/A

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

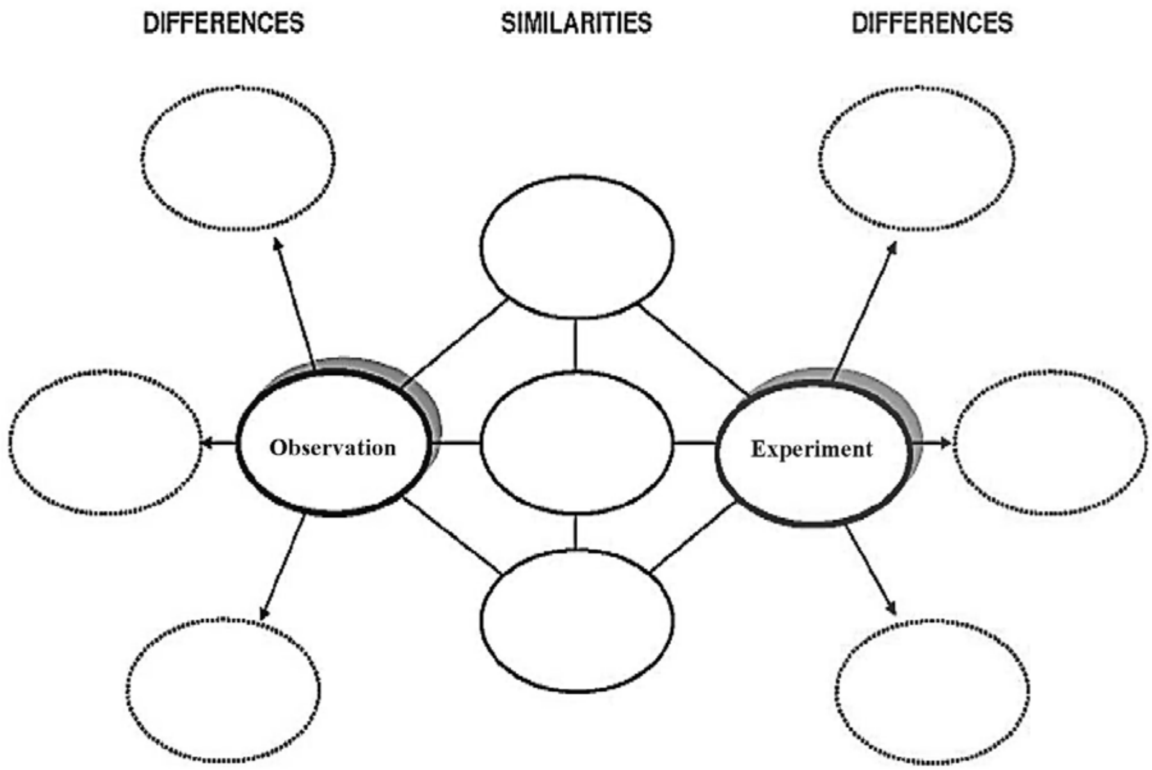
Joanne Muratori

Signature applied by Joanne Muratori on 01/16/2015 10:41:44 AM EST

IRB Coordinator

Page 1 of 1

APPENDIX B: SCIENCE COMPARE AND CONTRAST DOUBLE BUBBLE MAP


DIFFERENCES
SIMILARITIES
DIFFERENCES
Observation
Experiment

APPENDIX C: EXPERIMENTAL AND COMPARISON TEACHER PRE-SURVEY

Experimental & Comparison Teacher Pre-survey

Teacher Name:_______________________________

On average, how long are students writing in your science class? (Circle one).

(1) Never
(2) 5-10 minutes
(3) 10-20 minutes
(4) 20-30 minutes
(5) 30-40 minutes
(6) The entire class period

What type(s) of activities do these students participate in when they write in class? (Circle all that apply).

(a) Note taking while listening
(b) Note taking while reading
(c) Filling out a graphic organizer
(d) Filling out an outline
(e) Fill in the blank answers to questions
(f) Short answers to questions
(g) Writing about a picture
(h) Journal entry
(i) Lab report
(j) Writing a list
(k) Writing instructions or step by step procedure
(l) Writing a paragraph (4 or more sentences)
(m) Writing an essay (more than one paragraph)
(n) Synthesizing information from multiple sources
(o) Other ____

What type(s) of writing do these students participate in when they are assessed in science?

(a) Circling multiple choice answers
(b) Fill in the blank
(c) Short answer questions
(d) Questions that require a few sentences to answer (a paragraph)
(e) Essay questions (More than a paragraph answer

What type(s) of writing do these students participate in when they are completing homework in science?

(a) Note taking while reading
(b) Filling out a graphic organizer
(c) Filling out an outline
(d) Fill in the blank answers to questions
(e) Short answers to questions
(f) Writing about a picture
(g) Journal entry
(h) Lab report
(i) Writing a list
(j) Writing instructions or step by step procedure
(k) Writing a paragraph (4 or more sentences)
(l) Writing an essay (more than one paragraph)
(m) Synthesizing information from multiple sources
(n) Other ____

With regard to writing, what do you feel your students in this class struggle with the most? Circle all that apply.

(a) Punctuation
(b) Capitalization
(c) Spelling
(d) Vocabulary
(e) Using a complete sentence
(f) Using correct grammar (correct verb tense, plurals possessives etc.)
(g) Writing long complex sentences that are not run on sentences
(h) Staying on topic
(i) Maintaining attention to the writing task
(j) Motivation to write
(k) Knowing what to write as far as content

What type of writing do you model by writing in class for the students? (Circle all that apply)

(a) Spelling
(b) Vocabulary words
(c) Definitions
(d) Goals for the day or assignments (abbreviated)
(e) Goals for the day or assignments (written in full sentences)
(f) Phrases
(g) Complete sentences with correct grammar
(h) Sentences to answer a specific prompt or question
(i) A summary
(j) I use speaking the majority of the time to teach my class

Do you feel that incorporating writing into science is important for your students? (Circle one)

Yes   No   Why or why not?

My teaching degree is:

(a) Bachelor's degree in ____
(b) Master's degree in
____

I have been teaching for:

(a) less than 2 years
(b) 2-5 years
(c) 5-10 years
(d) 10-15 years
(e) 15+ years

I have been teaching science for:

(a) less than 2 years
(b) 2-5 years
(c) 5-10 years
(d) 10-15 years
(e) 15+ years

I completed the following courses in writing when obtaining my degree:

(a) No specific courses on how to teach writing
(b) One course that may have incorporated techniques on how to teach writing
(c) One or more courses that incorporated techniques on how to teach writing
(d) Specific explicit instruction on how to teach writing and supervised practice

I have completed the following professional development in writing outside of my degree:

(a) No professional development on how to teach writing
(b) Some informal training on how to teach writing
(c) One formal inservice or professional learning on how to teach writing
(d) A few inservices or professional learning on how to teach writing
(e) Several courses on writing

My school administration expectations of teaching writing as science teacher is: (Circle all that apply).

(a) Writing in my class in not part of my job
(b) Writing is something I should have the student's participate in, but the focus of the class is learning science content.
(c) Writing is just as important as learning the science content.
(d) Writing is the teacher's responsibility to teach at Legacy Middle.

I feel that teaching writing is: (Circle all that apply)

APPENDIX D: EXPERIMENTAL AND COMPARISON TEACHER POST-SURVEY

Experimental & Comparison Teacher Post Survey

What type of activities do students complete in your class that require writing?

Approximately how many minutes per day are students writing in your class?

What type of writing do you model in class for the students?

What kind of writing do they have to do on a test?

With regard to writing, what do you feel your students in this class struggle with the most?

What is the expectation of your school administration with regard to science teachers and the teaching of writing?

Do you feel that incorporating writing into science is important for your students? Why or why not?

My teaching degree is:

(a) Bachelor's degree in ____
(b) Master's degree in ____

I have been teaching for:

I have been teaching science for:

I feel that teaching writing is:

APPENDIX E: EXPERIMENTAL TEACHER SOCIAL VALIDITY POST-SURVEY

Experimental Teacher Post-Survey

1. **What are your overall impressions of Metalinguistic Sentence Combining (MSC)?**

2. **Do you think use MSC class had a positive impact on student writing in science? Why or why not?**

3. **Do you think students developed better metalinguistic awareness of the grammatical constructions that may be found in science for comparison and contrast? Why or why not?**

4. **What are your thoughts regarding the amount of time MSC required in science class?**

5. **How likely are you to use MSC again in future units? Why or why not?**

6. **How likely are you to recommend MSC to colleagues within your content area? Within other content areas?**

APPENDIX F: STUDENT PRE-SURVEY

Experimental Student Pre-Survey

Teacher:____

1. **I like to write**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

2. **I think writing is important in all my classes.**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

3. **I think writing is important in science.**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

Why may writing be important for your future?

APPENDIX G: STUDENT POST-SURVEY

Experimental Student Post-Survey

Teacher:____

1. **I like to write**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

2. **I think writing is important in all my classes.**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

3. **I think writing is important in <u>science.</u>**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

Why may writing be important for your future?

4. **I think that learning metalinguistic sentence combining (Writer) helped my writing in science class.**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

5. **I think that learning metalinguistic sentence combining (Writer) helped me to think more when I am writing.**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

6. **I think that metalinguistic sentence combining (Writer) helped me to read and learn the science concepts better.**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

7. **I think learning sentence-combining will help me to write in all my classes.**

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	4	3	2	1

How may have MSC affected your writing for the future?

APPENDIX H: ADVANCE ORGANIZER-IMAGERY AND PURPOSE

Metalinguistic Sentence Combining (MSC) ©

Metalinguistic means to THINK ABOUT LANGUAGE

Purpose: We will take simple sentences and combine them into mature complex sentences to <u>improve our writing</u>. We read, think, and then, write.

To remember how to do MSC we use:

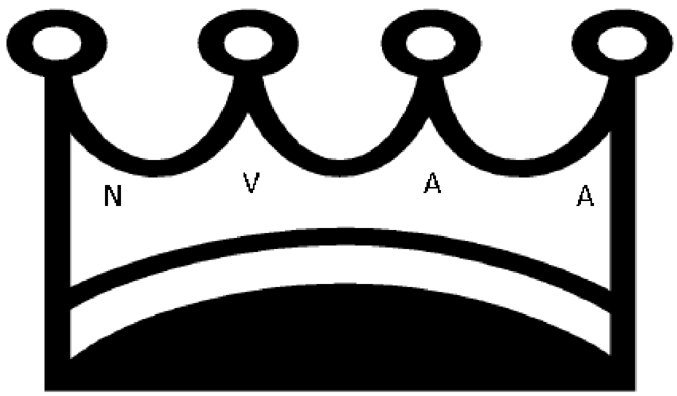
♕WRIT<u>Er</u>.

"You are the writer. You decide how you are going to say what you need or want to say. You are telling the reader how to read your message. <u>You are the writer</u>"

"For this class we are going to be science writers. Science writers have to read, think, and discover the clues that tell about similarities and differences, and then write clear, correct, and mature sentences." ©

APPENDIX I: ADVANCE ORGANIZER PARTS OF SPEECH REVIEW CROWN

"Put your crown on"



Nouns-person, place or thing

Verbs-action words or <u>being</u> words

Adjectives-words that describe a noun

Adverbs-words that describe a verb©

APPENDIX J: METALINGUISTIC SENTENCE COMBINING EXERCISE-PHASE ONE

Name:________________________________

MSC DAY 4

- Combine these 2 sentences into 1 sentence.
- You cannot use "because"
- You must use all the science information in your new sentence, but you can add or change words.
- You must use one of the connector words in the box below

Similarities	**Differences**
both	but
like	;however

Physical properties are unique to a particular substance.

Chemical properties are unique to a particular substance.

↘

APPENDIX K: METALINGUISTIC SENTENCE COMBINING EXERCISE-PHASE TWO

Name:_______________________________

MSC-DAY 9-PHASE 2

- Combine these 2 sentences into 4 sentences.
- You cannot use "because"
- You must use all the science information in your new sentence, but you can add or change words.
- You must use one of the connector words in the box below

Similarities	**Differences**
both	but
like	;however

#1	**# 2**
Protons are inside the nucleus of an atom. Neutrons are inside the nucleus of an atom. A proton has a positive charge. Neutrons have no electric charge.	Protons have a positive charge Electrons have a negative charge. An election is a subatomic particle. A proton is a subatomic particle.
↘	↘

APPENDIX L: METALINGUISTIC SENTENCE COMBINING EXERCISE-PHASE THREE

Name:________________________________

MSC-DAY 17 & 18-PHASE 3

- Combine these 3 sentences into 1 sentence.
- You must use all the science information in your new sentence, but you can add or change words.
- You must keep all underlined portions of the third sentence.
- Connect the third sentence with the word in (parenthesis)
- You must still use one of the connector words in the box below and you still cannot use "because"

Similarities	**Differences**
both	,but
like	;however,

DAY 17	**DAY 18**
Elements cannot be altered by physical changes. Physical changes cannot alter compounds. Physical changes happen in the environment.(that)	The identity of an element stays the same. The compound's identity stays the same. A physical change happens.(after)
↘	↘

APPENDIX M: WRITER MNEMONIC

©	**Goal:** 1. **Combine two sentences into one sentence to compare or contrast science information.**
W	**WORD** your message to achieve your goal.
R	**REMOVE** words you do not need.
I	**INTEGRATE** the words into a new sentence.
T	**TEST** your sentences.
Er. ©	**Erase** your mechanical errors. ©

APPENDIX N: METALINGUISTIC SCRIPT

Metalinguistic Cues & Overt Actions by Mnemonic Steps & Language Areas ©

Word

ASK (Metalinguistic Cues)	DO (Overt Actions)
Is this a similarity or a difference?	
How do you know?	
	Circle key words
Which conjunction should we use?	
Why?	
	Circle conjunction needed (Phase 1) Write S next to similarity sentences and D next to difference sentences (Phase 2 & 3)
What is the third sentence telling us more about? Is it about the noun or like an adjective? Is it about the verb or like an adverb? (Phase 3 only)	

Remove

ASK (Metalinguistic Cues)	DO (Overt Actions)
Which words can you remove?	
	Cross out words
Why?	
Which words need to stay?	
	Underline words
Why?	

Integrate

ASK (Metalinguistic Cues)	DO (Overt Actions)
Say aloud the sentence portions that will make up the entire sentence	
	Write sentence portions that will make up the entire sentence
Why?	
	Cross out portions as used
Why?	

Test

ASK (Metalinguistic Cues)	DO (Overt Actions)
Read the sentence aloud	
Does this sound correct?	
	Make corrections to sentence.
Read aloud all the noun/verb pairs in clauses.	
Do these sound correct?	
Why or why not?	
	Make corrections to sentence.

Erase.

ASK (Metalinguistic Cues)	DO (Overt Actions)
	Check from left to right. Make corrections as needed.
Is there a capital letter?	
Why do we need the capital?	
Do we need comma/semicolon?	
Why or why not?	
Is there a period?	
Why?	

Language Area	Metalinguistic Questions
Phonologic	If that word is spelled incorrectly, what may confuse the reader? Is the word somewhere on the paper that you can copy correctly?
Semantic	Is that a noun or a verb? Does that convey a similarity or a difference? What key words tell you it is a similarity or difference? What does the word/phrase mean? Is that clause like an adjective or an adverb?
Morphologic	How does that word ending change the word? Does the end of that word match the word before it? When you add that ending to the word, what does that mean?
Syntactic	Where do I put this part of the sentence? What part of the sentence does that clause refer to? What word should I start the sentence with? By putting the clause there, where do I have to put the other parts of the sentence?
Pragmatic	What message does that capital letter send to the reader? What message does that punctuation mark send to the reader? What message does that word send to the reader? What message might the reader perceive?

APPENDIX O: METALINGUISTIC SENTENCE COMBINING INTERVENTION OUTLINE

Metalinguistic Sentence Combining Intervention Outline

Pre-surveys to experimental and comparison teachers& experimental students			
Pretest: TOWL-4 Sentence Combining Subtest Form A & Science compare/contrast double bubble map			
Pretest: Science compare and contrast written prompt and Science expository essay			
day	**Instructional Phase**	**Concept**	**Instructional Level of Support**
1	**ONE**: 2 sentences. to 1 sentence	compare only	TEACHER MODEL
2	**ONE**: 2 sentences. to 1 sentence	compare only	TEACHER MODEL
3	**ONE**: 2 sentences. to 1 sentence	contrast only	TEACHER MODEL
4	**ONE**:: 2 sentences. to 1 sentence	contrast only	STUDENT ENLISTMENT MODEL
5	**ONE**: 2 sentences. to 1 sentence	compare/contrast	FADED GUIDED PRACTICE
6	**ONE**: 2 sentences. to 1 sentence	compare /contrast	GUIDED PRACTICE
7	**ONE**: 2 sentences. to 1 sentence	compare /contrast	PEER-SUPPORTED PRACTICE
8	**ONE**: 2 sentences. to 1 sentence	compare /contrast	INDEPENDENT PRACTICE
9	**TWO:** 4 sentences to 2 sentences	compare /contrast	TEACHER MODEL
10	**TWO:** 4 sentences to 2 sentences	compare /contrast	STUDENT ENLISTMENT MODEL
11	**TWO:** 4 sentences to 2 sentences	compare /contrast	GUIDED PRACTICE
12	**TWO:** 4 sentences to 2 sentences	compare /contrast	FADED GUIDED PRACTICE
13	**TWO:** 4 sentences to 2 sentences	compare /contrast	PEER-SUPPORTED PRACTICE
14	**TWO:** 4 sentences to 2 sentences	compare /contrast	INDEPENDENT PRACTICE
15	**THREE:** 3 sentences to 1 sentence	compare /contrast + clause	TEACHER MODEL
16	**THREE:** 3 sentences to 1 sentence	compare /contrast + clause	STUDENT ENLISTMENT MODEL
17	**THREE:** 3 sentences to 1 sentence	compare /contrast + clause	GUIDED PRACTICE
18	**THREE:** 3 sentences to 1 sentence	compare /contrast + clause	FADED GUIDED PRACTICE
19	**THREE:** 3 sentences to 1 sentence	compare /contrast + clause	PEER-SUPPORTED PRACTICE
20	**THREE:** 3 sentences to 1 sentence	compare /contrast + clause	INDEPENDENT PRACTICE
Posttest: TOWL-4 Sentence Combining Subtest Form B, Science compare/contrast double bubble map			
Posttest: Science compare and contrast written prompt and Science expository essay			
Post-surveys to experimental and comparison teachers & experimental students			

APPENDIX P: METALINGUISTIC SENTENCE COMBINING SCOPE AND SEQUENCE PLAN

Metalinguistic Sentence Combining Scope and Sequence Plan

PHASE ONE	OUTCOME	# OF DAYS
PRE-INTERVENTION	Teachers will answer questions in multiple-choice format related to student writing expectations, writing time, writing tasks, their perception of the importance of writing, and provide demographic background information. Students will answer questions related their perceptions of writing and writing in science.	1
PRETESTING	Students will demonstrate their skills in sentence combining, comparing and contrasting science information on a graphic organizer and in written form, as well as writing a science expository essay.	2
Combine two sentences into one sentence to compare and contrast	Students will combine two sentences into one sentence using comparing conjunctions with a teacher model.	2
	Students will combine two sentences into one sentence using comparing conjunctions with a teacher model.	1
	Students will combine two sentences into one sentence using contrasting conjunctions with a teacher model.	1
	Students will combine two sentences into one sentence using comparing conjunctions with visual, verbal, and yes/no question prompts (student enlistment model).	1
EDITING BEGINS	Students will correct mechanical or morpho-syntactic given uncoded feedback. Students will combine two sentences into one sentence using comparing/contrasting conjunctions with visual, verbal, and wh-question prompts (guided practice).	1
	Students will correct mechanical or morpho-syntactic given uncoded feedback. Students will combine two sentences into one sentence using comparing/contrasting conjunctions with faded teacher prompts (faded guided practice).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine two sentences into one sentence using comparing/contrasting conjunctions with peer support (peer-supported practice).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine two sentences into one sentence using comparing/contrasting conjunctions independently (independent practice).	1

PHASE TWO	OUTCOME	# OF DAYS
Combine two sentence into four sentences to compare and contrast	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using a comparing/contrasting conjunctions with a teacher model.	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using a comparing conjunction with visual, verbal, and yes/no question prompts (student enlistment model).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using a comparing/contrasting conjunction with visual, verbal, and wh-question prompts (guided practice).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using a comparing/contrasting conjunction with faded teacher prompts (faded guided practice).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using a comparing/contrasting conjunction with peer support (peer-supported practice).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using comparing/contrasting conjunctions independently (independent practice).	1

PHASE THREE	OUTCOME	# OF DAYS
Combine three sentence into one sentences to compare and contrast	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using comparing/contrasting conjunctions with a teacher model.	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using comparing conjunctions with visual, verbal, and yes/no question prompts (student enlistment model).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using comparing/contrasting conjunctions with visual, verbal, and wh-question prompts (guided practice).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using a comparing/contrasting conjunctions with faded teacher prompts (faded guided practice).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using a comparing/contrasting conjunctions with peer support (peer-supported practice).	1
	Students will correct mechanical or morpho-syntactic errors given uncoded feedback. Students will combine four sentences into two sentences using comparing/contrasting conjunctions independently (independent practice).	1
POSTTESTING	Students will demonstrate their skills in sentence combining, comparing and contrasting science information on a graphic organizer and in written form, as well as writing a science expository essay.	2
POST-SURVEYS	Teachers will answer questions in open-ended format related to student writing expectations, writing time, writing tasks, their perception of writing and provide demographic background information. Students will answer questions related their perceptions of writing, writing in science, as well as questions related to MSC intervention.	2

APPENDIX Q: INNOVATION CONFIGURATION (IC) MAP

Metalinguistic Sentence Combining Innovation Configuration Map

	Ideal Implementation (4)	**In Process (3)**	**In Process (2)**	**In Process (1)**	**No Implementation (0)**
Distribution of Materials	Teacher distributes the MSC folders with worksheets prior to the intervention.		Teacher may distribute the MSC folders with worksheets prior to the intervention.		Teacher forgets to distribute the MSC folders with worksheets prior to the intervention.
Advance Organizer Purpose	Teacher states the purpose of the intervention in relation to the discipline of science and the goal of comparison and contrast.	Teacher states the purpose, but forgets either to relate it to the discipline of science or the goal of comparison and contrast.		Teacher vaguely states the purpose of the intervention.	Teacher does not state the purpose of the intervention.
Background Knowledge "Put on your crown"	Teacher discusses crown of knowledge analogy and reviews all the parts of speech.	Teacher discusses the crown of knowledge analogy and parts of speech, but omits some points.	Teacher discusses the crown of knowledge analogy and parts of speech, but omits most of the points.	Teacher mentions the crown and that it is related to prior knowledge.	Teacher omits this step.
Mnemonic WRITEr.	Teacher reviews the mnemonic or asks students to recall the mnemonic parts.	Teacher presents the mnemonic but forgets to incorporate student recall of the mnemonic parts.	Teacher presents most of the mnemonic correctly, but forgets some parts.	Teacher presents the mnemonic incorrectly.	Teacher does not talk about the mnemonic.
Goals	Teacher presents the goals of the day or asks the students to recall previous goals.	Teacher presents the goals of the day, but forgets to incorporate student recall of previous goals.	Teacher presents only part of the goals of the day and forgets to incorporate students' recall of previous goals.	Teacher presents the goals unclearly or in a disorganized fashion.	Teacher omits the goals.

Instructions	Teacher reads the directions and the sentences aloud.		Teacher forgets to either read the directions or read the sentences aloud.		Teacher fails to read the directions or the sentences aloud.
Metalinguistic Script	Teacher follows all the metalinguistic cue questions.	Teacher remembers more than 75% of the metalinguistic cue questions.	Teacher remembers 50-75% of the metalinguistic cue questions.	Teacher forgets the majority of the metalinguistic cues questions.	Teacher ignores the metalinguistic cues to ask the students.
Overt Actions	Teacher performs all the overt actions in sync with the metalinguistic cue questions.	Teacher completes some of the overt actions correctly.	Teacher completes some of the overt actions correctly.	Teacher attempts to use overt actions, but forgets the majority or completes them incorrectly.	Teacher omits the overt actions.
Editing	Teacher explicitly introduces the editing process, explains the uncoded feedback system, and allows students to edit their work the first two minutes of the remainder of sessions	Teacher explicitly discusses the editing process and the uncoded feedback system, but forgets to provide editing time consistently.	Teacher explicitly discusses the editing process and the uncoded feedback system, but forgets to provide editing time more than half of the sessions.	Teacher mentions the editing process, but did not explicitly explain the uncoded feedback system.	Teacher does not introduce or use the editing process.
Monitor Student Engagement	(DAYS 4-20). Teacher consistently scans the room to check that all students are engaged and provides verbal prompts for students who may lose focus. Teacher is systematically inclusive when asking students to answer questions.	(DAYS 4-20) Teacher periodically scans the room to check that all students are engaged and provides verbal prompts for students who may lose focus. Teacher asks most students to answer questions.	(DAYS 4-20) Teacher occasionally scans the room to check that all students are engaged and provides verbal prompts for students who may lose focus. Teacher asks some students to answer questions.	(DAYS 4-20) Teacher may only notice student engagement when they are off task. Teacher will randomly ask students to answer questions.	(DAYS 4-20) Teacher does not check if students are engaged or off task.

Provide Intermittent Specific Feedback	Teacher provides intermittent specific feedback throughout the session.	Teacher provides some specific feedback.	Teacher provides intermittent feedback that is nonspecific.	Teacher may provide occasional general feedback.	Teacher provides no feedback throughout the session.
Collaboration	Teacher encourages and engages the classroom teacher during the intervention sessions.	Teacher attempts to engage and encourage the classroom teacher during the intervention sessions.	Teacher talks to classroom teacher outside the intervention but does not actively engage the classroom teacher during the intervention sessions.	Teacher acknowledges classroom teacher.	Teacher ignores the other teacher.
Closing Organizer	Teacher presents the goals of the day or asks the students to recall the goals and then reveals what will be accomplished the next session.	Teacher presents the goals of the day or asks the students to recall the goals but forgets to state what will be accomplished the next session.	Teacher presents only part of the goals of the day and forgets to incorporate student's recall of previous goals or state what will be accomplished the next class.	Teacher presents the goals unclearly or in a disorganized fashion.	Teacher omits the closing organizer.

APPENDIX R: FIDELITY OF IMPLEMENTATION CHECKLISTS

PHASE ONE-Interventionist Fidelity Daily Checklist-Metalinguistic Sentence Combining

Date:________ MSC-Day #:__________Class:_______________________________________

Metalinguistic Sentence Combining- Two sentences into one sentence DAYS 1-6		
Key Elements:	**+ present, absent –**	
	sentence 1	**sentence 2**
1. Provides student time to edit previous work (start day 4)		NA
2. Reviews/asks about purpose of MSC		NA
3. Reviews/writer's crown analogy (noun/verb/adj/adv)		NA
4. States/asks the goal of the day		NA
5. Reviews/asks each part of the WRITEr. © mnemonic		NA
6. Reads or reviews the instructions at the top of the worksheet		NA
7. Reads aloud the sentence set		
8. WORD- States/asks if sentences similar/different and any key words (circles sim/diff)		
9. WORD-States/asks which conjunction/transition to use (circles it)		
10. REMOVE States/asks to think about which words can and cannot be removed		
11. REMOVE Mentions/asks about key words or patterns related to what will be kept/removed		
12. INTEGRATE -Says sentence aloud while writing and crossing out as needed		
13. TEST-Reads the sentence aloud and asks if it sounds correct		
14. TEST Checks/asks if all nouns and verb match		
15. ERASE-Checks sentence from to LEFT to RIGHT-Capital, punctuation, period.		
16. ERASE-States/asks purpose capitals, punctuation, or period (started thought, took a pause, ended thought)		
17. States the goals of the day and states what they will be doing next session	NA	

Total out of 27: ________ _________%

PHASE TWO-Interventionist Fidelity Daily Checklist–Metalinguistic Sentence Combining

Date:________ MSC-Day #:__________Class:______________________________________

Metalinguistic Sentence Combining- Four sentences into two sentences DAYS 9-12

Key Elements: **+ present, absent**

	sentence 1	**sentence 2**
1. Provides student time to edit previous work		NA
2. Reviews/asks about purpose of MSC		NA
3. Reviews/writer's crown analogy (noun/verb/adj/adv)		NA
4. States/asks the goal of the day		NA
5. Reviews/asks each part of the WRITEr. © mnemonic		NA
6. Reads or reviews the instructions at the top of the worksheet		NA
7. Reads aloud sentence set		
8. WORD- States/asks if sentences similar/different and any key words (marks S and D)		
9. WORD-States/asks which conjunction/transition to use		
10. REMOVE States/asks to think about which words can and cannot be removed		
11. REMOVE Mentions/asks about key words or patterns related to what will be kept/removed		
12. INTEGRATE -Says sentence aloud while writing and crossing out as needed		
13. TEST-Reads the sentence aloud and asks if it sounds correct		
14. TEST Checks/asks if all nouns and verb agree		
15. ERASE-Checks sentence from to LEFT to RIGHT-Capital, punctuation, period.		
16. ERASE-States/asks purpose capitals, punctuation, or period (started thought, took a pause, ended thought)		
17. States the goals of the day and states what they will be doing next session	NA	

Total out of 27: ________ _________%

PHASE THREE-Interventionist Fidelity Daily Checklist-Metalinguistic Sentence Combining

Date:________ MSC-Day #:__________Class:______________________________________

Metalinguistic Sentence Combining- Three sentences into one sentence DAYS 15-18		
Key Elements: **+ present, absent –**		
	sentence 1	**sentence 2**
1. Provides student time to edit previous work		NA
2. Reviews/asks about purpose of MSC		NA
3. Reviews/writer's crown analogy (noun/verb/adj/adv)		NA
4. States/asks the goal of the day		NA
5. Reviews/asks each part of the WRITEr. © mnemonic		NA
6. Reads or reviews the instructions at the top of the worksheet		NA
7. Reads aloud the sentence set (may read all four at one time or two and two)		
8. WORD- States/asks if sentences similar/different and any key words (writes S or D)		
9. WORD-States/asks which conjunction/transition to use		
10. WORD-States/asks @ third sentence (subordinate clause) & its related part of speech (noun/verb/adj/adv)		
11. REMOVE States/asks to think about which words can and cannot be removed		
12. REMOVE Mentions/asks about key words or patterns related to what will be kept/removed		
13. INTEGRATE -Says sentence aloud while writing and crossing out as needed		
14. TEST-Reads the sentence aloud and asks if it sounds correct		
15. TEST Checks/asks if all nouns and verb match		
16. ERASE-Checks sentence from to LEFT to RIGHT-Capital, punctuation, period.		
17. ERASE-States/asks purpose capitals, punctuation, or period (started thought, took a pause, ended thought)		
18. Reviews one other sentence combination option (DAYS 16-18)		
19. States the goals of the day and states what they will be doing next session	NA	

Total out of 30: ________ _________%

PEER-SUPPORTED PRACTICE-Interventionist Fidelity Daily Checklist – Metalinguistic Sentence Combining

Date:________ MSC-Day #:__________Class:______________________________________

Metalinguistic Sentence Combining with PEER-supported Practice DAYS 7, 13, 19		
Key Elements: **+ present, absent –**		
	sentence	**sentence**
	1	**2**
1. Provides student time to edit previous work		NA
2. Reviews/asks about purpose of MSC		NA
3. Reviews/writer's crown analogy (noun/verb/adj/adv)		NA
4. States/asks the goal of the day		NA
5. Reviews/asks each part of the WRITEr. © mnemonic		NA
6. Reads or reviews the instructions at the top of the worksheet		NA
7. Explains that students can talk with each other complete two MSC exercises		NA
8. Reads the directions aloud		NA
9. Reads aloud the sentence set		
10. Tells each group which conjunction/transition they will be using.		
11. Scans the room, keeps time, and provides cues when needed.		
12. Chooses at least two different example sentence combinations to present and writes them on the board.		NA
13. Asks students questions about the examples, if they feel it is a correct, and why/why not if applicable		NA
14. States the goals of the day and states what they will be doing next session.	NA	

Total out of 17: ________ _________%

INDEPENDENT PRACTICE Interventionist Fidelity Daily Checklist – Metalinguistic Sentence Combining

Date:________ MSC-Day #:__________Class:_______________________________________

Metalinguistic Sentence Combining with INDEPENDENT Practice DAYS 8, 14, 20		
Key Elements:	**present, absent –**	
	sentence 1	**sentence 2**
1. Provides student time to edit previous work		NA
2. Reviews/asks about purpose of MSC		NA
3. Reviews/writer's crown analogy (noun/verb/adj/adv)		NA
4. States/asks the goal of the day		NA
5. Reviews/asks each part of the WRITEr. © mnemonic		NA
6. Reads or reviews the instructions at the top of the worksheet		NA
7. Explains that students will work independently		NA
8. Reads the directions aloud		NA
9. Reads aloud the sentence set		NA
10. Tells students to choose which conjunction they will be using.		
11. Scans the room, keeps time, and provides cues when needed.		
12. Provides at least two different example sentence combinations to present and writes them on the board.		NA
13. Asks students questions about the examples, if they feel it is a correct, and why/why not if applicable		NA
14. States the goals of the day and states what they will be doing next session.	NA	

Total out of 17: ________ _________%

APPENDIX S: INTER-RATER RELIABILITY TALLY SHEET

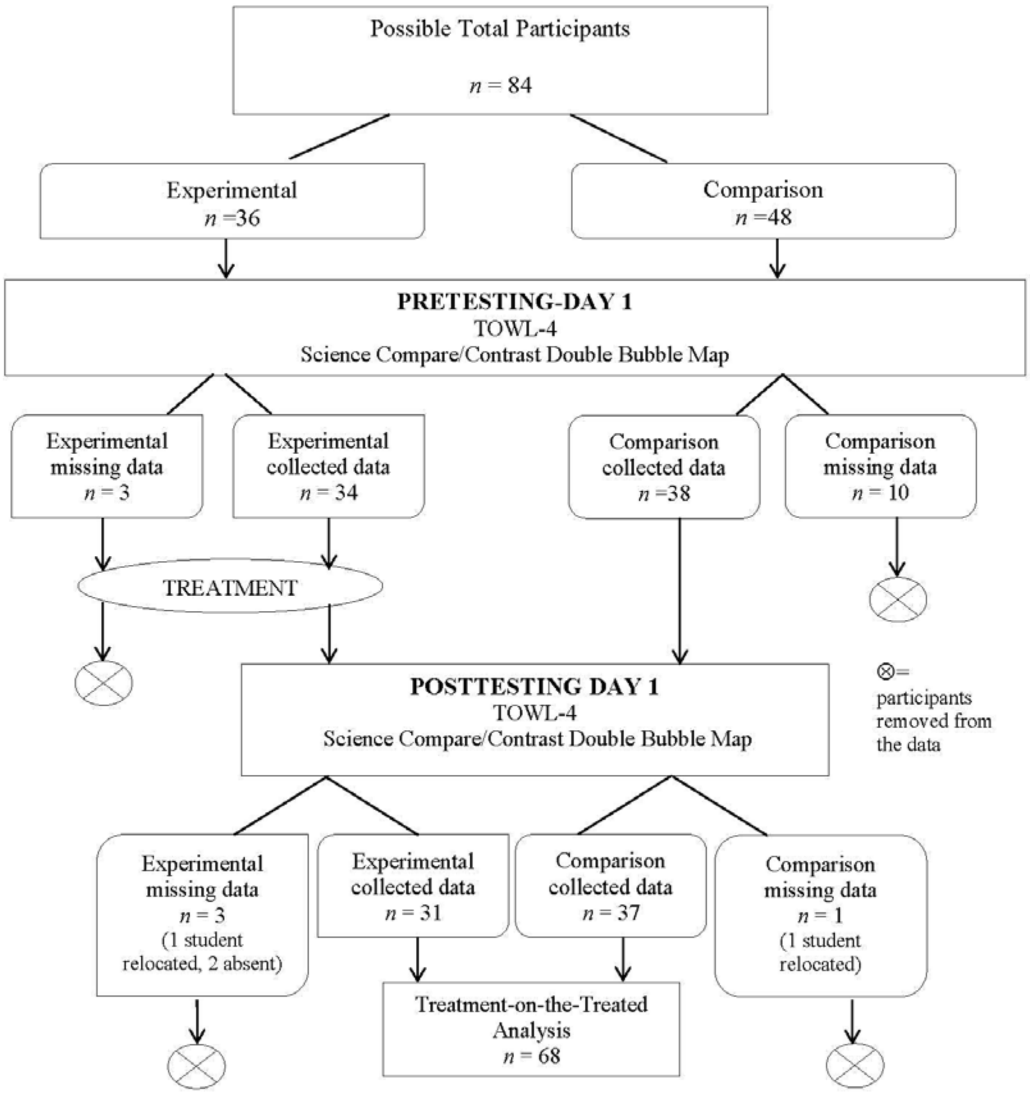
Inter-Rater Reliability Tally Sheet

INTER-RATER RELIABILITY CHECK FOR Assessment measure –
PRETEST/POSTTEST _______________________ GROUP #_________
(Raters – Researcher and 1 research assistant)

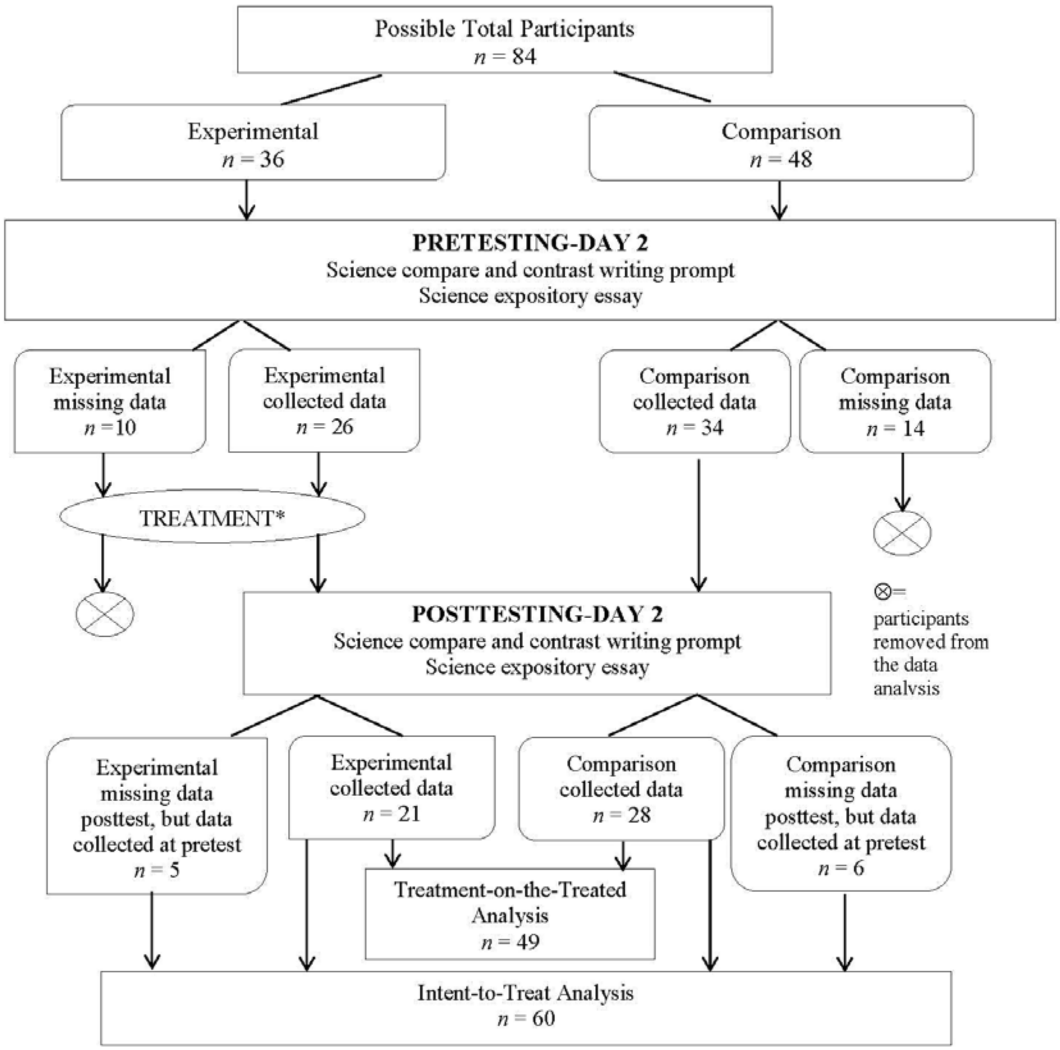
	Student Number	Rater 1	Rater 2	Match Yes/No
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

TOTAL RELIABILITY =

APPENDIX T: COLLECTED AND MISSING DATA PRETEST DAY ONE, POSTTEST DAY ONE


Possible Total Participants
n = 84
Experimental
n =36
Comparison
n =48
PRETESTING-DAY 1
TOWL-4
Science Compare/Contrast Double Bubble Map
Experimental missing data
n = 3
Experimental collected data
n = 34
Comparison collected data
n =38
Comparison missing data
n = 10
TREATMENT
POSTTESTING DAY 1
TOWL-4
Science Compare/Contrast Double Bubble Map
⊗= participants removed from the data
Experimental missing data
n = 3
(1 student relocated, 2 absent)
Experimental collected data
n = 31
Comparison collected data
n = 37
Comparison missing data
n = 1
(1 student relocated)
Treatment-on-the-Treated Analysis
n = 68

APPENDIX U: COLLECTED AND MISSING DATA PRETEST DAY TWO, POSTTEST DAY TWO


Possible Total Participants
n = 84
Experimental
n = 36
Comparison
n = 48
PRETESTING-DAY 2
Science compare and contrast writing prompt
Science expository essay
Experimental missing data
n = 10
Experimental collected data
n = 26
Comparison collected data
n = 34
Comparison missing data
n = 14
TREATMENT*
⊗= participants removed from the data analysis
POSTTESTING-DAY 2
Science compare and contrast writing prompt
Science expository essay
Experimental missing data posttest, but data collected at pretest
n = 5
Experimental collected data
n = 21
Comparison collected data
n = 28
Comparison missing data posttest, but data collected at pretest
n = 6
Treatment-on-the-Treated Analysis
n = 49
Intent-to-Treat Analysis
n = 60

APPENDIX V: RESEARCH QUESTION THREE-TREATMENT-ON-THE TREATED-DATA WITH OUTLIER CASES REMOVED

Research Question Three Treatment on the Treated Data with Outlier* Cases Removed for Science Expository Essay

Total Sample Size with 13 Outliers Removed *n* = 33
experimental *n* = 16 (5 cases removed)
comparison *n* = 20 (8 cases removed)
*outliers were above and below two standard deviations from the mean

Results of Assumptions Testing

Assumption	Test		Evidence	Assumptions Satisfied?
Independence	Matrix Scatterplots		No observable trends	Yes
Univariate Normality	Shapiro-Wilk		Not met for all (see Table 19)	No
	Boxplot/ Histogram/ Q-Q Plot		Relatively normal shape for all variables	Yes
	Skewness	Standardized Residuals	Met for all variables (all within value of absolute 2.0)	Yes
		DeCarlo (1997)	All were p < .05 except posttest agentless passive voice (*p* = .0332)	No
	Kurtosis	Standardized Residuals	Met for all variables except for posttest agentless passive voice comparison group (5.431) and noun phrase density comparison group (3.340)	No
		DeCarlo (1997)	All were p <.05 except for pre-test connectives (*p* =.0062), pretest (*p* = .0001) and posttest (*p* = .0054) agentless passive voice	No
	Scatterplot (Cook's vs. Unleveraged values)		Relatively normal shape for all variables	Yes
	Cook's Distance		All < 1.00	Yes
Multivariate Normality	Skewness	Small's test	χ^2 = 42.4471, *df* = 14.0000, *p* = .0001	No
		Srivastava's test	χ (b1p)=23.5710, *df* =14.0000, *p* = .0516	Yes
DeCarlo (1997)	Kurtosis	Small's variant	χ^2=13.3152, *df*=14.0000 , *p* = .5019	Yes
		Srivastava's test	χ = 3.3700, N(b2p) = 1.6956, *p* = .0900	Yes
		Mardia's test	b2p = 219.7918 , N(b2p) = -.5964, *p* = .5509	Yes
	Omnibus	Small's variant	χ^2= 55.7623, *df* = 28.0000 *p* = .0014	No
Linearity	Matrix Scatterplots		Straight positive linear shapes	Yes
Homogeneity of variances-covariances	Levene's Test		Not met for all (see Table 20)	No
	Spread-vs-level plots		Most boxplots have varying box and whisker lengths	No

Shapiro-Wilk Test of Univariate Normality for Treatment Groups for the Science Expository Essay for Treatment-on-the-Treated-Data with Outliers Removed

Dependent Variable	Group	S-W	df	Significance	Assumption Satisfied?
PreCONN	Experimental	.899	16	.077	Yes
	Comparison	.904	20	.050	Yes
PostCONN	Experimental	.939	16	.333	Yes
	Comparison	.909	20	.060	Yes
PreWBMC	Experimental	.946	16	.436	Yes
	Comparison	.946	20	.312	Yes
PostWBMC	Experimental	.864	16	.022*	No
	Comparison	.941	20	.249	Yes
PreAPV	Experimental	.784	16	.002*	No
	Comparison	.635	20	.000*	No
PostAPV	Experimental	.781	16	.002*	No
	Comparison	.496	20	.000*	No
PreVPD	Experimental	.895	16	.067	Yes
	Comparison	.951	20	.384	Yes
PostVPD	Experimental	.913	16	.132	Yes
	Comparison	.949	20	.354	Yes
PreNPD	Experimental	.718	16	.000*	No
	Comparison	.798	20	.001*	No
PostNPD	Experimental	.772	16	.001*	No
	Comparison	.923	20	.114	Yes
PrePPD	Experimental	.937	16	.311	Yes
	Comparison	.948	20	.336	Yes
PostPPD	Experimental	.942	16	.377	Yes
	Comparison	.960	20	.539	Yes
PreCIWS	Experimental	.939	16	.339	Yes
	Comparison	.858	20	.007*	No
PostCIWS	Experimental	.965	16	.754	Yes
	Comparison	.952	20	.404	Yes

Note. *Pre=pretest, Post=post-test, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences*

Levene's Test of Homogeneity of Variances for Independent Variables for the Science Expository Essay for Treatment-on-the Treated Data with Outliers Removed

Dependent variable	F	df1	df2	Sig.
PreCONN	2.151	4	31	.098
PostCONN	1.917	4	31	.133
PreWBMC	.711	4	31	.590
PostWBMC	3.161	4	31	.027*
PreAPV	3.842	4	31	.012*
PostAPV	3.535	4	31	.017*
PreVPD	1.703	4	31	.174
PostVPD	6.436	4	31	.001*
PreNPD	11.258	4	31	.000*
PostNPD	4.186	4	31	.008*
PrePPD	3.375	4	31	.021*
PostPPD	3.279	4	31	.024*
PreCIWS	2.338	4	31	.077
PostCIWS	2.286	4	31	.082

Note. Pre=pretest, Post=post-test, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences

Hierarchical Repeated Measures ANOVA Results-Treatment-on-the-Treated-Data with outliers removed

The results for the hierarchical repeated measures ANOVA suggested that there was not a statistically significant main effect (F_{test} = .529, *df* = 7, 25, *p* = .804)for the combined dependent variables from pretest to posttest . Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a medium effect and low power (partial η^2_{test} = .129, observed power = .185), as determined by Cohen (1988). Approximately 13% of the total variance of the combined dependent variables on the science expository essay can be accounted for by the within-subjects factor (i.e., time from pretest to posttest).

There was not a statistically significant interaction of treatment group and testing ($F_{treatment}$ = .200, *df* = 7, 25, *p* = .528). Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a large effect and low power (partial $\eta^2_{treatment}$ =.200, observed power = .303), as determined by Cohen (1988). Approximately 20% of the total variance of the combined dependent variables on the science expository measures can be accounted for by treatment group. The large effect indicated a large proportion of difference in scores between groups. The experimental group had higher marginal means on agentless passive voice (*M* = 2.302, *SE* = .874, *CI*= .519 to 4.085) than the comparison group (*M* = 1.392, *SE* =.747, *CI* = -.132 to 2.915); and correct versus incorrect word sequences (experimental *M* =71.986 , *SE* = 10.624, *CI* = 50.319 to 93.654), than the comparison group (*M* = 40.898, *SE* =9.077, *CI* = 22.386 to 59.410). The table below shows marginal mean, standard error, and confidence intervals for the dependent variables on the science expository essay for the treatment-on-the-treated data.

Marginal Means for Dependent Variables by Treatment Group for Science Expository Essay for Treatment-on-the-Treated Data with Outliers Removed

Variable	Treatment group	Mean	Standard Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
CONN	Experimental	87.136[a]	11.448	63.788	110.485
	Comparison	98.732[a]	9.781	78.784	118.680
WBMC	Experimental	2.868[a]	.462	1.927	3.809
	Comparison	3.030[a]	.394	2.226	3.835
APV	Experimental	2.302[a]	.874	.519	4.085
	Comparison	1.392[a]	.747	-.132	2.915
VPD	Experimental	186.942[a]	24.297	137.388	236.497
	Comparison	197.789[a]	20.759	155.452	240.127
NPD	Experimental	260.307[a]	34.758	189.419	331.196
	Comparison	316.661[a]	29.696	256.096	377.226
PPD	Experimental	51.874[a]	9.240	33.030	70.718
	Comparison	64.949[a]	7.894	48.849	81.049
CIWS	Experimental	71.986[a]	10.624	50.319	93.654
	Comparison	40.898[a]	9.077	22.386	59.410

[a] *Based on modified population marginal mean. Note. CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences*

In addition, there was no statistically significant interaction for testing and the between-subjects factor of the nested classrooms by teacher within treatment group ($F_{classroom}$= .376, *df* = 21, 81, *p* = .937). Multivariate partial eta squared for the effect of the between-subjects factor of the nested classrooms by teacher within treatment group indicated a small effect and low power (partial η^2_{test} =.125, observed power = .379), as determined by Cohen (1988). Approximately 13% of the total variance of the combined dependent variables on the science expository essay measures can be accounted for by the nested classrooms by teacher within treatment group. There was a small effect and notable differences in some of the marginal means. The experimental class one (*M* =80.273, *SE* =11.878, *CI* = 56.048 to 104.498) scored higher than the other classrooms on the marginal mean of the correct versus incorrect word sequences count. The other classes achieved the following means: experimental class 2 (*M* = 63.700, *SE* = 17.618, *CI* = 27.768 to 99.632), comparison class one (*M* = 37.944, *SE* =13.132, *CI* = 11.162 to 64.726), comparison class two (*M* =22.750, *SE* =16.083, *CI* = -10.051 to 55.551); and comparison class three (*M* = 62.000, *SE* =17.618, *CI* = 26.068 to 97.932). See table below for marginal means for correct versus incorrect word sequences by classroom for the science expository essay for treatment-on-the-treated data.

Marginal Means for Correct Versus Incorrect Word Sequences by Classroom for Science Expository Essay for Treatment-on-the-Treated Data with Outliers Removed

				95% Confidence Interval	
Variable	Classroom	Mean	Standard Error	Lower Bound	Upper Bound
CIWS	Experimental 1	80.273[a]	11.878	56.048	104.498
	Experimental 2	63.700[a]	17.618	27.768	99.632
	Comparison 1	37.944[a]	13.132	11.162	64.726
	Comparison 2	22.750[a]	16.083	-10.051	55.551
	Comparison 3	62.000[a]	17.618	26.068	97.932

[a] Based on modified population marginal mean. Note CIWS=correct versus incorrect word sequences

APPENDIX W: RESEARCH QUESTION THREE-ITT ANALYSIS -DATA WITH OUTLIER CASES REMOVED

Research Question Three ITT Analysis Data with Outlier* Cases Removed for Science Expository Essay

Total Sample Size with 16 Outliers Removed *n* = 44
experimental *n* = 17 (9 cases removed)
comparison *n* = 27 (7 cases removed)
*outliers were above and below two standard deviations from the mean

Results of Assumptions Testing

Assumption	Test		Evidence	Assumption Satisfied?
Independence	Matrix Scatterplots		No observable trends	Yes
Univariate Normality	Shapiro-Wilk		Not met for all variables (see Table 24)	No
	Boxplot/ Histogram/ Q-Q Plot		Relatively normal shape for all variables	Yes
	Skewness	Standardized Residuals	Met for all variables except for agentless passive voice pretest (2.475) and posttest (2.924)	No
		DeCarlo (1997)	All were p < .05 except for pretest and posttest agentless passive voice (*p* = .000) and pretest noun phrase density (*p* =.0261)	No
	Kurtosis	Standardized Residuals	Met for all variables except for agentless passive voice pretest (6.033) and posttest (9.135)	No
		DeCarlo (1997)	All were p <.05 except for pretest (*p* = .0012) and posttest (*p* = .0097) agentless passive voice	No
	Scatterplot (Cook's vs. Unleveraged values)		Relatively normal shape for all variables	Yes
	Cook's Distance		All < 1.00	Yes
Multivariate Normality	Skewness	Small's test	χ^2 = 59.4369, *df* = 14.0000, *p* = .0000	No
		Srivastava's test	χ (b1p) = 36.5250, *df* =14.0000, *p* =.0009	No
DeCarlo (1997)	Kurtosis	Small's test	χ^2=30.0695, *df*=14.0000 , *p* = .0075	No
		Srivastava's test	χ = 3.5538, N(b2p) = 2.8055, *p* = .0050	No
		Mardia's test	b2p = 231.6574, N(b2p) = 1.1999, *p* = .2302	Yes
	Omnibus	Small's test	χ^2= 89.5063, *df* = 28.0000 *p* = .0000	No
Linearity	Matrix Scatterplots		Straight positive linear shapes	Yes
	Levene's Test		Not met (see Table 25)	No
	Spread-vs-level plots		Most boxplots have varying box and whisker lengths	No

Shapiro-Wilk Test of Univariate Normality for Treatment Groups for the Science Expository Essay for Data Utilizing ITT Analysis with Outliers Removed

Dependent Variable	Group	S-W	df	Sig	Assumption Satisfied?
PreCONN	Experimental	.899	17	.064	Yes
	Comparison	.890	27	.008*	No
PostCONN	Experimental	.965	17	.734	Yes
	Comparison	.901	27	.014	No
PreWBMC	Experimental	.958	17	.602	Yes
	Comparison	.907	27	.019*	No
PostWBMC	Experimental	.962	17	.671	Yes
	Comparison	.948	27	.194	Yes
PreAPV	Experimental	.718	17	.000*	No
	Comparison	.568	27	.000*	No
PostAPV	Experimental	.766	17	.001*	No
	Comparison	.587	27	.000*	No
PreVPD	Experimental	.948	17	.433	Yes
	Comparison	.935	27	.090	Yes
PostVPD	Experimental	.972	17	.856	Yes
	Comparison	.936	27	.099	Yes
PreNPD	Experimental	.961	17	.647	Yes
	Comparison	.861	27	.002*	No
PostNPD	Experimental	.943	17	.361	No
	Comparison	.926	27	.055	Yes
PrePPD	Experimental	.963	17	.696	Yes
	Comparison	.934	27	.085	Yes
PostPPD	Experimental	.969	17	.796	Yes
	Comparison	.941	27	.126	Yes
PreCIWS	Experimental	.930	17	.216	Yes
	Comparison	.931	27	.072	Yes
PostCIWS	Experimental	.976	17	.912	Yes
	Comparison	.958	27	.336	Yes

Note. *Pre=pretest, Post=posttest, CONN=sentence connectives, WBMC=words before the main clause, APV = agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences*

Levene's Test of Homogeneity of Variances for Independent Variables for the Science Expository Essay for Data Utilizing ITT Analysis with Outliers Removed

Dependent variable	F	df1	df2	Sig.
PreCONN	9.155	4	39	.000*
PostCONN	2.929	4	39	.033*
PreWBMC	.347	4	39	.844
PostWBMC	.879	4	39	.485
PreAPV	2.149	4	39	.093
PostAPV	5.425	4	39	.001*
PreVPD	3.638	4	39	.013*
PostVPD	7.612	4	39	.000*
PreNPD	15.025	4	39	.000*
PostNPD	14.598	4	39	.000*
PrePPD	4.077	4	39	.007*
PostPPD	4.098	4	39	.007*
PreCIWS	3.145	4	39	.025*
PostCIWS	2.147	4	39	.093

Note. *Pre=pretest, Post=post-test, CONN=sentence connectives, WBMC=words before the main clause, APV=agentless passive voice, VPD=verb phrase density, NDP=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences*

Hierarchical Repeated Measures ANOVA Results-ITT Data with outliers removed

The results for the hierarchical repeated measures ANOVA suggest that there was not a statistically significant main effect (F_{test} = .615, *df* = 7, 33, *p* = .739) for the combined dependent variables from pretest to posttest . Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a medium effect and low power (partial η^2_{test} = .115, observed power =.224), as determined by Cohen (1988). Approximately 12% of the total variance of the combined dependent variables on the science expository essay can be accounted for by the within-subjects factor (i.e. time from pretest to posttest).

There was not a statistically significant interaction of treatment group and testing ($F_{treatment}$ = .868, *df* = 7, 33, *p* = .543). Multivariate partial eta squared for the effect of the within-subjects factor of test indicated a large effect and low power (partial $\eta^2_{treatment}$ =.155, observed power = .313), as determined by Cohen (1988). Approximately 16% of the total variance of the combined dependent variables on the science expository essay measures can be accounted for by treatment group. The effect was large, which indicated that there was a large proportion of difference in scores between the groups. The experimental group score on the word sequences was higher than the comparison group on the marginal mean for sentence connectives (experimental *M* = 107.033, *SE* =9.198, *CI* = 88.428 to 125.637; comparison *M* =99.040, *SE* =7.457, *CI* = 83.956 to 114.123); words before the main clause (experimental *M* = 3.811, *SE*

= .421, *CI* = 2.959 to 4.662; comparison *M* =2.914, *SE* =.341, 2.223 to 3.604); agentless passive voice (experimental *M* =2.296, *SE* =.814, *CI* = l.649 to 3.943; comparison *M* =1.589, *SE* =.660, *CI* = .254 to 2.924); verb phrases density (experimental *M* =245.825, *SE* = 20.287, *CI* = 204.792 to 286.859; comparison *M* =198.169, *SE* = 16.447, *CI* = 164.901 to 231.437); noun phrase density (experimental *M* =341.802, *SE* = 29.852, *CI* =281.421 to 402.183; comparison *M* =313.647, *SE* = 24.203, *CI* = 264.692 to 362.601); and correct versus incorrect word sequences (experimental *M* =69.163, *SE* = 9.994, *CI* =48.948 to 89.378; comparison *M* =40.984, *SE* = 8.103, *CI* =24.595 to 57.374). The table below shows marginal means, standard error, and confidence intervals for the dependent variables on the science expository essay for the ITT data.

Marginal Means and Standard Error for Treatment Groups for Pretest and Posttest of the Science Expository Essay for Data Utilizing ITT Analysis with Outliers Removed

			Standard	95% Confidence Interval	
Dependent variable	Treatment group	Mean	Error	Lower Bound	Upper Bound
CONN	Experimental	107.033[a]	9.198	88.428	125.637
	Comparison	99.040[a]	7.457	83.956	114.123
WBMC	Experimental	3.811[a]	.421	2.959	4.662
	Comparison	2.914[a]	.341	2.223	3.604
APV	Experimental	2.296[a]	.814	.649	3.943
	Comparison	1.589[a]	.660	.254	2.924
VPD	Experimental	245.825[a]	20.287	204.792	286.859
	Comparison	198.169[a]	16.447	164.901	231.437
NPD	Experimental	341.802[a]	29.852	281.421	402.183
	Comparison	313.647[a]	24.203	264.692	362.601
PPD	Experimental	61.186[a]	9.761	41.443	80.929
	Comparison	65.530[a]	7.913	49.524	81.537
CIWS	Experimental	69.163[a]	9.994	48.948	89.378
	Comparison	40.984[a]	8.103	24.595	57.374

[a] Based on modified population marginal mean. Note. CONN=sentence connectives, WBMC=words before the main clause, APV agentless passive voice, VPD=verb phrase density, NPD=noun phrase density, PPD=prepositional phrase density, CIWS=correct versus incorrect word sequences

In addition, there was no significant interaction for testing and the between-subjects factor of the nested classrooms by teacher within treatment group $F_{classroom}$= .376, *df* = 21, 81, *p* = .937). Multivariate partial eta squared for the effect of the between-subjects factor of the nested classrooms by teacher within treatment group indicated a medium effect and moderate power (partial η^2_{test} = .127, observed power = .528), as determined by Cohen (1988). Approximately 13% of the total variance of the combined dependent variables on the science expository essay measures can be accounted for by the nested classrooms by teacher within treatment group. The medium effect indicated a medium proportion of difference in scores between the groups. The experimental class one (*M* =81.409, *SE* =11.875, *CI* = 57.390 to 105.428) scored higher than the other classrooms on the marginal mean of the correct versus incorrect word sequences count. The other classes scored as follows: experimental class two *(M* =56.917, *SE* =16.079, *CI* = 24.395 to 89.439); comparison class one (*M* =36.036, *SE* =10.526, *CI* = 14.745 to 57.326), comparison class two (*M* = 19.500, *SE* = 14.886, *CI* = -10.610 to 49.610); and comparison class three (*M* =67.417, *SE* =16.079, *CI* = 34.895 to 99.939). See the table below for marginal means for correct versus incorrect word sequences by classroom for science expository essay for the ITT analysis data with outliers removed.

Marginal Means for Correct Versus Incorrect Word sequences by Classroom for Science Expository Essay for Data Utilizing ITT Analysis with Outliers Removed

				95% Confidence Interval	
Variable	Classroom	Mean	Standard Error	Lower Bound	Upper Bound
CIWS	Experimental 1	81.409[a]	11.875	57.390	105.428
	Experimental 2	56.917[a]	16.079	24.395	89.439
	Comparison 1	36.036[a]	10.526	14.745	57.326
	Comparison 2	19.500[a]	14.886	-10.610	49.610
	Comparison 3	67.417[a]	16.079	34.895	99.939

[a] Based on modified population marginal mean. Note CIWS=correct versus incorrect versus incorrect word sequences

REFERENCES

Achieve, Inc. (2005). *Rising to the challenge: Are high school graduates prepared for college and work?* Washington, DC: Author.

Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an inquiry-based approach known as the Science Writing Heuristic to traditional science teaching practices: Are there differences? *International Journal of Science Education*, *29*(14), 1745-1765.

American College Test (ACT), Inc. (2000). *Workplace essential skills. Resources related to the SCANS competencies and foundational skills*. Iowa City, IA: Author.

ACT, Inc. (2005). *Crisis at the core. Preparing all students for college and work.* Iowa City, IA: Author.

ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City, IA: Author.

ACT, Inc. (2007). *Aligning postsecondary expectations and high school practice: The gap defined.* Iowa City, IA: Author.

ACT, Inc. (2009). *Focusing on the essentials for college and career readiness. Policy implications of the ACT national curriculum survey results 2009*. Iowa City, IA: Author.

ACT, Inc. (2013a). *ACT National Curriculum Survey 2012. Policy implications on preparing for higher standards.* Iowa City, IA: Author.

ACT, Inc. (2013b). *The condition of college and career readiness.* Iowa City: Author.

Alvermann, D. E. (2002). Effective literacy instruction for adolescents. *Journal of Literacy Research*, *34*(2), 189-208.

Amato, J. M., & Watkins, M. W. (2009). The predictive validity of CBM writing indices for eighth-grade students. *The Journal of Special Education*, 44(4) 195-204. doi: 10.1177/0022466909333516

American Diploma Project. (2004). Ready *or not. Creating a diploma that counts.* Washington DC: Achieve, Inc.

American Management Association. (2012). AMA *2012 Critical skills survey*. Retrieved from http://playbook.amanet.org/wp-content/uploads/2013/03/2012-Critical-Skills-Survey-pdf.pdf

American Speech-Language Hearing Association. (2010). *Roles and responsibilities of speech-language pathologists in schools* [Professional Issues Statement]. Retrieved April 11, 2015 from www.asha.org/policy

American Speech-Language-Hearing Association. (2001). *Roles and responsibilities of speech-language pathologists with respect to reading and writing in children and adolescents* [Guidelines]. doi: 10.1044/policy.GL2001-00062

Anderson, R. C. and others. (1985). *Becoming a nation of readers.* (Report of the Commission on Reading). Champaign, IL: University of Illinois.

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., & Zhu, D. (2006). The effect of grammar teaching on writing development. *British Educational Research Journal*, *32*(1), 39-55.

Applebee, A. N. (1977). ERIC RCS: Writing and Reading. *Journal of Reading*, *20*(6)534-537.

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research, 54*(4), 577–596.

Applebee, A. N. (2000). Alternative models of writing development. In R. Indirino, J. R. Squire, (Eds.), *Writing: Research/theory/practice* (pp. 90-110). Newark: International Reading Association.

Applebee, A. N., & Langer, J. A. (2009). EJ Extra: What is happening in the teaching of writing? *English Journal*, *98*(5), 18-28.

Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle and high schools. *English Journal, 100* (6), 14–27.

Archer, A. L., & Hughes, C. A. (2004). *Explicit instruction: Effective and efficient teaching*. New York, NY: Guilford Press.

Armbruster, B. B., Echols, C. H., & Brown, A. L. (1983). *The role of metacognition in reading to learn: A developmental perspective* (Reading Education Report No. 40). Urbana, IL: Center for the Study of Reading.

Armijo-Olivo, S., Warren, S., & Magee, D. (2009). Intention to treat analysis, compliance, dropouts and how to deal with missing data in clinical research: a review. *Physical Therapy Reviews*, *14*(1), 36-49.

Arndt, K. B., & Schuele, C. M. (2013). Multiclausal utterances aren't just for big kids: A framework for analysis of complex syntax production in spoken language of preschool- and early school-age children. *Topics in Language Disorders*, *33*(2), 125-139.

Balgopal, M. M., & Wallace, A.M. (2009). "Decisions and dilemmas: Using writing to learn activities to increase ecological Literacy." *The Journal of Environmental Education, 40* (3), 13-26.

Balthazar, C., & Scott, C. (in press). The place of syntax in language assessment and intervention. In T. Ukrainetz (Ed.), *School-age language intervention: Principles and procedures for oral and written language success.* Austin, TX: Pro-Ed.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29-58.

Basturkmen, H., Loewen, S., & Ellis, R. (2002). Metalanguage in focus on form in the communicative classroom. *Language Awareness*, *11*(1), 1-13.

Bazerman, C., Little, J., Bethel, L., Chavkin, T., Fouquette, D., & Garufis, J. (2005). *Writing across the curriculum. A reference guide to rhetoric and composition.* West Lafayette, Ind: Parlor Press, WAC Clearinghouse.

Beaufort, A. (2006). Writing in the professions. In P. Smagorinsky (Ed.), *Research on composition: Multiple perspectives on two decades of change* (pp. 217–242). New York, NY: Teachers College Press.

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

Berninger, V. W. (2000). Development of language by hand and its connections with language by ear, mouth, and eye. *Topics in Language Disorders*, *20*(4), 65-84.

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology*, *102*(3), 635.

Berninger, V. W., Abbott, R. D., Jones, J., Wolf, B. J., Gould, L., Anderson-Youngstrom, M., ... & Apel, K. (2006). Early development of language by hand: Composing, reading, listening, and speaking connections; three letter-writing modes; and fast mapping in spelling. *Developmental Neuropsychology*, *29*(1), 61-92.

Berninger, V. W., Abbott, R. D., Swanson, H. L., Lovitt, D., Trivedi, P., Lin, S. J. C., ... & Amtmann, D. (2010). Relationship of word-and sentence-level working memory to reading and writing in second, fourth, and sixth grade. *Language, Speech, and Hearing Services in Schools*, *41*(2), 179-193.

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology*, *102*(3), 635-651. doi:10.1037/a0019319.

Berninger, V. W., & Winn, W. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In MacArthur, C. A., Graham, S., & Fitzgerald, J. (Eds.), *Handbook of writing research* (pp. 96-114). New York, NY: Guilford Press.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development*, 498-510.

Biancarosa, G., & Snow, C. E. (2006). *Reading Next: A vision for action and research in middle and high school literacy.* (Report from Carnegie Corporation of New York, Alliance for Excellent Education, 2nd ed.). New York, NY: Carnegie Corporation.

Biber, D. (1988). *Variation across speech and language.* Cambridge, UK: Cambridge University Press.

Bill & Melinda Gates Foundation. (2008). *All students ready for college, career, and life: Reflection of the education investments 2000-2008.* Retrieved from https://docs.gates foundation.org/Documents/reflections-foundations-education-investments.pdf

Billman, A., & Pearson, P. (2013). Literacy in the disciplines. *Literacy Learning: The Middle Years*, *21*(1), 25.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, *14*(3), 191-205.

Bloom, M. R., & Lafleur, B. (1999). *Turning skills into profit: Economic benefits of workplace education programs.* New York, NY: The Conference Board Inc.

Blosser, J., Roth, F. P., Paul, D. R., Ehren, B. J., Nelson, N.W., Sturm, J. M. (2012). Integrating the core. *The ASHA Leader*, 17, 12-15. doi:10.1044/leader.FTR1.17102012.12

Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry, 71*(1), 52-63.

Brea-Spahn, M. (2014). Bilingual children with language learning disabilities: Convergence in conceptual, linguistic, and cultural circles of knowledge. In C. A. Stone, E. R. Silliman,

B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders (*2nd ed., pp. 359-377). New York: NY, Guilford Press.

Breaux, K. C., & Frey, F. E. (2009). Assessing writing skills using correct-incorrect word sequences: A national study. Retrieved from http://images. pearsonclinical.com/images/ products/wiat-iii/WIAT-III_NASP_Poster.pdf

Britton, J. (1980). *Language and learning across the curriculum.* Retrieved from http://comppile.org/archives/fforum/fforum2%282%29files/fforum2%282%29Britton.pdf

Britton, J., Burgess, A., Martin, N., McLeod, A., & Rosen, H. (1975). T*he development of writing abilities (11–18): Schools Council Project on written language of 11–18 year olds.* London, England: Macmillan

Brown, B. A., Reveles, J. M., & Kelly, G. J. (2005). Scientific literacy and discursive identity: A theoretical framework for understanding science learning. *Science Education*, *89*(5), 779-802.

Brozo, W. G., Moorman, G., Meyer, C., & Stewart, T. (2013). Content area reading and disciplinary literacy: A case for the radical center. *Journal of Adolescent & Adult Literacy*, *56*(5), 353-357.

Bullock, S. (2006). Building concepts through writing-to-learn in college physics classrooms. *Ontario Action Researcher*, *9*(2), 1-8.

Burkhalter, Nancy. (1996). Assessing grammar teaching methods using a metacognitive framework. *Journal of Teaching Writing*, *15*(2), 259-283.

Business Roundtable. (2009). *Getting ahead—staying ahead: Helping America's workforce succeed in the 21st century*. Retrieved from http://businessroundtable.org/sites/default/ files/ Getting_Ahead_Executive_Summary_1.pdf

Bybee, B. (2010). Advancing STEM Education: A 2020 vision. *Technology and Engineering Teacher*, *70*(1), 30-35.

Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching*, *43*(4), 349-352.

Cain, K. (2007). Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics*, *28*(4), 679-694.

Cain, K., & Oakhill, J. (2007). Reading comprehension difficulties: Correlates, causes, and consequences. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41–74). New York, NY: Guilford Press.

Cairns, H. S., Schlisselberg, G., Waltzman, D., & McDaniel, D. (2006). Development of a metalinguistic skill judging the grammaticality of sentences. *Communication Disorders Quarterly*, *27*(4), 213-220.

Callan, P. M., Finney, J. E., Kirst, M. W., Usdan, M. D., & Venezia, A. (2006). *Claiming common ground: State policymaking for improving college readiness and success.* (National Center Report No. 06-1). San Jose, CA: National Center for Public Policy and Higher Education.

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand-McNally.

Carnevale, A. P., Smith, N., Melton, M. (2011). *Science, Technology, Engineering, Mathematics.* Retrieved from cew.georgetown.edu/STEM

Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York, NY: Carnegie Corporation of New York.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century US workforce*. (Executive Summary). U.S.A.: The Conference Board, Corporate Voices for Working Families, Partnership for 21st Century Skills and Society for Human Resource Management.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Ellis-Weismer (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech*, *Language*, *and Hearing Research*, *48*(6), 1378-1396.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1142-1157.

Cazden, C. B. (1974). Play and metalinguistic awareness: One dimension of language experience. *The Urban Review*, *7*(1), 28-39.

Cazden, C. B. (1976). Play with language and meta-linguistic awareness. In J. Bruner, A. Jolly, & K. Sylva (Eds.), *Play: Its role in development and evolution*. New York, NY: Basic Books.

Center for Workforce Preparation. (2002). *Higher skills, bottom-line results: A chamber guide to improving workplace literacy.* Washington, DC: U.S. Chamber of Commerce.

Cervetti, G. N., Barber, J., Dorph, R., Pearson, P. D., & Goldschmidt, P. G. (2012). The impact of an integrated approach to science and literacy in elementary school classrooms. *Journal of Research in Science Teaching*, 49(5), 631-658.

Cervetti, G., & Pearson, P. D. (2012). Reading, writing, and thinking like a scientist. *Journal of Adolescent & Adult Literacy*, *55*(7), 580-586.

Change the Equation. (2011). *Improving teaching and learning in science, technology, engineering, and mathematics (STEM).* (Vital Signs Report*)*. Retrieved from http://www.changethe equation.org/sites/default/files/CTEqVitalSigns2011 National O.pdf

Chen, H., & Jones, P. (2013). Understanding metalinguistic development in beginning writers: A functional perspective. *Journal of Applied Linguistics and Professional Practice*, *9*(1), 81-104.

Chen, X., & Weko, T.(2009). *Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education.* Stats in Brief. NCES 2009-161. Retrieved from http://nces.ed.gov/pubs2009/2009161.pdf

Chen, Y. C., Park, S., Hand, B. (2013). Constructing and critiquing arguments. *Science and Children, 50*(5), 40-45.

Chomsky, N. (1965). *Aspects of a theory of syntax.* Cambridge, MA: M. I. T Press.

Coker, D., & Lewis, W. E. (2008). Beyond writing next: A discussion of writing research and instructional uncertainty. *Harvard Educational Review*, *78*(1), 231-251.

College Board. (2014). *2014 College Board program results: SAT.* Retrieved on January 26, 2014 from https://www.collegeboard.org/program-results/2014/sat

College Board. (2015). *SpringBoard English Language Arts.* Retrieved on February 1, 2015 from http://springboardprogram.collegeboard.org/english-language-arts/

Combs, W. E. (1976). Further effects of sentence-combining practice on writing ability. *Research in the Teaching of English*, 137-149.

Combs, W. E. (1977). Sentence-combining practice aids reading comprehension. *Journal of Reading, 21*(1), 18-24.

Comings, J., Reder, S., Sum, A. (2001). *Building a level playing field. The need to expand and improve the national and state adult education and literacy systems.* (National Center for the Study of Adult Learning and Literacy Occasional Paper). Cambridge, MA: Harvard University Graduate School of Education.

Committee on Conceptual Framework for the New K-12 Science Education Standards. (2012). *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: National Academies Press.

Committee on STEM Education National Science and Technology Council. (2013). *Federal science, technology, engineering, and mathematics (STEM) education. 5-year strategic plan*. (Report). Washington DC: Office of the Executive Council.

Connors, R. J. (2000). The erasure of the sentence. *College Composition and Communication, 52*(1), 96-128.

Crawford, C. C., & Royer, M. M. (1935). Oral drill versus grammar study. *The Elementary School Journal*, *36*(2), 116-119.

Datchuk, S. M., & Kubina, R. M. (2013). A review of teaching sentence-level writing skills to students with writing difficulties and learning disabilities. *Remedial and Special Education, 34*(3), 180-192.

DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, *37*(6), 582-601.

De Carlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*(3), 292.

Deno, S. L. (1985). Curriculum-based measurement: the emerging alternative. *Exceptional Children*, *52*(3), 219-232.

Deshler, D. D. Alley, G. R., Warner, M. M., & Schumaker, J. B. (1981). Instructional practices for promoting skill acquisition and generalization in severely learning disabled adolescents. *Learning Disability Quarterly, 4* (4), pp. 415-421

Deshler, D. D., & Lenz, B. K. (1989). The Strategies Instructional Approach. *International Journal of Disability, Development and Education, 36* (3), 203-224, doi:10.1080/01566 55893603004

Dispezio, M. A., Frank, M., Heithaus, M. R., & Ogle, D. M. (2012). *Florida Science Fusion Text Book McDougal.* Orlando, FL: Houghton Mifflin Publishing Company.

Dockrell, J. E. (2014). Developmental Variations in the production of written text: Challenges for students who struggle with writing. In C. A. Stone, E. R. Silliman, B .J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders* (2nd ed., pp. 505-523). New York, NY: Guilford Press.

Dodge, J. (2009). *25 Quick formative assessments for a differentiated classroom: Easy, low-prep assessments that help you pinpoint students' needs and reach all learners*. New York: Scholastic Inc.

Dow, P. (1997). Sputnik revisited: Historical perspectives on science reform. In *A symposium reflecting on Sputnik: Linking the past, present, and future of educational reform.* Washington, DC: National Academy of Sciences.

Dugger, W. E. (2010). *Evolution of STEM in the United States.* A paper presented at the 6th Biennial international conference on technology education research. Abstract retrieved from http://www.iteea.org/Resources/PressRoom/AustraliaPaper.pdf

Ebbels, S. H., & van der Lely, H. (2001). Meta-syntactic therapy using visual coding for children with severe persistent SLI. *International Journal of Language and Communication Disorders, 36*, 345–350.

Ebbels, S. H., van der Lely, H. K., & Dockrell, J. E. (2007). Intervention for verb argument structure in children with persistent SLI: A randomized control trial. *Journal of Speech, Language, and Hearing Research, 50*(5), 1330-1349.

Eberhardt, N. C. (2013). Syntax: Somewhere between words and text. *Perspectives on Language and Literacy*, *39*(3), 43-48.

Edmonds, W. A., & Kennedy, T. D. (2013). *An applied reference guide to research designs: Quantitative, qualitative, and mixed methods*. Thousand Oaks, CA: Sage.

Ehren, B. J. (2009). Looking through an adolescent literacy lens at the narrow view of reading. *Language, Speech, and Hearing Services in Schools*, *40*(2), 192-195.

Ehren, B. J. (2000). Maintaining a therapeutic focus and sharing responsibility for student success: Keys to in-classroom speech language services. *Language, Speech, and Hearing Services in Schools*, *31*, 219–229.

Ehren, B. J., & Ehren, T. C. (2001). New or expanded literacy roles for speech-language pathologists: Making it happen in the schools. *Seminars in Speech and Language*, 22, 233–243.

Ehren, B. J., Lenz, B. K. B. H, Deshler, D. (2014). Adolescents who struggle with 21st-century literacy. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of* *language and literacy: Development and disorders* (2nd ed., pp. 619-636). New York, NY: Guilford Press.

Ehren, B. J., Murza, K. A. (2010). The urgent need to address workforce readiness in adolescent literacy intervention. *Perspectives on Language Learning and Education 17*, 93-99. doi:10.1044/lle17.3.93

Ehren, B. J., Murza, K. A., & Malani, M. D. (2012). Disciplinary literacy from a speech–language pathologist's perspective. *Topics in Language Disorders, 32*(1), 85-98.

Ehren, B. J., & Whitmire, K. (2009). Speech-language pathologists as primary contributors to response to intervention at the secondary level. *Seminars in Speech and Language, 30* (2), 90-104.

Eisenberg, S. L. (2013). Grammar intervention: Content and procedures for facilitating children's language development. *Topics in Language Disorders*, *33*(2), 165-178.

Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus-on-form. *System*, *30*(4), 419-432.

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication (28)*, 2 122-128.

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.

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

Espin, C., Wallace, T., Campbell, H., Lembke, E. S., Long, J.D., and 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.

Evans, R., Venetozzi, R., Bundrick, M., & McWilliams, E. (1988). The effects of sentence-combining instructions on writing and on standardized test scores. *The Journal of Educational Research*, 53-57.

Faggella-Luby, M. N., Graner, P. S., Deschler, D. D., & Drew, S. V. (2012). Building a house on sand: Why disciplinary literacy is not sufficient to replace general strategies for adolescent learners who struggle. *Topics in Language Disorders, 32*(1), 69-84.

Fang, Z. (2005). Scientific literacy: A systemic functional linguistics perspective. *Science Education*, *89*(2), 335-347.

Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education*, 28(5), 491-520.

Fang, Z. (2008). Going beyond the fab five: Helping students cope with the unique linguistic challenges of expository reading in intermediate grades. *Journal of Adolescent & Adult Literacy, 51(*6), 476-487.

Fang, Z. (2012a). The challenges of reading disciplinary texts. In T. L. Jetton & C. Shanahan (Eds.), *Adolescent literacy in the academic disciplines. General principles and practical strategies* [Kindle version]. Retrieved from Amazon.com

Fang, Z. (2012b). Language correlates of disciplinary literacy. *Topics in Language Disorders, 32*(1), 19-34.

Fang, Z., & Schleppegrell, M. J. (2008). *Reading in secondary content areas: A language-based pedagogy*. Ann Arbor, MI: University of Michigan Press.

Fang, Z., & Schleppegrell, M. J. (2010). Disciplinary literacies across content areas: Supporting secondary reading through functional language analysis. *Journal of Adolescent & Adult Literacy, 53*(7), 587-597.

Fang, Z., Schleppegrell, M. J., & Cox, B. E. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research, 38*(3), 247-273.

Fang, Z., Schleppegrell, M. J., & Moore, J. (2014). The linguistic challenges of learning across academic disciplines. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders (*2nd ed., pp. 545-568). New York, NY: Guilford Press.

Farrall, M. L. (2013). The assessment of written syntax. *Perspectives on Language and Literacy*, *39*(3), 31-37.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1301-1318.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Fey, M. E., Long, S. H., & Finestack, L. H. (2003). Ten principles of grammar facilitation for children with specific language impairments. *American Journal of Speech-Language Pathology*, *12*(1), 3-15.

Finestack, L. H. (2014). Language learning of children with typical development using a deductive metalinguistic procedure. *Journal of Speech, Language, and Hearing Research*, *57*(2), 509-523.

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist, 35*(1), 39-50.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., and Shelley, B. E. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context* (NCES Report No. 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Flood, J., & Salus, M. W. (1982). Metalinguistic awareness: its role in language development and its assessment. *Topics in Language Disorders*, *2*(4), 56-64.

Florida Department of Education. (2014). *Understanding FCAT 2.0 assessment tests.* Tallahassee, FL: Author.

Florida Department of Education. (2015). *PK-12 Public school data publications and reports.* Retrieved on May 18, 2015 from http://www.fldoe.org/accountability/data-sys/edu-info-accountability-services/pk-12-public-school-data-pubs-reports/students.stml

Florida Department of Education (2014) *Next Generation Sunshine State Standards-Grade 8 Science.* Retrieved on May 5, 2015 from http://www.cpalms.org/Public/Search/CriteriaSearch?search=Standards&subjectId=29&gradeId=13

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4) 365-387.

Fromkin, V., Rodman, R., & Hyams, N. (2014). *An introduction to language* (10th ed.). Boston, MA: Wadsworth.

Fry, S. W., & Villagomez, A. (2012). Writing to learn: benefits and limitations. *College Teaching*, *60*(4), 170-175.

Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education*, *92*(5), 799-824.

Fusaro, J. A. (1993). A meta-analysis of the effect of sentence combining on reading comprehension. *Reading Improvement*, *30*(4), 228-31.

Gall, M. D., Gall, J. P., & Borg, W. R. (2006). *Educational research: An introduction* (8th ed.). Boston, MA: Allyn & Bacon.

Gardner, D. P., Larsen, Y. W., Baker, W. O., Campbell, A., Crosby, E. A., Foster C. A., Jr., ...Kirk, A. Y. (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people.* (Report to the nation and the Secretary of Education). Washington DC: Government Printing Office.

Garrison, C., & Ehringhaus, M. (2013). *Formative and Summative Assessments. Retrieved from* National Middle School Association and Measured Progress. Retrieved from http://schools.nyc. gov/NR/rdonlyres /33148188-6FB5-4593-A8DF8EAB8CA002AA /0/2010_11_Formative_Summative_Assessment.pdf

Gast, D. L., & Ledford, J. (Eds.). (2010). *Single-subject research methodology in behavioral sciences.* New York, NY: Routledge.

Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, *45*(8), 922-939.

Gilbert, J., & Graham, S. (2010). Teaching writing to elementary students in grades 4–6: A national survey. *The Elementary School Journal*, *110*(4), 494-518.

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

Gillespie, A., Graham, S., Kiuhara, S., & Hebert, M. (2014). High school teachers' use of writing to support students' learning: a national survey. *Reading and Writing*, *27,* 1043-1072.doi10.1007/s11145-013-9494-8

Gillam, R. B., & Johnston, J. R. (1992). Spoken and written language relationships in language/learning-impaired and normally achieving school-age children. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1303-1315.

Gersten, R., & Baker, S. (2001). Teaching expressive writing to students with learning disabilities: A meta-analysis. *The Elementary School Journal*, *101*(3), 251-272.

Gombert, J. E. (1992). *Metalinguistic development*. Chicago, IL: University of Chicago Press.

Gombert, J. E. (1993). Metacognition, metalanguage and metapragmatics. *International Journal of Psychology*, *28*(5), 571-580.

Gonzalez, H. B., & Kuenzi, J. J. (2012). S*cience, Technology, Engineering, and Mathematics (STEM) Education: A prime.* (CRS Report for Congress R42642). Retrieved from http://digital.library.unt.edu/ark:/67531/metadc122233/m1/1/high_res_d/R42642_2012Aug01.pdf

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*(2), 371-398.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193-202.

Graham, S. (2006). Writing. In P. L. Alexander & P. Winne (Eds.) *Handbook of educational psychology* (2nd ed., pp. 457-478). Mahwah, NJ: Erlbaum.

Graham, S. (2006). Strategy Instruction and the teaching of writing. In MacArthur, C. A., Graham, S., & Fitzgerald, J. (Eds.), *Handbook of writing research* (pp. 187-207). New York, NY: Guilford Press.

Graham, S. (2013). *It all starts here. Fixing our national writing crisis from the foundation.* (Reprinted white paper). Columbus, OH: Sapperstein Associates.

Graham, S., Capizzi, A., Harris, K. R., Hebert, M., & Morphy, P. (2014a). Teaching writing to middle school students: A national survey. *Reading and Writing, 27*, 1015-1042. doi: 10.1007/s11145-013-9495-7

Graham, S., Early, J., Wilcox, K. (2014b). Adolescent writing and writing instruction: introduction to the special issue. *Reading and Writing, 27*, 969-972. doi:10.1007/s11145-014-9497-0

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: importance, development, and instruction. *Reading and Writing, 26*(1), 1-15.

Graham, S., & Harris, K. R. (2011). *Informing writing: The benefits of formative assessment.* (Report from Carnegie Corporation of New York). New York, NY: Carnegie Corporation.

Graham, S., & Harris, K. R. (2013). Common Core State Standards, writing, and students with LD: Recommendations. *Learning Disabilities Research & Practice*, *28*(1), 28-37.

Graham, S., Harris, K. R., Fink-Chorzempa, B., & MacArthur, C. (2003). Primary grade teachers’ instructional adaptations for struggling writers: A national survey. *Journal of Educational Psychology, 95*(2), 279–292. doi:10.1037/0022-0663.95.2.279

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve reading.* (Report from Carnegie Corporation of New York). New York, NY: Carnegie Corporation.

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

Graham, S., Harris, K., & Hebert, M. (2011). *Informing Writing: The benefits of formative assessment (*Report from Carnegie Corporation of New York). New York, NY: Carnegie Corporation.

Graham, S., Harris, K. R., & Larsen, L. (2001). Prevention and intervention of writing difficulties for students with learning disabilities. *Learning Disabilities Research & Practice*, *16(*2), 74-84.

Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445.

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

Greenleaf, C., Hanson, T., Herman, J., Litman, C., Rosen, R., Schneider, S., & Silver, D. (2011a). *Grant title: A study of the efficacy of reading apprenticeship professional development for high school history and science teaching and learning.* (Final Report to Institute for Education Sciences National Center for Education Research Teacher Quality/Reading and Writing). Retrieved from http://readingapprenticeship.org/wp content/uploads/2013/12/IES-history-biology-final-report1.pdf

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., ... & Jones, B. (2011b). Integrating literacy and science in biology teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal*, *48*(3), 647-717.

Gummersall, D. M., & Strong, C. J. (1999). Assessment of complex sentence production in a narrative context. *Language, Speech, and Hearing Services in Schools, 30*(2), 152-164.

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*(3), 109–112. doi:10.4103/2229-3485.83221

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, *8*(3), 381-391.

Guthrie, J. T., Anderson, E., Alao, S., & Rinehart, J. (1999). Influences of concept-oriented reading instruction on strategy use and conceptual learning from text. *The Elementary School Journal*, *99*(4), 343-366.

Hadjioannou, X., & Hutchinson, M. C. (2010). Putting the" G" back in English: Preparing pre-service teachers to teach grammar. *English Teaching: Practice and Critique*, *9*(3), 90-105.

Hahs-Vaughn, D. L. (in progress). *Multivariate statistics.* New York, NY: Routledge/Taylor & Francis.

Hall, G. E., & Hord, S. M. (2011). *Implementing change: Patterns, principles, and potholes*. (Vol. 3). Boston: Pearson/Allyn & Bacon.

Halliday, M. A. K. (1985). *Spoken and written language.* Oxford, UK: Oxford, University Press.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education*, *5,* 93–116.

Halliday, M. A. K. (2004). The language of science. In J. Webster (Ed.) *The collected works of M.A.K. Halliday* (Vol. 5). Webster, London: Continuum.

Halliday, M. A. (2007). Language and education. In J. Webster (Ed.), *The collected works of M.A.K. Halliday* (Vol. 5). London, UK: Continuum.

Hammill, D. D., & Larsen, S. C. (2009). *Test of Written Language* (4th ed.). Austin, TX: Pro-ed.

Hand, B., Alvermann, D. E., Gee, J., Guzzetti, B. J., Norris, S. P., Phillips, L. M., ... & Yore, L. D. (2003). Message from the "Island Group": What is literacy in science literacy? *Journal of Research in Science Teaching*, *40*(7), 607-615.

Hand, B., Wallace, C. W., & Yang, E. M. (2004). Using a Science Writing Heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: quantitative and qualitative aspects. *International Journal of Science Education*, *26*(2), 131-149.

Harris, K. R., & Graham, S. (2013). "An adjective is a word hanging down from a noun": Learning to write and students with learning disabilities. *Annals of Dyslexia*, *63*(1), 65-79.

Harris, K. R., Graham, S., Friedlander, B., Leslie, L. (2013). Bring powerful writing strategies into your classroom. Why and how. *The Reading Teacher, 66*(7), 538-542.

Haug, B. S., & Ødegaard, M. (2014). From words to concepts: Focusing on word knowledge when teaching for conceptual understanding within an inquiry-based science setting. *Research in Science Education, 44*(5), 777-800.

Haynes, E. F. (1978). Using research in preparing to teach writing. *The English Journal, 67*(1), 82-88.

Haystead, M. W., & Marzano, R. J. (2009). *Meta-analytic synthesis of studies conducted at Marzano Research Laboratory on instructional strategies.* Englewood, CO: Marzano Research Laboratory.

Hebert, M., Gillespie, A., & Graham, S. (2013). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing*, *26*(1), 111-138.

Herber, H. L. (1970). *Teaching reading in the content areas*. Englewood Cliffs, NJ: Prentice Hall.

Heller, R., & Greenleaf, C. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement.* Washington, DC: Alliance for Excellent Education.

Hess, F. M., & McGuinn, P. J. (2002). Seeking the mantle of" opportunity": Presidential politics and the educational metaphor, 1964-2000. *Educational Policy*, *16*(1), 72-95.

Hiebert, E. H., & Van Sluys, K. (2014). Examining three assumptions about text complexity: Standard 10 of the Common Core State Standards. In K. S. Goodman, R. C. Calfee, & Y. M. Goodman (Eds.), *Whose knowledge counts in government literacy policies? Why expertise matters* (pp. 144-160). New York, NY: Routledge

Hillocks, G. (1986). *Research on written composition: New directions for teaching.* Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Hirschman, M. (2000). Language repair via metalinguistic means. *International Journal of Language & Communication Disorders*, *35*(2), 251-268.

Hodgson, J. (1992). The status of metalinguistic skills in reading development. *Journal of Learning Disabilities*, *25*(2), 96-101.

Hohenshell, L. M., & Hand, B. (2006). Writing-to-learn Strategies in Secondary School Cell Biology: A mixed method study. *International Journal of Science Education*, *28*(2-3), 261-289.

Hourigan, M. M. (1994). *Literacy as social exchange: Intersections of class, gender, and culture*. Albany: State University of New York Press.

Hughes, C. A. (2011). Effective instructional design and delivery for teaching task-specific learning strategies to students with learning disabilities. *Teaching Exceptional Children, 44*(2), 1-15.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Urbana, IL: National Council of Teachers of English.

Hurd, P. D. (1997). Scientific literacy: New minds for a changing world. *Science Education, 82*(3), 407-416.

Hyerle, D. (n.d.). Retrieved from http://www.thinkingfoundation.org/resources/presentations/pdf-presentations/tf-broadcast.pdf

International Reading Association. (2012). *Adolescent literacy.* [Position Statement Revised]. Newark, DE: Author.

Jellinek, C. (2012). *21st century skills: A global imperative.* (United States Department of Education. Official Web blog post). Retrieved on September 1, 2014 from http://www.ed.gov/blog /2012/03/21st-century-skills-a-global-imperative/

Jetton, T. L., & Lee, R. (2012).A model for teacher planning with text in the academic disciplines. In T. L. Jetton & C. Shanahan (Eds.), *Adolescent literacy in the academic disciplines. General principles and practical strategies* [Kindle version]. Retrieved from Amazon.com.

Jewett, P. (2013). Content-area literacy: Recognizing the embedded literacies of science and mathematics. *Journal of Reading Education*, *38*(2), 18-24.

Johanningmeier, E. V. (2010). " A Nation at Risk" and" Sputnik": Compared and Reconsidered. *American Educational History Journal*, *37*(2), 347-365.

Johnston, W. B., & Packer, A. E. (1987). *Workforce 2000: Work and workers for the twenty-first century*. Indianapolis, IN: Hudson Institute Inc.

Kamhi, A. G. (2014). Improving clinical practices for children with language and learning disorders. *Language, Speech, and Hearing Services in Schools*, *45*(2), 92-103.

Kamhi, A. G., & Catts, H. W. (2012). Language and reading: convergences and divergences. In A. G. Kamhi & H. W. Catts (Eds.), *Language and reading disabilities* (3rd ed., pp. 1-23). Boston, MA: Pearson.

Kamhi, A. G., & Koenig, L. A. (1985). Metalinguistic awareness in normal and language-disordered children. *Language, Speech, and Hearing Services in Schools*, *16*(3), 199-210.

Kamhi, A. G., Ward, M. F., & Mills, E. A. (1995). Hierarchical planning abilities in children with specific language impairments. *Journal of Speech, Language, and Hearing Research*, *38*(5), 108-116.

Kane, M., Berryman, S., Goslin, D., Meltzer, A. (1990). *The Secretary's Commission on Achieving Necessary Skills. Identifying and describing the skills required by work.* Washington, DC: Pelavin Associates, Inc.

Kelly, D., Xie, H., Nord, C. W., Jenkins, F., Chan, J. Y., and Kastberg, D. (2013). *Performance of U.S. 15-Year-Old students in mathematics, science, and reading literacy in an international context: First look at PISA 2012* (Report No. NCES 2014-024). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., ... & Kristapovich, P. (2014). *The condition of education 2014*. NCES 2014-083. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, *101*(1), 136.

Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, *36*(10), 1065-1084.

Klein, P. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, *11*(3), 203–270.

Klein, P. D., Piacente-Cimini, S., Williams, L. A. (2007). The role of writing in learning from analogies. *Learning and Instruction*, *17*(6), 595-611.

Klein, P. D., & Yu A. M. (2013). Best practices in writing to learn. In S. Graham, C.A. MacArthur, J. Fitzgerald, J. (Eds.), *Best practices in writing instruction*, (2nd ed., pp. 166-189), New York, NY: Guilford Press.

Koret Task Force for K-12 Education (2003). *Are we still at risk?* Retrieved on January 26, 2015 at http://educationnext.org/are-we-still-at-risk/

Kristapovich, P. (2014). *The condition of education 2014* (Report No. NCES 2014-083). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Kuenzi, J. J. (2008). *Science, technology, engineering, and mathematics (STEM) education: Background, federal policy, and legislative action*. (Congressional Research Service Report. Paper 35). Retrieved from http://digitalcommons.unl.edu/crsdocs/35

Langdon, D., McKittrick, G., Beede, D., Khan, B.& Doms, M. (2011). *STEM: Good jobs now and for the future.* Executive Summary. Retrieved from http://www.esa.doc.gov/sites/default/files/stemfinalyjuly14_1.pdf

Leikin, M. (2002). Processing syntactic functions of words in normal and dyslexic readers. *Journal of Psycholinguistic Research*, *31*(2), 145-163.

Lemke, J. (1990). *Talking science: Language learning, and values.* Norwood, NJ: Ablex.

Levin, H. M. (1994). Education and workplace needs. *Theory into Practice*, *33*(2), 132-138.

Levy, H., & Friedmann, N. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Language*, *29*(1), 15-49.

Lightsey, G. E., & Frye, B. J. (2004). Teaching metalinguistic skills to enhance early reading instruction. *Reading Horizons*, *45*(1), 27-37.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teac*hing (9th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Loban, Walter. (1976). Language development: *Kindergarten through grade twelve.* (NCTE Committee on Research Report No. 18, CS 202 933). Urbana, IL: National Council of Teachers of English.

Lomax, R. G. (2007). *An introduction to statistical concepts* (2nd ed.). Mahwah, NJ: Lawrence and Earlbaum.

Lomax, R. G., & Hahs-Vaughn, D. L. (2010). *An introduction to statistical concepts*. (3rd ed.) New York, NY: Routledge.

Long, D., & Carlson, D. (2011). Mind the map: How Thinking Maps affect student achievement. *Networks, 13*(2), 1-7.

MacArthur, C.A., Philippakos, Z., Graham, S, Harris, K. (2012). Writing Instruction. In Wong, B., Butler, D. L. (Eds.), *Learning about learning disabilities* (4th ed., pp. 243-270). San Diego, CA: Elsevier Academic Press.

Maize, R. C. (1954). A writing lab for retarded students. *College English, 16*(1), 44-48.

Martin, M. O. Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 International results in science.* Executive summary. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, and Amsterdam: the Netherlands International Association for the Evaluation of Educational Achievement [IEA].

Marx, R. W., & Harris, C. J. (2006). No Child Left Behind and science education: Opportunities, challenges, and risks. *The Elementary School Journal*, *106*(5), 467-478.

Marzano, R. J. (2010). Representing knowledge nonlinguistically. *Educational Leadership, 67*(8), 84-86.

Marzano, R. J. (2012). Writing to learn. *Educational Leadership, 69*(5), 82-83.

Marzano, R.J. (2014). *2014 Marzano teacher evaluation learning map.* Learning Sciences International.

Mason, L. H. (2009). Effective Instruction for Written Expression. *Perspectives of Language and Literacy, 35(*3), 21-24.

McFarlane, D. A. (2013). Understanding the Challenges of Science Education in the 21st Century: New Opportunities for Scientific Literacy. *International Letters of Social and Humanistic Sciences*, *4*, 35-44.

McLeod, S. H. (1992). Writing across the curriculum: An introduction. In S.H. McLeod, & M. Soven (Eds.), *Writing across the curriculum: A guide to developing programs*(pp. 1-11). Retrieved on from http://wac.colostate.edu/books/mcleod_soven/mcleod_soven.pdf

McLeod, S. H., Miraglia, E., Soven, M., & Thaiss, C. (2001). *WAC for the New millennium: strategies for continuing writing-across-the-curriculum programs*. Urbana, IL: National Council of Teachers of English.

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2013). *Coh-Metrix version 3.0.* Retrieved on January 15, 2015 from http://cohmetrix.com

McNamara, D. S., Graesser, A. C., McCarthy, P. M., Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix.* [Kindle version]. Retrieved from Amazon.com

Mellon, J. C. (1969). *Transformational sentence combining: A method for enhancing the development of syntactic fluency in English composition* (No. 10-13). Urbana, IL: National Council of Teachers of English.

Meltzer, J. (2001). *Supporting adolescent literacy across the content areas. Perspectives on policy and practice.* Washington DC: Office of Educational Research and Improvement.

Menyuk, P. (1976). Relations between acquisition of phonology and reading. *Aspects of Reading Acquisition.* Baltimore, MD: Johns Hopkins.

Menyuk, P., & Flood, J. (1981). Linguistic competence, reading, writing problems and remediation. *Annals of Dyslexia*, *31*(1), 13-28.

Miller, J. D. (1989, January). *Scientific Literacy.* Paper presented at the annual meeting of the American Association for the Advancement of Science, San Francisco, California.

Miller, B., & McCardle, P. (2011). Reflections on the need for continued research on writing. *Reading and Writing*, *24*(2), 121-132. doi:10.1007/s11145-010-9267-6

Miller, B. D., & Ney, J. W. (1968). The effect of systematic oral exercises on the writing of fourth-grade students. *Research in the Teaching of English*, *2*(1), 44-61.

Mo, Y., Kopke, R. A., Hawkins, L. K., Troia, G. A., & Olinghouse, N. G. (2014). The neglected "R" in a time of Common Core. *The Reading Teacher*, *67*(6), 445-453.

Mokhtari, K., & Thompson, H. B. (2006). How problems of reading fluency and comprehension are related to difficulties in syntactic awareness skills among fifth graders. *Literacy Research and Instruction*, *46*(1), 73-94.

Moje, E. B. (2007). Developing socially just subject-matter instruction: A review of the literature on disciplinary literacy teaching. *Review of Research in Education*, *31*(1), 1-44.

Moje, E. B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy*, *52*(2), 96-107.

Mokhtari, K., & Thompson, H. B. (2006). How problems of reading fluency and comprehension are related to difficulties in syntactic awareness skills among fifth graders. *Literacy Research and Instruction*, *46*(1), 73-94.

Myhill, D. A., Jones, S. M., Lines, H., & Watson, A. (2012). Re-thinking grammar: The impact of embedded grammar teaching on students' writing and students' metalinguistic understanding. *Research Papers in Education*, *27*(2), 139-166.

Myhill, D., Jones, S., & Watson, A. (2013). Grammar matters: How teachers' grammatical knowledge impacts on the teaching of writing. *Teaching and Teacher Education*, *36*, 77-91.

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics*, *21*(02), 229-241.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*(4), 342-356.

National Academy of Sciences. (2011). *Rising above the gathering storm, revisited: Rapidly approaching category 5*. Condensed Version. Retrieved from http://www.bradley.edu/dotAsset/187205.pdf.

National Center for Educational Statistics. (2007) *Comparing TIMSS with NAEP and PISA in mathematics and science*. Retrieved from http://nces.ed.gov/timss/pdf/Comparing_TIMSS_NAEP_%20PISA.pdf

National Center for Education Statistics (2011). *The nation's report card: Reading 2011* (NCES 2012–457). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Center for Educational Statistics. (2012a). *The nation's report card: Science 2011*. Washington, DC: U.S. Department of Education, Institute for Education Sciences.

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

National Center for Educational Statistics. (2013). *The nation's report card: Trends in academic progress 2012. (*NCES 2013 456).Washington, DC: U.S. Department of Education, Institute for Education Sciences.

National Center for Educational Statistics. (2014a). *The nation's report card: Are the nation's 12th-graders making progress in mathematics and reading? Trends in academic progress 2012.* (NCES 2014 087). Washington, DC: U.S. Department of Education, Institute for Education Sciences.

National Center for Educational Statistics. (2014b). *The nation's report card: A first look: 2013 mathematics and reading (*NCES 2014 451). Washington, DC: U.S. Department of Education, Institute for Education Sciences.

National Center for Education Statistics, U.S. Department of Education, Institute of Education Sciences. (2014). *Status dropout rates.* Retrieved on September 6, 2014 from http://nces.ed.gov/programs/coe/indicator_coj.asp#info

National Center for Public Policy and Higher Education and the Southern Regional Education Board. (2009). *States, schools, and colleges. Policies to improve student readiness for college and strengthen coordination between schools and colleges*. Retrieved on January 27, 2015 from http://www.highereducation.org/reports/reports_center_2009.shtml

National Center for Public Policy and Higher Education and the Southern Regional Education Board. (2010*). Beyond the rhetoric. Improving college readiness through coherent state policy and higher education.* (Special Report). Retrieved on September 15, 2014 from http://www.highereducation.org/ reports/college_readiness/CollegeReadiness.pdf

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal*, *84*(2), 112-130.

National Commission on the High School Senior Year. (2001). *Raising our sights: No high school senior left behind.* (Final Report). Princeton, NJ: Woodrow Wilson National Fellowship Foundation.

National Commission on Writing for America's Schools and Colleges. (2003). *The neglected "R": The need for a writing revolution*. (Report of the National Commission on Writing, The College Board) Retrieved from http://www.collegeboard.com/prod_downloads/ writingcom/neglectedr.pdf

National Commission on Writing for America's Families, Schools, and Colleges. (2004). *Writing; A ticket to work…or a ticket out.* (Report of the National Commission on

Writing, The College Board). Retrieved from http://www.collegeboard.com/prod_ downloads/writingcom/writing-ticket-to-work.pdf

National Commission on Writing for America's Families, Schools, and Colleges. (2005). *Writing: A powerful message from state government. (*Report of the National Commission on Writing, The College Board). Retrieved from http://www.collegeboard.com/prod_downloads/writingcom/powerful-message-from-state.pdf

National Commission on Writing for America's Families, Schools, and Colleges. (2006). *Writing and school reform.* (Report of the National Commission on Writing, The College Board). Retrieved from http://www.collegeboard.com/prod_downloads/writingcom/writing-school-reform-natl-comm-writing.pdf

National Conference of State Legislatures. (2015). *Hot topics in higher education. Reforming remedial education.* Retrieved on January 27, 2015 at http://www.ncsl.org/research/ education/improving-college-completion-reforming-remedial.aspx

National Council of Teachers of English. (2006). *NCTE principles of adolescent literacy reform.* (Policy research brief). Retrieved from http://www.ncte.org/library/nctefiles/resources/ policyresearch/adollitprinciples.pdf

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010a). *Common Core State Standards*. Washington, DC: Authors.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010b). *Common Core State Standards*. *Appendix A.* Washington, DC: Authors.

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

National Literacy Act, Pub. L. 102-73, 105 Stat. 333. (1991).

National Research Council. (1996). *National Science Education Standards.* Retrieved from http://www.nap.edu/catalog/4962.html

National Research Council, Committee on a Conceptual Framework for New K-12 Science Education Standards, Board on Science Education, Division of Behavioral and Social Sciences and Education. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

National Academies Committee on Science, Engineering, and Public Policy. (2005). *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* Executive Summary. Retrieved from https://www.nsf.gov/attachments/105652/public/NAS-Gathering-Storm-11463.pdf

National Science and Technology Council. (2013). *Federal science, technology, engineering, and mathematics (STEM) education 5-year strategic plan*. (Report from the Committee on STEM Education). Retrieved from http://www.whitehouse.gov/sites/default/files /microsites/ostp/stem_stratplan_2013.pdf

Nelson, N.W. (2014). Classroom based writing assessment. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders (*2nd ed., pp. 524-544). New York, NY: Guilford Press.

Nelson, N.W. (2013a). Implications for assessment and intervention. *Perspectives on Language and Literacy*, *39*(3), 9-17.

Nelson, N.W., (2013b). Simple to complex-Targeting elaboration before correctness. *Topics in Language Disorders*, *33*(2), 125-139.

Nelson, N.W., Roth, F., Scott, C., Van Meter, A., Troia, G. (2006). *Written language assessment and intervention. It targets more than writing.* Paper presented at the 2006 American Speech-Language-Hearing Association Annual Conference, Miami, FL. Retrieved from http://search.asha.org/default.aspx?q=Written%20Language% 20Assessment%20%26%20Intervention

Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading & Writing Quarterly*, *23*(3), 287-309.

Neville, D. D., & Searls, E. F. (1985). The effect of sentence combining and kernel-identification training on the syntactic component of reading comprehension. *Research in the Teaching of English*, *19*(1) 37-61.

Neville, D. D., & Searls, E. F. (1991). A meta-analytic review of the effect of sentence-combining on reading comprehension. *Literacy Research and Instruction*, *31*(1), 63-76.

Next Generation Science Standards Lead States. (2013a). *Next Generation Science Standards: for states, by states. Appendix C- College and Career Readiness.* Washington, DC: The National Academies Press.

Next Generation Science Standards Lead States. (2013b). *Next Generation Science Standards: for states, by states. Appendix D- All standards, All students: Making the Next Generation Science Standards accessible to all students.* Washington, DC: The National Academies Press.

Next Generation Science Standards Lead States. (2013c). *Next Generation Science Standards: for states, by states. Appendix G- Crosscutting concepts.* Washington, DC: The National Academies Press.

Next Generation Science Standards Lead States. (2013d). *Next Generation Science Standards: for states, by state*s. *Appendix M- Connections to the Common Core State Standards for Literacy in Science and Technical Subjects.* Washington, DC: The National Academies Press.

Next Generation Science Standards Lead States. (2013e). *Next Generation Science Standards: For States, By States. Executive Summary.* Washington, DC: The National Academies Press.

Next Generation Science Standards Lead States. (2013f). *Next Generation Science Standards: for states, by state*s. *Topic Arrangements of the Next Generation Science Standards* Washington, DC: The National Academies Press.

Neville, G. E. (2006). Writing to learn. How alternative theories of school writing account for student performance. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 235-247). New York, NY: Guilford Press.

Nippold, M. A. (1993). Developmental markers in adolescent language. Syntax, semantics, and pragmatics. *Language, Speech, and Hearing Services in Schools, 24*(1), 21-28.

Nippold, M. A. (2000). Language development during the adolescent years: Aspects of pragmatics, syntax, and semantics. *Topics in Language Disorders*, *20*(2), 15-28.

Nippold, M. A. (2007). *Later language development: School-age children, adolescents, and young adults*, (3rd ed.). Austin, TX: Pro Ed.

Nippold, M. A. (2010). *Language sampling with adolescents.* San Diego, CA: Plural Publishing.

Nippold, M. A. (2014). Language intervention at the middle school: Complex talk reflects complex thought. *Language, Speech, and Hearing Services in Schools*, *45*(2), 153-156.

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse a study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research, 48*(5), 1048-1064.

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2008). Expository Discourse in adolescents with language impairments. Examining syntactic development. *American Journal of Speech-Language Pathology*, *17, 356-366*.

Nippold, M. A., & Scott, C. M. (2010). Overview of expository discourse. In Nippold, M. A. & Scott, C. M. (Eds.). *Expository discourse in children, adolescents, and adults: Development and disorders* (pp.-11). New York, NY: Psychology Press/Taylor & Francis.

Nippold, M. A., Ward-Lonergan, J. M., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults. A study of syntactic, semantic, and pragmatic development. *Language, Speech, and Hearing Services in Schools*, *36*(2), 125-138.

No Child Left Behind Act of 2001, Pub. L. No. 107–110, 155 Stat. 1425 (2002).

O'Donnell, R. C. (1967). *Syntax of kindergarten and elementary school children: A transformational analysis.* Champaign, IL: National Council of Teachers of English.

O'Hare, F. (1973). *Sentence combining: Improving student writing without formal grammar instruction.* Urban, IL: National Council of Teachers of English.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, *69(*348), 894-908.

Organisation for Economic Co-operation and Development [OECD]. (2014). *PISA 2012 Results in focus: What 15-year-olds know and what they can do with what they know*. Retrieved from the OECD website http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Owens, R. E. (2008). *Language development: An introduction* (7th ed.). Boston, MA: Pearson Education Inc.

Owens Jr, R. E., Farinella, K. A., & Metz, D. E. (2015). *Introduction to Communication disorders: A lifespan evidence-based perspective*. (5th ed.) New York, NY: Pearson Higher Ed.

Partnership for 21st Century Skills. (2011). *Framework for 21st century learning.* (Brochure). Retrieved from Partnership for 21st century learning website: http://www.p21.org/storage/documents/1.__p21_framework_2-pager.pdf

Partnership for 21st Century Skills. (2014). Our history. Retrieved on September 4, 2014 from Partnership for 21st century skills website: http://www.p21.org/about-us/our-history

Paul, R., & Norbury, C. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating*.(4th ed.). St. Louis, MO: Elsevier Health Sciences.

Pearson, P. D., & Gallagher, M. C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology*, *8*(3), 317-344.

Pearson, P. D., Moje, E., Greenleaf, C. (2010). Literacy and science: Each in service of the other. *Science*, *328*(5977), 459-463.

Peets, K. F. (2014). Cognitive processes underlying typical and atypical second-language and literacy development. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders (*2nd ed., pp. 545-568). New York, NY: Guilford Press.

Peha, S. (1995). *Writing across the curriculum. Tips, tools, and techniques for improving student writing in the content areas.* Retrieved from http://elementary.nrms.wikispaces.net/file/view/Writing+Across+the+Curriculum.pdf

Pence-Turnbull, K. L., & Justice, L. M. (2013). *Language development from theory to practice* (2nd ed.). Upper Saddle River, NJ: Pearson Higher Ed.

Perera, K. (1984). *Children's writing and reading: Analysing classroom language*. Oxford, England: Blackwell.

Perin, D. (2013). Best practices in teaching writing for college and career readiness. In S. Graham, C. A. MacArthur, J. Fitzgerald, J. (Eds.), *Best practices in writing instruction,* (2nd. ed., pp. 48-72), New York, NY: Guilford Press.

Phelps-Gunn, T., & Phelps-Terasaki, D. (1982). *Written language instruction: Theory and remediation*. Rockville, MD: Aspen Systems Corporation.

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common Core State Standards: The new US intended curriculum. *Educational Researcher*, *40*(3), 103-116.

Powell-Smith, K. A., & Shinn, M. R. (2006). *Administration and scoring on written expression curriculum-based measurement (WE-CBM) for use in general outcome measurement.* Eden Prairie, MN: Edformation Inc.

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. (2008). Assessing the microstructure of written language using a retelling paradigm. *American Journal of Speech-Language Pathology*, *17*(2), 107–120. doi:10.1044/1058-0360(2008/012)

Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*(1), 103-121.

Quinn, H., Keller, T., Moulding, B., Eberle, F. (2012). *A framework for K-12 science education: Practices, crosscutting concepts and core ideas.* Webinar from the National Science

Teachers Association. Retrieved from http://learningcenter.nsta.org/products/symposia_seminars/NGSS/webseminar.aspx

Rijlaarsdam, G., Janssen, T., Braaksma, M., Van Steendam, E., Van den Branden, K., Couzjin, M., Verheyden, L. (2014). Developmental variations in the production of written text: Challenges for students who struggle with writing. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders (*2nd. ed., pp. 545-568). New York, NY: Guilford Press.

Rogers, L. A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology*, *100*(4), 879-906.

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, *113*(3), 553-565.

Romance, N. R., & Vitale, M. R. (2012a). Expanding the role of K-5 science instruction in educational reform: Implications of an interdisciplinary model for integrating science and reading. *School Science and Mathematics*, *112*(8), 506-515.

Romance, N. R., & Vitale, M. R. (2012b). Science IDEAS: A research-based K-5 interdisciplinary instructional model linking science and literacy. *Science Educator*, *21*(1), 1-12.

Rosenshine, B. V. (1986). Synthesis of research on explicit teaching. *Educational Leadership, 43*(7), 60-69.

Roth, F. P., Speece, D. L., Cooper, D. H., & De La Paz, S. (1996). Unresolved mysteries: How do metalinguistic and narrative skills connect with early reading? *The Journal of Special Education*, *30*(3), 257-277.

Routman, R. (2005). *Writing essentials: Raising expectations and results while simplifying teaching*. Portsmouth, NH: Heinemann.

Saddler, B. (2012). Teacher's guide to effective sentence writing. In K. R. Harris & S. Graham (Eds.). *What works for special needs learners.* [Kindle version]. Retrieved from Amazon.com

Saddler, B. (2013). Best practices in sentence construction skills. In In S. Graham, C.A. MacArthur, J. Fitzgerald, J. (Eds.), *Best practices in writing instruction*, (2nd. ed., pp. 238-256), New York, NY: Guilford Press.

Saddler, B., & Asaro-Saddler, K. (2009). Writing better sentences: Sentence-combining instruction in the classroom. *Preventing School Failure: Alternative Education for Children and Youth*, *54*(3), 159-163.

Saddler, B., & Asaro-Saddler, K. (2013). Response to Intervention in writing: A suggested framework for screening, intervention, and progress monitoring. *Reading & Writing Quarterly*, *29*(1), 20-43.

Saddler, B., Asaro, K., & Behforooz, B. (2008). The effects of peer-assisted sentence-combining practice on four young writers with learning disabilities. *Learning Disabilities: A Contemporary Journal*, *6*(1), 17-31.

Saddler, B., Behforooz, B., & Asaro, K. (2008). The effects of sentence-combining instruction on the writing of fourth-grade students with writing difficulties. *The Journal of Special Education*, *42*(2), 79-90.

Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence-combining instruction on the writing of more and less skilled young writers. *Journal of Educational Psychology*, *97*(1), 43-54.

Saddler, B., & Preschern, J. (2007). Improving sentence writing ability through sentence-combining practice. *Teaching Exceptional Children*, *39*(3), 6-11.

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The Nation's Report Card: Writing 2007* (NCES 2008–468). Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Sanders, M. (2009). Stem, stem education, stemmania. *The Technology Teacher*, *68*(4), 20-26.

Scarborough, H. S. (1991). Early syntactic development of dyslexic children. *Annals of Dyslexia*, *41*(1), 207-220.

Science Pioneers. (2014). *Why STEM education is important for everyone.* Retrieved on September 2, 2014 from http://www.sciencepioneers.org/parents/why-stem-is-important-to-everyone

Scholl, D. M., & Ryan, E. B. (1980). Development of metalinguistic performance in the early school years. *Language and Speech*, *23*(2), 199-211.

Schleppegrell, M. J. (2007). At last: The meaning in grammar. *Research in the Teaching of English*, *42*(1), 121-128.

Schleppegrell, M. J. (2013). Exploring language and meaning in complex texts. *Perspectives on Language and Literacy*, *39*(3), 37-40.

Scholl, D. M., & Ryan, E. B. (1980). Development of metalinguistic performance in the early school years. *Language and Speech*, *23*(2), 199-211.

Schumaker, J. (1989). Strategies Instruction Model. *Strategram*, *1*(4), Lawrence, KS: University of Kansas Center for Research on Learning.

Schumaker, J. B., Deshler, D. D., Woodruff, S. K., Hock, M., Bulgren, J. A., & Lenz, B. K. (2006). Reading strategy interventions: Can literacy outcomes be enhanced for at-risk adolescents? *Teaching Exceptional Children*, *38*(3), 64-68.

Scott, C. M. (1988). Producing complex sentences. *Topics in Language Disorders*, *8*(2), 44-62.

Scott, C. M. (2004). Syntactic contributions to literacy learning. In C.A. Stone, EAR. Silliman, B. J. Ehren, & K. Apel (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 340–362). New York, NY: Guilford Press.

Scott, C. M. (2009a). A case for the sentence in reading comprehension. *Language, Speech, and Hearing Services in Schools, 40(*2), 184-191.

Scott, C. M. (2009b). Language-Based Assessment of Written Expression. In G. A. Troia, (Ed.). *Instruction and assessment for struggling writers: Evidence-based practices (*2nd ed.). New York, NY: Guilford Press.

Scott, C. M. (2012). Learning to write. In A. G. Kamhi & H. W. Catts (Eds.), *Language and reading disabilities (*pp. 244-268). Boston, MA: Pearson Education Inc.

Scott, C. M. (2014). One size does not fit all: Improving clinical practice in older children and adolescents with language and learning disorders. *Language, Speech, and Hearing Services in Schools*, *45*(2), 145-152.

Scott, C. M., & Balthazar, C. H. (2010). The grammar of information: Challenges for older students with language impairments. *Topics in Language Disorders, 30*(4), 288-307.

Scott, C. M., & Balthazar, C. H. (2013). The role of complex syntax knowledge in children with reading and writing difficulties. *Perspectives on Language and Literacy, 39*(3), 18-26.

Scott, C. M., & Koonce, N. M. (2014). Syntactic contributions to literacy learning. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders* (2nd. ed., pp. 283–301). New York, NY: Guilford Press.

Scott, C. M., & Nelson, N. W. (2009). Sentence combining: Assessment and intervention applications. *Perspectives on Language Learning and Education, 16*(1), 14-20.

Scott, C. M., & Nelson, N. W., Anderson, S. A., Zelinski K. (2006). Poster presentation at the 2006 American Speech-Language-Hearing Association Annual Conference, Miami, FL. Retrieved from http://search.asha.org/default.aspx?q=!Development+of+Written+ Sentence +Combining+Skills+in+School-Age+Children

Scott, C. M., & Stokes, S. L. (1995). Measures of syntax in school-age children and adolescents. *Language, Speech, and Hearing Services in Schools, 26*(4), 309-319.

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

Secretary's Commission on Achieving Necessary Skills. (1991). *What work requires of schools.* (SCANS Report for America 2000). Washington, DC: U.S. Department of Labor.

Secretary's Commission on Achieving Necessary Skills. (1992*). Learning a living. A blueprint for high performance.* (SCANS Report for America 2000). Washington, DC: U.S. Department of Labor

Secretary's Commission on Achieving Necessary Skills. (1999). *Skills and tasks for jobs.* (SCANS Report for America 2000). Washington, DC: U.S. Department of Labor.

Seltman, H. J. (2012). *Experimental design and analysis.* Retrieved from http://www. stat. cmu. edu/, hseltman/309/Book/Book. pdf.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shanahan, C. (2012a). How disciplinary experts read. In T. L. Jetton & C. Shanahan (Eds.), *Adolescent literacy in the academic disciplines. General principles and practical strategies* [Kindle version]. Retrieved from Amazon.com.

Shanahan, C. (2012b). Learning with text in science. In T. L. Jetton & C. Shanahan (Eds.), *Adolescent literacy in the academic disciplines. General principles and practical strategies* [Kindle version]. Retrieved from Amazon.com.

Shanahan, T. (2006b). Relations among oral language, reading, and writing development. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 171-183). New York: Guilford Press.

Shanahan, T. (2009). Connecting reading and writing instruction for struggling writers. In G. A. Troia, (Ed.), *Instruction and assessment for struggling writers. Evidence-based practices* (pp. 358-385). New York: Guilford Press.

Shanahan, T. (2013). Best practice in writing about text. In S. Graham, C.A. MacArthur, J. Fitzgerald, (Eds.), *Best practices in writing instruction* (2nd. ed., pp. 334-350), New York, NY: Guilford Press.

Shanahan, T., Fisher, D., & Frey, N. (2012). The challenge of challenging text. *Educational Leadership, 69*(6), 58-62.

Shanahan. T. & Lomax, R. G. (1986). An analysis and comparison of theoretical models of the reading–writing relationship. *Journal of Educational Psychology 78*(2), 116-123.

Shanahan, T., & Lomax, R. G. (1988). A developmental comparison of three theoretical models of the reading-writing relationship. *Research in the Teaching of English*, *22*(2)196-212.

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

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

Sheils, M. (1975). Why Johnny can't write. *Newsweek, 92*(8), 58-65.

Singer, B. D., & Bashir, A. S. (1999). What are executive functions and self-regulation and what do they have to do with language-learning disorders? *Language, Speech, and Hearing Services in Schools, 30*(3), 265-273.

Singer, B. D., & Bashir, A. S. (2004). Developmental variations in writing composition skills. In Stone, C. A., Silliman, E. R., Ehren B. J., Apel, K. (Eds.), *Handbook of language and literacy,* (pp. 559-582). New York, NY: Guilford Press.

Smagorinsky, P. (Ed.). (2006). *Research on composition*. New York, NY: Teachers College.

Smith, C. (1988). Does it help to write about your reading? *Journal of Reading*, *32*(3), 276–277.

Smith, N. B. (1964). Patterns of writing in different subject areas. *Journal of Reading*, *8*(1), 31-37.

Smith, M. W., Cheville, J., Hillocks Jr., G. (2006). "I guess I'd better watch my English" Grammars and the teaching of English language arts." In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 263-274). New York, NY: Guilford Press.

Somerville, J., & Yi, Y. (2002). Curriculum and assessment systems. In State Higher Education Officers P-16 Initiative, *Student success: Statewide P-16 systems* (pp. 27-35). Denver, CO: State Higher Education Officers.

Southern Educational Board. (2013). *Get it in writing. Making adolescent writing an immediate priority in Texas.* Atlanta, GA: Author.

Stage, E. K., Asturias, H., Cheuk, T., Dero, P.A., Hampton, S. B. (2013). Opportunities and challenges in Next Generation Standards. *Science, 340*(6130), 276-277. doi: 10.1126/ science.1234011

STEM Education Coalition. (2013). *The case for STEM education as a national priority: Good jobs and American competitiveness.* Retrieved from http://www.stemedcoalition.org/wp-content/uploads/2013/10/Fact-Sheet-STEM-Education-Good-Jobs-and-American-Competitiveness-June-2013.pdf

Stevens, J. P. (2007). *Intermediate statistics: A modern approach* (3rd ed.). New York, NY: Routledge.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental Psychology*, *38*(6), 934.

Stotsky, S. L. (1975). Sentence combining as a curricular activity: its effect on written language development and reading comprehension. *Research in the Teaching of English*, *9*(1)30-71.

Stotsky, S. (1983). Research on reading/writing relationships: A synthesis and suggested directions. *Language Arts*, *60*(5), 627-642.

Straw, S. B., & Schreiner, R. (1982). The effect of sentence manipulation on subsequent measures of reading and listening comprehension. *Reading Research Quarterly*, 339-352.

Street, C., & Stang, K. (2008). Improving the teaching of writing across the curriculum: A model for teaching in-service secondary teachers to write. *Action in Teacher Education*, *30*(1), 37-49.

Strong, W. (1976). Sentence combining: back to basics and beyond. *The English Journal*, *65*(2), 56-64.

Strong, W. (1986). *Creative approaches to sentence combining*. Urbana, IL: ERIC Clearinghouse on Reading and Composition Skills and the National Conference on Research in English. (ERIC Document Reproduction Service No. ED274985)

Sun, L., & Nippold, M. A. (2012). Narrative writing in children and adolescents: Examining the literate lexicon. *Language, Speech, and Hearing Services in Schools, 43*(1), 2-13.

Sutter, J. C., & Johnson, C. J. (1990). School-age children's metalinguistic awareness of grammaticality in verb form. *Journal of Speech, Language, and Hearing Research,* 33(1), 84-95.

Swanson, H. L., & Deshler, D. (2003). Instructing adolescents with learning disabilities converting a meta-analysis to practice. *Journal of Learning Disabilities*, *36*(2), 124-135.

Symonds, P. M. (1931). Practice versus grammar in the learning of correct English usage. *Journal of Educational Psychology*, *22*(2), 81.

Tharp, R.G., & Gallimore, R. (1988). *Rousing minds to life. Teaching, learning, and schooling in a social context.* New York, NY: Cambridge University Press.

Thompson, C. K., & Shapiro, L. P. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology*, *16*(1), 30-42.

Torgerson, D. J., & Torgerson, C. J. (2008). *Designing randomized trials in health, education and the social sciences: An introduction.* New York, NY: Palgrave MacMillan.

Troia, G. A. (2006). Writing instruction for students with learning disabilities. In MacArthur. C.A., Graham, S., Fitzgerald. (Eds.), *Handbook of writing research*, (pp. 324-336). New York, NY: Guilford Press.

Troia, G. A. (2009). Tactics to help students use writing to express their comprehension of content area texts. *Perspectives of Language and Literacy*, *35*(3), 35-41.

Troia, G. A. (2011). How might pragmatic language skills affect the written expression of students with language learning disabilities? *Topics in Language Disorders*, *31*(1), 40-53.

Troia, G. A., & Graham, S. (2003). Effective writing instruction across the grades: What every educational consultant should know. *Journal of Educational and Psychological Consultation*, *14*(1), 75-89.

Troia, G. A., & Olinghouse, N. G. (2013). The common core state standards and evidence-based educational practices: The case of writing. *School Psychology Review*, *42*(3), 343-357.

Tunmer, W. E. (1984). The development of metalinguistic awareness: A conceptual overview. In W. E. Tunmer, C. Pratt, & M. L. Herrimah (eds.) *Metalinguistic awareness in children* (pp. 12-35). Berlin, Heidelberg: Springer.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly*, *23*(2), 134-158.

United States Department of Labor. (n.d.) *Executive summary. Workplace essential skills: Resources related to the SCANS competencies and foundation skills.* Retrieved from wdr.doleta.gov/opr/fulltext/00-wes_es.pdf

United States Chamber of Commerce. The Center for Workforce Proficiency. (n.d.) *The chamber guide to workforce literacy*. Retrieved from https://www.workforce3one.org /view/ 1860

United States Department of Education. (1991). *Adult learning and literacy. Adult Education Act: 1964-1991*. Retrieved from http://www2.ed.gov/about/offices/list/ovae/pi/AdultEd/ anniv40/ae-act.pdf

United States Department of Education. (2014a). *Science, technology, engineering and math: (STEM) Education.* Retrieved from http://www2.ed.gov/about/overview/budget/ budget15/crosscuttingissues/stem.pdf

United States Department of Education. (2014b). *Science, technology, engineering and math: Education for global leadership.* Retrieved on August 30, 2014 from http://www.ed.gov/stem.

Van Kleeck, A. (1982). The emergence of linguistic awareness: A cognitive framework. *Merrill-Palmer Quarterly, 28*(2), 237-265.

Varnum, R. (1986). From crisis to crisis: The evolution toward higher standards of literacy in the United States. *Rhetoric Society Quarterly, 16*(3), 145-165.

Vavra, E. (1987). Grammar and syntax: The student's perspective. *English Journal, 75*(6), 42-48.

Venezia, A., Callan, P. M., Finney, J. E., Kirst, M. W., & Usdan, M. D. (2005). *The governance divide: A report on a four-state study on improving college readiness and success.* (National Center Report No. 05-3). San Jose, CA: The National Center for Public Policy and Higher Education.

Videen, J., Deno, S., & Marston, D. (1982). *Correct word sequences: A valid indicator of proficiency in written expression* (Research Report 84). Minneapolis, MN: Institute for Research on Learning Disabilities.

Villaneuva, G. A., & Hand, B. (2011). Science for all: Engaging students with special needs in and about science. *Learning Disabilities Research & Practice, 26*(4), 233–240.

Vocke, K., Jewell, E., Doherty-McDowell, C., Wolford, D., Ehst, S., Hovey, A. M., ... & Gonzalez, K. (2012). Grammar: Navigating teaching without crashing and burning. *Language Arts Journal of Michigan, 27*(2), 9.

Vygotsky, I. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky* (Vol. 1, pp. 37-285). New York: Plenum.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology, 102*(4), 860.

Walker, B. J., Mokhtari, K., Sargent, S. (2006). Reading fluency: More than fast and accurate reading. In Rasinski, T., Blachowicz, C. Lems, K. (Eds). *Fluency instruction: Research based best practices (pp. 86-105).* New York, NY: Guilford Press.

Wallach, G. P. (2014). Improving clinical practice: A school-age and school-based perspective. *Language, Speech, and Hearing Services in Schools, 45*(2), 127-136.

Wallach, G. P., Charlton, S., Bartholomew, J. C. (2014). The spoken-written comprehension connection. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders* (2nd ed., pp. 619-636). New York, NY: Guilford Press.

Wallach, G., & Weis-Liebergott, J. (1984). Who shall be called "learning disabled": Some new directions. In Wallach, G. & Butler, K. (Eds.), *Language and learning disabilities in school-age children* (pp.-13). Baltimore, MD: Waverly Press.

Weaver, C. (1996). Teaching grammar in the context of writing. *English Journal, 85*(7), 15-24.

Weaver, C., Bush, J., Anderson, J., & Bills, P. (2006). Grammar Intertwined throughout the Writing Process: An" Inch Wide and a Mile Deep." *English Teaching: Practice and Critique*, *5*(1), 77-101.

Wellman, H. M. (1990). *The child's theory of mind*, Cambridge, MA, MIT Press.

West, S. G., Finch J. F., Curran P. J. (1996). Structural equation models with nonnormal variables: problems and remedies. In: Hoyle RH, editor. *Structural equation modeling: Concepts, issues and applications*. Newbery Park, CA: Sage; 1995. pp. 56–75.

Westby, C. (2012). In A.G. Kamhi & H. W. Catts (Eds.), *Language and reading disabilities (*3rd ed., pp. 163-221). Boston, MA: Pearson Education Inc.

Westby, C. (2014). A language perspective on executive functioning, metacognition, and self-regulation in reading. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders* (2nd. ed., pp. 619-636). New York, NY: Guilford Press.

White House Office of Science and Technology. (2014). Preparing Americans with 21st century skills. Science, technology, engineering, and mathematics (STEM) education in the 2015 budget. Retrieved from http://www.whitehouse.gov/sites/default/files/ microsites/ ostp/fy_2015_stem_ed.pdf

Wilkinson, P. A., & Patty, D. (1993). The effects of sentence combining on the reading comprehension of fourth-grade students. *Research in the Teaching of English, 27*(1)104-125.

Williams, J. P., Hall, K. M., & Lauer, K. D. (2004). Teaching expository text structure to young at-risk learners: Building the basics of comprehension instruction. *Exceptionality*, *12*(3), 129-144.

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality**.** *College Composition and Communication*, *32*(2), 189-204.

Wise, B. (2008). High schools at the tipping point. *Educational Leadership. Reshaping High Schools*, *65*(8), 8-13.

Wyse, D. (2001). Grammar for writing? A critical review of the empirical evidence. *British Journal of Educational Studies, 49*(4), 411-427.

Yore, L., Bisanz, G. L., & Hand, B. M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, *25*(6), 689-725.

Yore, L. D., Florence, M. K., Pearson, T. W., & Weaver, A. J. (2006). Written discourse in scientific communities: A conversation with two scientists about their views of science, use of language, role of writing in doing science, and compatibility between their epistemic views and language. *International Journal of Science Education, 28*(2-3), 109-141.

Yore, L. D., Pimm, D., & Tuan, H. L. (2007). The literacy component of mathematical and scientific literacy. *International Journal of Science and Mathematics Education*, *5*(4), 559-589.

Yule, G. (2010). *The study of language*. New York: Cambridge University Press.

Zumbrunn, S., & Kraus, K. (2012). Conversations with leaders: Principles of effective writing instruction. *The Reading Teacher, 65*(5), 346-353. doi:10.1002/TRTR.01053

Zygouris-Coe, V. (2012). Disciplinary literacy and the common core state standards. *Topics in Language Disorders*, *32*(1), 35-50.