METHODS AND MEASURES FOR USING WRITING TO TRANSFORM KNOWLEDGE IN SCIENCE CLASSES

A Dissertation

by

KATHERINE ELIZABETH LANDAU WRIGHT

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Erin M. McTigue
Committee Members,	Li-Jen Kuo
	Zohreh R. Eslami
	Patricia Goodson
	Sharon Matthews
Head of Department,	Lynn Burlbaw

August 2016

Major Subject: Curriculum and Instruction

ABSTRACT

This dissertation documents three connected studies addressing critical issues in writing-to-learn research: a) how to measure students' feelings about writing, b) how to assess scientific writing, and c) how to integrate writing-to-learn into current secondary science curriculum. Considered in concert, this work seeks to provide measures and methods for using writing as a tool to transform knowledge in secondary science classes.

Students' emotions about writing impact achievement, and therefore tools capturing motivation, attitude, and self-efficacy are needed to determine the extent that writing achievement is a result of skill development, affective issues, or a combination of both. Specifically, my first study describes the validation of a newly developed measure of self-efficacy towards writing for middle grades students called the *Student Writing Affect Survey* (SWAS). Findings indicate the SWAS yields reliable and valid scores to measure middle grades students' self-efficacy towards writing.

The purpose of the second study was to create and validate a rubric, known as the *Rubric for Scientific Writing* (RSW), which can be used to support writing instruction in science classes and evaluate scientific writing. This rubric assesses both students' general writing skills and their ability to write appropriately within the scientific genre. My findings demonstrate that the RSW produces valid and reliable scores for two factors of students' scientific writing – scientific argumentation and English rhetoric. The RSW has the potential to aid both science teachers who may lack training in the teaching and assessment of writing as well as researchers who need a stable measure of students' scientific writing.

Finally, the third study uses these tools to measure the effectiveness of a writing-to-learn intervention in middle and high school science classes. Prior literature posits that writing-to-learn strategies are less effective for younger students; however, few studies have implemented similar strategies across grade levels. Therefore, this study combines established best-practices to create a writing-to-learn intervention that can be implemented into existing science classes at various grade levels. While high school students did slightly outperform their middle-grade peers, further cluster analysis demonstrated that students who created visuals and used scientific vocabulary during the intervention made the most growth, regardless of grade level.

DEDICATION

In memory of my beloved grandmother, Adela Aleksandra Landau

Chemist, Polyglot, and Survivor

Who never lost her sense of curiosity about the world.

And for my husband, Micah W. Wright, who helped me find the courage to embark on this journey.

ACKNOWLEDGEMENTS

In *The Alchemist*, Paulo Coelho states that “when you want something, all the universe conspires in helping you to achieve it.” This has certainly never been truer for me than it has been while I have worked on this dissertation and degree. I would like to thank all those in my universe who have been part of this achievement. There are more of you than can be listed here, but please know that you have my extreme and humble gratitude – thank you for conspiring in my success.

First, I would like to thank my committee members, who have been much more than simply advisors on the dissertation: Dr. Zohreh Elsami, Dr. Li-Jen Kuo, Dr. Patricia Goodson, and Dr. Sharon Matthews. You have all provided guidance, both in and out of academia, and my success is a direct result of your mentorship. I have learned much from each of you, and I hope you know how much I value your support.

Thank you to the administration, faculty, staff, and students, at the local schools who opened their doors to me. I especially want to acknowledge Jamie Scott and Glenn Finley, two of the most inspiring science teachers I know. Thank you for welcoming me into your classrooms and letting me learn from you.

My appreciation also goes out to the numerous TLAC students who have been both colleagues and friends along this unpredictable path of graduate school – Sunni, Samantha, Beth, Jennifer, Matthew, Megan, Bethany, and many others. I especially want to thank my fellow members of Team MCT – Tracey, Chyllis, Amanda, Diane, and April – for inspiring me to work harder every day.

To my dear friends, Carrie Silver, Sarah Mullin, Reneé Callison, and Amialya Durairaj. You four have been my guideposts and anchors since I was a teenager. Thank you for believing in me when I did not.

To my entire family, for always loving and supporting me as I marched to my own unique beat. Special thanks to my aunts, Barbara Mullin and Anne Wilson, for being strong and inspiring women.

To my committee chair, Dr. Erin McTigue – "thank you" does not even begin to express my gratitude. You have been (and continue to be) the best mentor and role model I could ever have asked for. I will be a better academic, teacher, and person because I have your example to follow.

To my mom, for crawling up to my childhood bunk bed, equipped with *A Wrinkle in Time*, and helping me find a love of reading. It was not until I became a teacher myself that I realized all you did for me, and this dissertation and degree is a testament to all your hard work. Thank you.

And to Micah, for telling me I could do this long before I thought it was possible, for caring enough to give me the feedback I did not always want to hear, and for loving me throughout the entire process. This accomplishment means so much more to me because we have achieved it together. I love you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

In 1988, James Britton, a leading scholar on classroom writing, argued that "when talking, reading, and writing are orchestrated in the classroom in such a way that each can make its unique contribution to a single end, we have surely harnessed language to learning as powerfully as possible" (The National Institute of Education, 1988, p. 6). While reading has certainly taken the dominant role in literacy education, writing research has grown as a field in recent decades. In fact, some go as far as to argue that writing instruction should take precedence over reading as the latter can be accomplished without "conscious comprehension" (Dunn, 2000, p. 169), whereas the writing product provides unequivocal evidence that a literacy task has been completed (Konopak, Martin & Martin, 1987).

Writing in science class allows students a space to build "their own knowledge through questioning, reprocessing, reflecting, analyzing… and drawing conclusions" (Alev, 2010, p. 1343). While some models for writing-to-learn in science class, such as the Science Writing Heuristic (Akkus, Gunel & Hand, 2013), have shown great promise in supporting student content knowledge and critical thinking skills, these effective interventions often require a complete shift in instructional approaches. Unfortunately, by the time students reach high school most of their writing instruction occurs in English class, and little writing occurs in science class (Graham & Harris, 2012). This dissertation describes three connected studies which address three critical issues in writing-to-learn research: a) how to measure students' feelings about writing, b) how to assess scientific writing, and c) how to integrate writing-to-learn into current secondary science curriculum.

Chapter II describes the development of the Student Writing Affect Survey. The effects of writing-to-learn on student achievement have been inconsistent, and the reasons for this phenomenon remain unclear (Klein, 1999). Often researchers focus on the age and grade of their participants, factors which have shown to impact writing-to-learn interventions (Bangert-Drowns, Hurley & Wilkinson, 2004), but students' feelings towards writing are not often described in existing literature. If a child is not motivated to write, has a negative attitude towards the writing process, or does not feel he or she is a proficient writer, their writing progress will likely be hindered, decreasing the impact of writing-to-learn. A child's motivation to and self-efficacy for writing can vary based upon the topic and situation, and therefore affective measures need to be utilized to help determine whether poor writing is a result of underdeveloped skills, lack of background knowledge, or poor affect. Therefore, the goal of the first study was to create and validate an instrument to measure writing affect and to examine how students' motivation, attitude, and self-efficacy mediate or moderate the impact of writing achievement.

The second study (Chapter III) focuses on the creation of a rubric that can be used to score scientific writing samples across grades and subject foci. The methods of scoring writing achievement in science class vary both in research and classroom practice, making cross-study comparisons difficult. Many researchers and teacher find it difficult to evaluate both the written conventions of student work as well as the overall scientific content and therefore prioritize one construct to the detriment of the other. Thus, I created a writing rubric aligned with research-based best practices for scientific writing, Common Core writing anchors (National Governors Association, 2010), and Next Generation Science Standards (NGSS Lead States, 2013). My goal was to develop a rubric that can be used to score scientific writing samples across grades and subject foci. Chapter III discuss both the development and mixed-methods validation of this rubric, which will become a stand-alone publishable measure for other researchers and teachers to use. This rubric provides a tool for science teachers who may not feel prepared to score student writing (Graham, Harris, Fink & MacArthur, 2001) as well a system for future researchers to score scientific writing using a universal metric.

Finally, using the instruments developed in the first two studies, in Chapter IV I designed and measured the impact of a quasi-experimental writing intervention. While research in the area of writing within content classes is limited, convergent evidence indicates that writing is most successful when students a) have multiple opportunities to practice writing (Miller, 2014), b) use evidence to form arguments (Klein & Rose, 2010) and c) write for authentic audiences (Choi, Notebaert, Diaz & Hand, 2010). This intervention study aimed to combine these key aspects of writing-to-learn instruction to create a viable (i.e., low time and training costs) writing intervention that can simultaneously support students' content knowledge development and scientific writing skills. My goal was to develop strategies for integrating writing-to-learn with existing instructional practices, so the success of the intervention can be easily replicated by classroom teachers.

Theoretical Framework

This research is grounded in Bereiter and Scardamalia's (1987) concept of *knowledge telling* versus *knowledge transforming* writing. According to the authors, *knowledge telling* writing consists of using existing knowledge to report information, which constitutes most of the writing students do for science class (Choi, et al., 2010). My goal is to move students from *knowledge telling* to *knowledge transforming* writing, which involves using writing to engage in a self-interaction that can build new understandings (Bereiter & Scardamalia, 1987). That is, whereas sociocognitive theory posits that knowledge is built through social interactions with others (Vygotsky, 1980), the process of writing can serve as a self-interaction and help an individual organize, reformulate, and essentially transform existing knowledge into new concepts.

By using writing to transform knowledge, students are engaging in authentic scientific literacy. Science literacy can be conceptualized under two distinct definitions: the *fundamental sense* and the *derived sense* (Norris & Phillips, 2002). The fundamental sense refers to an individual's ability to read and write when the subject is science, such as reading a scientific article and composing a response. In the derived sense, by contrast, science literacy refers to being knowledgeable and informed about science. However, a successful scientist's literacy skills must be strong in both domains, as there is a deep-seated relationship between scientific texts, literate thought, and scientific literacy (Norris & Phillips, 2002). I argue that this relationship implies that improving students' scientific literacy in the *fundamental sense*, and thus developing stronger readers and writers in science class, will make them better able to acquire the content necessary to become scientifically literate in the *derived sense.*

Operational Definitions

Before diving into the details of each study, I will define several key constructs to ensure consistency.

General Writing

In its simplest form, writing involves recording thoughts and ideas on paper (Miller, 2014); however writing for effective communication requires knowledge of and adherence to specific forms and conventions. For the purpose of this study, *general writing skills* refers to those most commonly highlighted in English/Language arts class and emphasized on standardized tests of academic writing (e.g., SATs), including adherence to English conventions, organization of writing, and writing style. These skills are necessary for effective written communication in all genres.

Scientific Writing

Within the field of science, students must be able to create sound connections between questions, claims, and evidence (Akkus, et al., 2013) – skills that are not often addressed in English/language arts class. The ability to write effectively in science falls under the larger instructional umbrella of disciplinary literacy (Shanahan & Shanahan, 2008). Unlike the more generic content-area literacy, disciplinary literacy focuses on the unique aspects of literacy within a genre. Therefore, *scientific writing skills* are understood to be the ability to make an effective scientific argument in writing – a disciplinary literacy skill.

Affect for Writing

Not only must students have the skills and background knowledge to write, they must also have some sort of intrinsic or extrinsic force driving them to complete the task. *Affect for writing* includes students' motivation, attitude, and self-efficacy for writing that support or hinders the writing process.

Content Knowledge

As established by Graham (2006), an author's knowledge of the content to be described in writing greatly impacts the overall quality of writing. Furthermore, the critical thinking required for writing is domain specific – writing skills without content knowledge will not contribute to learning (Willingham, 2007). For this research, *content knowledge* refers to students' general understanding of science as well as their comprehension of the specific scientific concepts being taught.

CHAPTER II

DEVELOPING THE STUDENT WRITING AFFECT SURVEY: A MEASURE OF ADOLESCENTS' SELF-EFFICACY TOWARDS WRITING

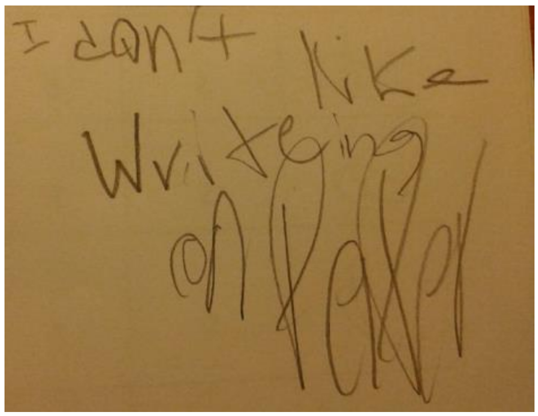


Figure 2.1. I don't like Writeing on Peper [sic], Student comment spontaneously included on writing affect survey

The sentiment expressed by a middle school student in Figure 2.1 would not surprise most teachers. In order to be effective writers, students must have an intrinsic or extrinsic force driving them to complete the task. A child's motivation to write can vary based on the topic and situation, and therefore affective measures need to be used to help determine whether poor writing is a result of underdeveloped skills and knowledge, or whether there are affective issues at play (e.g., self-efficacy, attitude, motivation). Motivation can help students persevere through difficult literacy activities (Fulmer & Frijters, 2011), and support struggling students with the more demanding practice of writing (Harris & Graham, 2013). Furthermore, poor motivation for literacy activities can possibly predict those who may be heading on a downward trajectory (Johnston & Costello, 2005), thus allowing teachers to intervene with at-risk students.

The manner that writing is currently addressed in the adolescent grades does not always support positive motivation and self-efficacy for writing. At the early grades, writing focuses primarily on mechanics, allowing students to write on topics of their choice. By the time students enter middle school, this autonomy of topic diminishes as the focus shifts to content-area writing, requiring students to learn specific skills to prepare for high school and college writing tasks (Graham & Harris, 2012). Moreover, in primary grades, students are encouraged to write freely and often; however, in secondary grades, most students complete very little extended writing as well as minimal writing requiring analysis and critical thinking (Graham & Harris, 2012). This focus has created an environment where, rather than being an expressive art form, writing becomes a task to be completed.

At the same time that writing becomes less intrinsically enjoyable for students, it is also often emphasized as part of test preparation. In many states, writing is only tested in selected grade levels, in comparison to reading and math which are assessed every year. Because of the growth of the value-added movement (Chetty, 2012; 2014) and the tendency for schools to be "punished" for failing scores (Heilig & Darling-Hammond, 2008), educators may be focusing on tested content at the expense of other academic material. This is causing writing to often be over-emphasized in tested grades and neglected in others, rather than allowing for yearly growth and increased sophistication called for by the Common Core state standards (National Governor's Association, 2010).

This varying pressure and absence of writing may be impacting students' views of writing and their own writing skills, and the literacy research field requires tools to measure and monitor this effect.

The purpose of the current study is to develop and validate a measure of students' self-efficacy toward writing. This measure can be used to monitor how school interventions impact students' motivation and attitude toward writing, as well as identify variables that mediate and moderate student achievement. In the following sections, I outline the theoretical framework underpinning this study and describe existing research in this vein. Drawing from prior research and social cognitive theory, I developed the *Student Writing Affect Survey* (SWAS) to specifically measure writing affect. I then worked to validate the SWAS through multiple types of factor analyses, as well as establish reliability coefficients for the instrument scores. Specifically, this study seeks to answer the following questions:

1. What aspects of students' affect toward writing are measured by the *Student Writing Affect Survey?*
2. How reliable are the scores produced by the *Student Writing Affect Survey*?
3. How valid are the constructs measured by the *Student Writing Affect Survey*?

Theoretical Framework: Affect, Agency, and Learning to Write

This study takes a social cognitive perspective to understand why students may be more or less engaged in writing tasks, and therefore how to measure that engagement. Social cognitive ideals have supported the development of a variety of theories to explain student engagement in learning, including Expectancy-Value Theory (EVT, Eccles & Wigfield, 2002) and attitude acquisition (Fishbein & Ajzen, 1975). To date, much of the published research related to motivation and literacy does not explicitly define target constructs, which impacts how these constructs are investigated and interpreted (Conradi, Jang, & McKenna, 2014). Therefore, I am dedicating time here to thoroughly define my constructs as they relate to engagement in writing. In the following section, I detail aspects of social cognitive theory as they relate to writing engagement, specifically the concept of *human agency.* Additionally, I draw connections to other, complementary, theories of learning engagement. Finally, I apply these theories to help explain how a child's experiences with writing influence his or her decisions to engage or avoid the task.

Under a social cognitive framework, human actions are simultaneously influenced by the environment as well as constantly changing their environment to make it suit their needs and desires (Bandura, 1986). In this perspective, human behavior is viewed as a result of the interaction between personal factors and the environment. That is, humans are not simply passive beings waiting for something to happen *to* them, they are active agents who make choices and engage in activities to achieve specific goals. This human agency is at the center of the decision-making process (Bandura, 2001), and an understanding of agency is essential to examine why children choose, or not, to engage in writing activities.

At the core of human agency are four features: intentionality, forethought, self-reactiveness, and self-reflectiveness. These features are generally sequential – that is, an individual making a choice will go through each of these phases in order. When applied to writing, *intentionality* implies that students do not involve themselves by accident – they are motivated by the idea that writing will help them achieve a specific goal (Bandura, 2001). According to Expectancy-Value Theorists, motivation is closely tied to an individual's expectancies (or how one feels he or she will do on a task) and values (the reasons why someone chooses to engage in a task) (Eccles & Wigfield, 2002). Combining Expectancy-Value Theory (EVT) with Bandura's beliefs surrounding intentionality, an individual must believe something will be gained by writing tasks, and that he or she is capable of realizing this reward, in order to intentionally engage in writing tasks.

Forethought involves intentionally choosing a specific path of action to best achieve the desired results. An individual's objectives may not be in their long-term best interest (as is demonstrated by a student avoiding a writing assignment), however they can still develop a plan to meet their perceived needs. This means students have the capacity to examine a task (such as a writing assignment) and confidently develop a plan for completing (or avoiding) the task. Bandura (2001) argues that even when faced with competing influences, humans have a tendency to weigh the options and select the path that will best suit their desired objectives.

In addition to intentionally developing plans for action, personal agency involves *self-reactiveness*, or the ability to execute a self-created plan (Bandura, 2001). This feature of personal agency also involves the ability to make changes along the way to achieve the individual's goals, which are rooted in a personal value system. After each experience, individuals engage in the *self-reflectivness* stage of agency and, essentially, decide whether or not the effort was worthwhile.

These features of personal agency help explain why students engage in or avoid writing tasks, and suggest that students must be motivated in order to complete the task effectively. Bandura (2001) argues that the responsibility of personal agency can be exhausting, leading some to feel that the effort was not worth the reward:

> There is an onerous side to direct personal control that can dull the appetite for it. The exercise of effective control requires mastery of knowledge and skills attainable only through long hours of arduous work. Moreover, maintaining proficiency under the ever-changing conditions of life demands continued investment of time, effort, and resources in self-renewal. (p. 13)

Learning to write well is a task that requires “long hours of arduous work” and “continued investment” in its practice, and attitude acquisition theory can help explain why students would develop negative feelings towards this task. Attitude is learned and predisposes consistent, predictable actions (Fishbein & Ajzen, 1975). Therefore, if a young child has a negative experience with writing, he or she will learn to dislike the activity and likely choose to not invest the arduous practice needed for mastery. Additionally, as social cognitive theory suggests that human actions are simultaneously influenced by the environment as well as constantly changing their environment to make it suit their needs and desires (Bandura, 1986), a child who has not felt successful in writing may alter their environment to avoid the task. This could be accomplished indirectly by associating with peers who also avoid writing, or directly, by engaging in other activities during potential writing times.

Furthermore, social cognitive theory posits that individuals can learn through second-hand experiences of their peers or those in "symbolic environments" like television (Bandura, 1986, xxi). This learning is not limited to academic content – individuals can vicariously acquire attitudes or beliefs by observing models in person or in the media. For example, if a child spends time with someone else who does not like writing, or perhaps sees writing portrayed as something negative on television, he or she can learn to dislike the task without having personal negative experiences. On the other hand, if a child spends time with someone else who enjoys writing, or if the child sees writing portrayed as inspirational, he or she can learn to enjoy the task of writing.

According to Bandura (2001), "among the mechanisms of personal agency, none is more central or pervasive than people's beliefs in their capacity to exercise some measure of control over their own functioning…" (10), implying that a sense of autonomy allows individuals to develop their personal agency. Bandura (1986) even goes so far as to define freedom as "the exercise of self-influence" (p. 39). This viewpoint is supported by other researchers, arguing that humans have an instinctual need for competence and personal causation (Eccles & Wigfield, 2002). When learning to write in school, students are not frequently given options – they have to complete a writing task to avoid a negative consequence (poor grades). With this vital component of agency no longer at play, researchers must examine how other aspects of agency and affect are influencing student writing development.

Certain intervention strategies tend to have more positive effects on reading motivation development (see Wright, Hodges, & Franks, 2015), and it can therefore be assumed that a similar effect exists for writing. A deeper understanding of students' affect towards writing in the context of educational interventions will help identify best practices for both teaching and research. In essence, there are many factors influencing an individual's feelings towards writing. A child who has had negative experiences will likely develop a poor attitude towards writing (Fishbein & Ajzen, 1975) and possibly believe that there is little value in the task (Eccles & Wigfield, 2002). Additionally, external factors, such as the level of writing autonomy and the experiences of peers, can impact how a student feels about writing and possibly predict his or her writing achievement. However, without a valid and reliable tool for measuring students' affect towards writing, these effects cannot be quantified.

Established Writing Affect Measures

While I consulted many related tools in the development of the SWAS (see Table 1.1), and despite the importance of affective factors related to writing, published measures specific to writing affect are limited. Kear, Coffman, McKenna, and Ambrosio (2000) published the *Writing Attitude Survey*, which was normed for students in first through 12th grade. This measure, modeled after McKenna and Kear's (1990) *Reading Attitude Survey*, presents a cartoon character expressing different emotions (from very happy to very upset), and asks the student to indicate how certain statements about literacy would make them feel. Items were developed using existing instruments and by reviewing college-level textbooks used in language arts methods classes. However, the *Writing Attitude Survey* focuses solely on writing attitude, requiring a broader analysis to fully explain students' writing behaviors.

The *Writing Activity and Motivation Scales* (WAMS) (Troia, Harbaugh, Shankland, Wolbers & Lawrence, 2013) is a relatively new Likert-style questionnaire that has been administered to students in grades four through 10. This 30-item questionnaire measures six facets of writing motivation: self-efficacy, success attribution, task interest/value, mastery goals, performance goals, and avoidance goals. This measure also contains 10 items designed to measure how frequently students engage in writing activities in a month, and the researchers used this tool to correlate writing behaviors with motivation. The theoretical model of the WAMS is closely aligned to that of the present study; the creators even argue that there is a need for a writing motivation scale that honors the multidimensional nature of motivation. Unfortunately, the reliability estimates for many of the measures were quite low (less than .60 in some instances), suggesting that items may not be measuring their intended construct (Troia, et al., 2013).

Affective constructs are situation-dependent (Guthrie et al., 2006b), meaning that a student's feelings towards academics will vary depending upon the skill or subject area being examined. Therefore, an affect measure specific to writing is required to understand the relationship between students' feelings towards writing and overall writing performance. While other tools measuring literacy and academic affect do exist (for instance, see Hodges, McTigue, Weber, Douglas, & Wright, 2015; Steinmayer & Spinath, 2009; Wigfield & Guthrie, 1995), a valid and reliable measure of adolescents' self-efficacy towards writing is still needed. Thus, the goal of this study was to develop a theoretically-sound instrument that reliably measured the writing self-efficacy and affect for students at this critical age.

Methods

In the following sections I outline briefly the instrument development, my participant selection and administration procedures, and the statistical analyses I conducted.

Instrumentation

After examining the existing tools for measuring writing affect, I decided to use the *Motivation for Reading Questionnaire* (MRQ, Wigfield & Guthrie, 1995) to begin instrument development. While this instrument does not pertain to writing specifically, it provides a theoretically supported model for assessing student motivation. This questionnaire includes 53 items with four choices per item in which students rate the extent that items were similar to or different from them. This tool has been repeatedly used to measure students' motivation towards reading, and consistently produces valid and reliable scores (e.g., Guthrie, Hoa, Wigfield, Tonks, & Perencevich, 2006a; Guthrie, Wigfield, & VonSecker, 2000; Mason, 2004). Additionally, Kear and colleagues (2000) found that using a structure similar to their reading attitude survey produced similar results in the creation of a writing attitude survey. Therefore, I decided to use the structure of the MRQ in the initial creation of the SWAS.

I collected several existing instruments that measure aspects of reading and writing affect as well as those measuring academic affect generally (see Table 2.1). Working with another researcher, I rephrased items from the instruments to directly relate to writing. For example one item from Wigfield and Guthrie's (1995) MRQ asks participants the degree to which they agree with the statement "I know I will do well in reading next year". This question was rephrased as "I know I will do well in writing next year". The resulting item bank consisted of 172 possible items. I removed redundant items and those which would not be applicable to classroom research. For instance, the *Preservice Teacher Self-Efficacy for Writing Inventory* (PT-SWI), measures writing self-efficacy but was designed for undergraduate preservice teachers (Hodges, et al. 2015). Therefore, items such as "effective teachers must be proficient at writing" were removed.

The developed instrument contained 41 items aimed to measure writing affect. Using the MRQ as a model, each of the items provided a statement about writing and asked the students to rate statements on a scale of 1 to 4. Selecting 1 indicates that the statement is "very different from me" and selecting a 4 indicates that the statement is "a lot like me". The option of selecting a neutral response was intentionally avoided with an even number of options, as is the norm with literacy affect surveys for children (e.g., McKenna & Kear, 1990; Wigfield & Guthrie, 1995).

Procedures

I conducted two separate administrations of the SWAS to allow me to conduct different statistical analyses. In both instances, the administration time was approximately 15 minutes.

Table 2.1

Existing measures referenced to create Student Writing Affect Survey

Measure	Citation	Description
Elementary Reading Attitude Survey	McKenna & Kear, (1990)	Likert-style survey of Elementary students' attitude towards reading
Motivation for Reading Questionnaire	Wigfield & Guthrie (1995)	Likert-style survey of students motivations for reading
Writing Attitude Survey	Kear et al., (2000)	Much like the Elementary Reading Attitude Survey, but rephrased to measure students' attitudes towards writing
Scales for the Assessment of Learning and Performance Motivation (translated from German title)	Steinmayer & Spinath (2009)	Likert-style scale assessing affective and personality factors that may impact students' academic performance
Writing Activity and Motivation Scales	Troia et al., (2013)	Survey which measures both students motivation for writing and writing behaviors
Preservice Teacher Self-Efficacy for Writing Inventory	Hodges, et al., (2015)	Survey evaluating preservice teachers' self-efficacy for writing and for teaching writing

Administration 1. I first administered the SWAS in January of 2015 at a large middle school (grades six through eight), in the southwest United States. Teachers at the school reported that the English curriculum in grades six and eight focused on reading comprehension with little writing instruction, whereas seventh grade emphasized writing nearly as much as reading in order to prepare students for the state writing exam at that grade. The school represented a diverse sample of students. The majority (75%) of students was eligible for free or reduced lunch, and approximately 10% of the students were identified as English Language Learners. Additionally, this school housed the district's middle school gifted and talented programs, representing students working above grade-level expectations. In total, 517 of the more than 800 students enrolled provided assent and completed the survey. Table 2.2 shows the basic demographic information collected on these students.

Table 2.2

Participant Descriptive Information (Group 1)

Total	517	Percentage
Males	225	43.5%
Females	220	42.5%
Gender not reported	72	13.9%
6th Grade	192	37.1%
7th Grade	156	30.1%
8th Grade	150	29.0%
Grade not reported	19	3.6%

After obtaining both university and local school-level Institutional Review Board approval, I sent notices home to parents; however, as the SWAS did not examine sensitive matters, collect identifying information, or place the students at risk for harm, individual parental permission was not required. I administered paper-based versions of the SWAS during a time agreed upon by school officials as being least intrusive. The students surveyed included all regular and special education populations present on the day of administration.

Each copy of the survey had a student assent form stapled to the front which informed students that their participation was voluntary. This form was read aloud by classroom teachers and students who wished to participate provided their initials on the form. The entire survey was then read aloud to students to ensure that results were not impacted by students' reading ability. All surveys were collected at the end of the period, however only those from students who initialed the assent form were retained.

Administration 2. To establish external validity, I administered the SWAS to a second group of participants. This second group consisted of 53 students in grades 6 through 11 who attended a private independent school in the same area as the other participants. While this group differed greatly from the other sample in socio-economic status, the student body is comprised of over 23 different nationalities, thus representing a different type of diverse population.

In addition to completing the SWAS, the students' teachers were asked to provide ratings of their students' attitude towards writing, self-confidence for writing, self-efficacy for writing, and tendency to avoid writing. Descriptive information for this second group is available in Table 2.3.

Measurement Validity

Following the model of Brackett, Reyes, Rivers, Elbertson, and Salovey (2012), I conducted the validity procedures by splitting the first sample at random to conduct both exploratory (EFA) and confirmatory factor analyses (CFA). Multiple EFA and CFA models, based on theory and psychometrics, were tested to identify the model that best fit the data. Additionally, I conducted a higher-order factor analysis to establish that the factors were related and measured one overarching, latent construct (Thompson, 2004). To establish external validity, I correlated the second groups' self-reported scores with their teachers' observational scores. Finally, writing samples from the second group's science journals were collected to examine the relationship between students' self-reported data and their writing behaviors.

Table 2.3

Participant Descriptive Information (Group 2)

Total	53	Percentage
Males	32	62%
Females	21	38%
6^{th} Grade	18	34%
7^{th} Grade	7	13%
9^{th} Grade	1	2%
10^{th} Grade	13	25%
11^{th} Grade	14	26%

Results

In the following sections, I outline my results by statistical analysis. First, I describe the EFA models and results. Next, I explain the results of my CFA models. Finally, I compare the second group's self reported scores with teacher reported measures and student writing behaviors.

Exploratory Factor Analyses

The first group of students was randomly split into two groups ($n1$ = 258 and $n2$ = 251, respectively) using SPSS software. For the EFAs, I used the $n1$ group. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) was .918, and the Bartlett sphericity test was less than .000, indicating I could reject the null hypothesis (the items were not correlated) and thus proceeded with a factor analysis. I conducted multiple

Principal Components Analysis (PCA) with Varimax rotation. Consulting writing theory and psychometric principles, I tested seven EFA models to determine which best fit my data (see Table 2.4) starting with data-driven models and moving to theoretically based models. A detailed summary of each EFA follows.

Table 2.4

Summary of EFA Models

EFA #	Description	Number of Survey Items	Number of Factors	Mean items per factor	Percent Variance Explained	Mean Factor Reliability
1	Eigenvalues over 1	41	8	5.12 (4.91)	59.96	.69 (.194)
2	Repeat EFA #1, removing items with structure coefficients <.5	31	6	5.16 (4.26)	59.68	.679 (.194)
3	Forced 3 factors	41	3	13.67 (7.50)	44.31	.84 (.10)
4	Repeat EFA #3, removing items with structure coefficients <.5	29	3	9.67 (5.03)	52.74	.84 (.09)
5	Forced 5 factors	41	5	8.2 (5.16)	52.04	.78 (.10)
6	Repeat EFA #5, removing items with structure coefficients <.5	30	5	6 (4.06)	58.16	.74 (.15)
7	Repeat EFA #5, five items removed	36	5	7.2 (3.56)	55.94	.79 (.09)

Note: When appropriate, standard deviations are displayed next to means in parenthesis

EFA #1. I first examined the model including all factors with eigenvalues over one. This structure yielded eight factors and explained 59.96% of the total variance (See

Table 2.4). However, many of the factors had few items, and nearly half of the variance explained (28.5%) was accounted by the first two factors.

EFA #2. I repeated the first EFA analysis removing any survey items which did not have a structure coefficient of at least 0.5 on any of the factors. This resulted in the removal of 10 items, and yielded a model with six factors explaining 59.68% of the variance. Upon examining the items in each factor, I discovered that there was little theoretical foundation for the factors, so I proceeded to a theory-based model.

EFA #3. Since I originally hypothesized that the SWAS would measure motivation, attitude, and self-efficacy, I conducted a PCA analysis with Varimax rotation forcing the items into three factors. This initial model, utilizing all 41 original items, only explained 44.31% of the variance, and many items had structure coefficients below 0.5.

EFA #4. Repeating EFA #3, I removed 12 items with structure coefficients less than 0.5. This resulted in a model which explained 52.74% of the variance, however the items did not factor as expected. For instance, items that theoretically measure self-efficacy, such as *I can write good papers because writing is easy for me* and *I feel confident in my overall writing abilities,* correlated on the same factor as items that theoretically relate to attitude such as *I like to write*.

EFA #5. Examining the results of EFA #1 revealed that 77% of the total variance explained was explained by the first five factors. I therefore hypothesized that a five-factor model could be a strong fit for the data. Thus, I conducted a PCA with Varimax rotation forcing the items into five factors. This model explained 52.04% of the variance, but there were still a number of items with low structure coefficients.

EFA #6. Next, I removed the 11 items from EFA #5 that had low structure coefficients and repeated the analysis. This explained 58.16% of the variance, and I found that theoretically similar items aligned on the same factor. However, as score reliability is heavily influenced by the number of items (Thompson, 2003) removing so many items from the analysis lowered the reliability of each factor.

EFA #7. Finally, I examined the items removed in each iteration and discovered that five items did not factor well in any of the models. I therefore removed these five items, repeated the five factor EFA, and the resulting model explained 55.94% of the total sample variance while maintaining reasonable reliability estimates. Furthermore, examining the items in each factor revealed a strong, theoretical foundation for the factors (further described below).

After removing the five consistently low items, the SWAS included 36 items and yielded a Cronbach's α of .945. Based on the results of the seven EFAs, the model best fitting the data contained five factors explaining 55.94% of the total sample variance. The descriptive properties of the five factors are detailed in Table 2.5.

Higher Order Factor Analysis

Because the model demonstrated correlation between the factors, I conducted a higher-order factor analysis to better describe the relationship between the factors. These procedures confirmed that the five factors were related and measured one overarching latent construct (Thompson, 2004). I named this overarching construct *writing self-*

efficacy, focusing specifically on five factors which contribute to self-efficacy. The higher-order factor model explained 62.651% of the variance for the EFA sample (see Table 2.6).

Table 2.5

Factors measured by SWAS

Factor	Items *n*	Eigenvalue	Total Variance Explained (%)	Cronbach's α
Attitude Towards Writing	13	6.6	18.35	.916
Value of Writing Achievement	8	5.23	14.53	.846
Social Persuasion for Writing	4	2.79	7.77	.683
Confidence in Writing	6	2.49	6.93	.776
Writing Avoidance	5	3.00	8.35	.733
Overall Self-Efficacy	36		55.94	.944

Table 2.6

Higher-order factor pattern coefficient matrix (EFA Sample)

First-Order Factor	Pattern Coefficient (*A*)	h^2
Attitude Towards Writing	.894	.798
Value of Writing Achievement	.822	.675
Social Persuasion for Writing	.802	.644
Confidence in Writing	.837	.700
Writing Avoidance	.561	.315

Confirmatory Factor Analysis

Using the CFA subgroup (*n2* = 251*)* of students, I conducted a confirmatory factor analysis using AMOS software, and examined the model fit indices to determine how well the models fit the data. I analyzed the EFA #7 model, and it continued to be a strong fit. I examined both absolute (i.e., Chi square divided by degrees of freedom test;

Root Mean Square Error of Approximation) and relative (i.e., Comparative Fit Index) model fit indexes. I compared results to the benchmarks established by Meyers, Gamst, and Guarino (2013) to describe the goodness of fit for each model (see Table 2.7). I also repeated the higher order factor analysis, using the CFA sample, which explained 61.157% of the variance (see Table 2.8). Because model #7 with the higher order factor demonstrated an acceptable goodness of fit, explained a noteworthy amount of the overall variance, and had a strong theoretical foundation in existing literature, this was retained as the best model for this data.

Table 2.7

Confirmatory Factor Analyses Model Fit Indexes

Model #	χ^2/df	RMSEA	CFI
7	1.88 (good)	.059 (acceptable)	.864 (adequate)
CFA #7 with Higher Order Factor	1.98 (good)	.063 (acceptable)	.845 (adequate)

Table 2.8

Higher-order factor pattern coefficient matrix (CFA Sample)

First-Order Factor	Pattern Coefficient (*A*)	h^2
Attitude Towards Writing	.895	.784
Value of Writing Achievement	.811	.657
Social Persuasion for Writing	.791	.626
Confidence in Writing	.835	.697
Writing Avoidance	.542	.294

External Validity: Teacher Ratings and Student Behaviors

Using the best fit model, I analyzed the results of the second SWAS administration ($n = 53$) on a new sample of students recruited from a local independent school with unique demographics. The reliability coefficients for each factor are consistent with the first administration indicating that across diverse samples, the instrument yields reliable scores (see Table 2.9).

Table 2.9

Score reliability for second SWAS administration (n = 53)

Factor	Cronbach's α
Attitude Towards Writing	.918
Value of Writing Achievement	.829
Social Persuasion for Writing	.649
Confidence in Writing	.804
Writing Avoidance	.746
Overall Reliability	.945

To consider external validity for the SWAS, the students' English teachers were asked to rate each child on four aspects of writing which corresponded with four of the factors identified by the SWAS: attitude towards writing, self-confidence in writing, overall self-efficacy in writing, and writing avoidance. The teachers were provided with brief definitions of these constructs as well as a rubric for scoring students in each construct (see Figure 2.2).

I calculated the Pearson's r two-tailed correlations between the teacher ratings and student self-reported scores on the SWAS (see Table 2.10). All items measuring the same construct (i.e., students' reported attitude towards writing and teachers' rating of writing attitude) were statistically significant at the .01 level. This indicates that measures of individual factors on the SWAS match teacher perceptions of students' affective constructs.

Attitude towards writing - A student's *general* tendency to respond favorably or unfavorably to writing tasks. A student with a high attitude score will react positively when presented with a writing task, while his peer with a negative attitude will likely groan and complain about the task

1 – Very negative attitude	2 – Slightly negative attitude	3 – Slightly positive attitude	4 – Very positive attitude

Self-Confidence in Writing - A student's perception of his or her writing abilities.

1 – Very unconfident	2 – Somewhat unconfident	3 – Somewhat confident	4 – Very confident

Self-efficacy in writing – A student's belief that he or she can accomplish a writing task, even when it is difficult. A student with high self-efficacy will not become frustrated and persevere when challenged, whereas a student with low self-efficacy is likely to give up.

1 – Very poor self-efficacy (i.e., gives up very quickly)	2 – Somewhat poor self efficacy (i.e., will make an effort before giving up)	3 – Somewhat strong self-efficacy (i.e., will seek some help or resources before giving up)	4 – Very strong self-efficacy (i.e., advocates for self, asks for help, and takes other steps to not give up)

Writing avoidance – A child's tendency to avoid writing activities (e.g., attempt to work in groups, not complete writing assignments, choose alternative projects, etc.)

1 – Almost always tries to avoid writing tasks	2 – Frequently tries to avoid writing tasks	3 – Occasionally tries to avoid writing tasks	4 – Almost never tries to avoid writing tasks

Figure 2.2. Teacher rubric for rating students' affect towards writing

While teacher-reported data provides evidence of external validation for the SWAS, it is also important to also examine student behaviors in relation to their scores on the measure. In order to further validate the SWAS, I completed word counts of approximately 10 of each student's journal entries. These writing samples had been collected as part of a larger intervention, and required students to summarize what they learned that day in science class to a hypothetical friend who was absent. I calculated correlations between students' average number of words written per assignment with their scores on the SWAS (see Table 2.11). While word count is not a measure of writing quality, a student who is highly motivated to engage in writing tasks would likely produce more text for this low-stakes writing activity than his or her peers.

Table 2.10

Correlations between student self-report on SWAS and teacher ratings

	Teacher Ratings			
Student Self-Report	Attitude towards Writing	Self-Confidence for Writing	Writing Avoidance	Self-Efficacy for Writing
SWAS Attitude	.403**	.374**	.371**	.319*
SWAS Confidence	.468**	.554**	.462**	.409**
SWAS Avoidance	.406**	.282*	.529**	.465**
SWAS Overall Self-Efficacy	.518**	.464**	.523**	.443**
SWAS Value	.561**	.418**	.565**	.482**
SWAS Social	.332**	.332**	.327*	.220
Average (SD)	.448 (.084)	.404 (.097)	.463 (.095)	.389 (.101)

** Correlation is significant at the .05 level (2-tailed)*
*** Correlation is significant at the .01 level (2-tailed)*
Note. Underlined correlations indicate those items measuring the same constructs.

Similar to the teacher ratings, student behaviors had modest, but statistically significant, correlations with their self-reported feelings toward writing on the SWAS. The small sample size of this group (n = 53) combined with the statistically significant findings indicates that these modest correlations are likely real effects and not a result of sampling error (Thompson, 2006).

Table 2.11

Correlations between average number of words written per journal entry and student self-reported scores on SWAS

	SWAS Attitude	*SWAS* Value	*SWAS* Social	*SWAS* Avoidance[a]	*SWAS* Confidence	*SWAS* Overall
Average # of words	.389*	.337*	.325**	.374**	.287*	.424**

* *Correlation is significant at the .05 level (2-tailed)*
** *Correlation is significant at the .01 level (2-tailed)*
[a] *Items for this factor were reverse coded so a higher score indicates a student who is not likely to avoid writing assignments.*

Discussion

Based on my results, I found three major themes: (1) the SWAS is measuring five aspects of self-efficacy toward writing; (2) the SWAS produces reliable scores for writing affect; and (3) external validity analyses show that the SWAS is capturing true elements of writing affect. Considering the three research questions, I describe my conclusions in the following sections.

Affect Measured by the SWAS

To answer the first research question (*What aspects of students' affect towards writing are measured by the Student Writing Affect Survey*), I examined seven EFA and CFA models. The model that best fit the data for this study was a higher order factor analysis measuring students' self-efficacy for writing. This model explained an average of 61.904% of the overall variance, and aligns with relevant theories of self-efficacy. Figure 2.3 shows this model and the resulting factors.

Factor	Description	Sample *SWAS* Item
Attitude Towards Writing	A child with a positive attitude towards writing will respond favorably when presented with a writing task, while his peer with a negative attitude will likely groan and complain about the task	I think it would be fun to be an author who writes books
Value of Writing Achievement	A child who values writing does not necessarily enjoy the task, but understands that it is important.	I feel most successful if I see that my writing has really improved
Social Persuasion of Writing	A child who scores highly on this factor will likely work hard to make sure his or her writing is coherent and clear and will enjoy sharing work with others.	I like when my classmates read something I wrote
Writing Avoidance	A child's tendency to avoid writing activities. A child with a high score on this factor will not tend to avoid writing activities.	I don't like having to rewrite my paper
Confidence in Writing	A child's perception of his or her writing abilities.	When writing a paper, it's easy for me to decide what goes 1st, 2nd, 3rd, and so on

Figure 2.3. SWAS Factors

According to Bandura's (2001) definitions, self-efficacy originates from four sources: (1) interpreting the results of previous performance, (2) models and observing others, (3) social persuasions, and (4) emotional states. The first of these sources is highly related to attitude, which is a "predisposition to respond in a consistently favorable or unfavorable manner" based upon previous experiences (Fishbein & Ajzen, 1975, p. 6). Therefore, the factor *attitude towards writing* suggests that how students have performed in the past affects their belief about future writing achievement.

Bandura's (2001) second source of self-efficacy comes from existing models and observations of others. As children can learn from the experiences of others (Bandura, 1986), if a child has seen someone else succeed in (or as a result of) writing, he or she will likely value writing and see a purpose to developing strong writing skills. This aligns with the second factor, *value of writing achievement*. Items in this factor measure how important a child feels writing skills are and how he or she values improving his or her writing ability.

The third source of self-efficacy, social persuasions, define the inherent social pressures that encourage someone to engage in an activity. As writing is an expressive task meant to communicate ideas with others, it is inherently a social activity even when the audience is unknown. This third factor, *social persuasion for writing*, measures how much others' opinions of a student's writing influence that student's efforts. Finally, a student's *confidence in writing*, or emotional states, affects self-efficacy. The items in this fourth factor measure a child's beliefs about his or her writing abilities. This factor differs from attitude, because a child may not *enjoy* the task but still believe he or she will be successful at writing.

One final factor emerged from the results of the SWAS – *writing avoidance*. Items from this factor asked students about their writing habits and whether or not they chose to engage in writing activities. Bandura's (2001) theory of self-efficacy posits that humans are constantly impacting their environment, and thus avoidance would be one method for making an environment more favorable for a student who does not enjoy writing. This factor may help researchers and teachers identify students at-risk for writing failure as those avoiding the task will likely not improve their skills and continue to have negative experiences with writing.

Reliability of Scores for the SWAS

Reliability is a necessary precondition for score validity (Thompson, 2003). Reliability can vary drastically from one administration to another (Viswesvaran & Ones, 2000), and therefore, coefficients from multiple administrations must be examined separately. To establish the reliability of the scores produced by the SWAS (and answer the second research question), I examined both the individual factors along with overall reliability coefficients from each administration.

The factor scores produced by the SWAS administrations yielded Cronbach's α reliability coefficients ranging from .649 to .918. In both administrations, the *Social Persuasion for Writing* factor scores had the lowest relative reliability coefficient (α = .683 and .649). Because reliability is influenced by the number of items in the measure, it is logical that the factor with the fewest number of items would have the lowest reliability coefficient. Nevertheless, this reliability estimate is still higher than some of those found for other, similar measures. For instance, the WAMS (Troia et al., 2013) yielded reliability estimates ranging from .51 to .88 for scales measuring writing motivation and achievement goal orientations. This indicates that the SWAS is an improvement over existing tools to measure writing affect.

The reliability coefficients of the scores from all items for each administration (.944 and .945, respectively) indicate that nearly 95% of the variance in survey responses can be attributed to true human variance and not measurement error (Cumming, 2012). Although the instrument was administered to two groups of students, further administrations with different populations of students are required before I can generalize this finding. However, this preliminary result indicates that the SWAS produces reliable scores for the various factors and for students' overall self-efficacy for writing.

External Validity for the SWAS

Psychometric validity indicates how well a tool measures the intended construct, in this case self-efficacy for writing. As all self-reported data is subject to validity threats, including social desirability, it is essential to compare the results of such measures to other non-self-reported types of data (Mundai, 2011). To establish the validity of the SWAS, and answer the third research question, I examined how students' self-reported scores correlated with other, seemingly unrelated, measures of self-efficacy for writing.

Students' self-reported data was highly correlated with teacher ratings of students' feelings towards writing. While the limited number of published studies in this area prohibits direct comparison, examining existing research on other aspects of literacy motivation and affect makes these findings are particularly noteworthy. Guthrie and colleagues (2007) examined how teacher ratings correlated with students' self reported motivation for reading and found that correlations with teacher data varied widely. For instance, Guthrie and colleagues' correlations of student motivation (as demonstrated through coded interviews) with teacher ratings ranged from 0.17 to 0.77. Additionally, teachers' rating of student motivation yielded an average correlation with MRQ survey data of 0.36 (SD = 0.12). To contrast, the correlations found between teacher ratings and student self-report on the SWAS yielded much less variance and all correlations were statistically significant (ranging from 0.389 to 0.463). This shows that the scores from the SWAS can be externally validated based upon teacher observations.

Additionally, I compared students' SWAS scores to their actual writing habits. In theory, a student who reports strong, positive feelings towards writing would be less likely to avoid the task and more likely to produce written content. The correlations between students' SWAS scores and the average number of words written were modest but statistically significant across all factors.

Conclusion

Writing continues to receive varying levels of support and interest from policy-makers. For example, the new Common Core State Standards encourage teachers to promote writing in their classrooms and allow students opportunities to write texts from multiple perspectives and genres (National Governors' Association, 2010). Disciplinary and content-area literacy movements note the importance of writing and its impact on student comprehension and achievement. Moreover, American businesses spend billions of dollars on writing remediation (Cutler & Graham, 2008; National Commission on Writing, 2004). Yet, as teachers are faced with more demands on their teaching time, writing often is only a focus during grades that culminate in a writing standardized test.

Taking these factors in unison suggests that writing does not consistently receive adequate attention. The result of this unsteady approach to writing is that students' affect towards the task may suffer.

The SWAS is a tool teachers, policy-makers, and researchers can utilize to assess the students' self-efficacy for writing. This is especially important for researchers as they must measure how interventions impact student affect towards writing, ensuring that additional instruction does have a positive impact on long-term feelings about writing. Future research must investigate how classroom and intervention practices can support students' self-efficacy for writing. Additionally, scores from this instrument can be used to inform future legislation to combat the common trajectory of decreased motivation as students progress through their education. Finally, my validation of this instrument is a starting point for researchers who work in writing research. Teachers can utilize results from this instrument to measure their students' self-efficacy for writing and better understand student strengths and weaknesses in writing affect.

In Bandura's (2001) view of self-efficacy, an individual's sense of their ability to complete a task is not only good for the individual, but also for the community at large. Bandura argues that a sense of self-efficacy has shown to lead to prosocial behavior, including a cooperative, helpful nature with people invested in each other's welfare. Furthermore, this self-efficacy helps to develop the individual's sense of personal agency, which allows the individual to adapt to diverse environments, circumvent constraints, and behave in manners that help them realize desired outcomes. In essence, a strong sense of self-efficacy for writing, which can be quantified using the SWAS, may be beneficial for both individual students and the larger community of learners.

CHAPTER III

THE RUBRIC FOR SCIENTIFIC WRITING: DEVELOPMENT AND VALIDATION OF A TOOL TO SUPPORT RESEARCH, ASSESSMENT, AND INSTRUCTION

When students are given the opportunity to write in science, they take part in an authentic scientific activity, and increase their proficiency in science (Duschl, Schweingruber & Shouse, 2007). This type of instruction then naturally leads to a more inquiry-based approach to education, which supports higher order thinking (Fordham, Wellman & Sandman, 2002). Despite these benefits, writing is not receiving due attention in middle and high schools. By the time students reach high school, most of their writing instruction occurs in English class (Graham & Harris, 2012). However, in 2007, English teachers reported spending less time on writing then they had in the past, and only 23% of 12^{th} grade students' writing skills were deemed "proficient" by the National Assessment of Educational Progress (Applebee & Langer, 2009). Furthermore, a recent survey indicates that most of the writing students complete in science class requires filling in the blanks or short responses, with nearly a third of middle and high school science teachers reporting they *never* assign theory, argumentation, or research papers (Drew, Olinghouse, & Faggella-Luby, 2014).

Addressing this dearth in writing instruction and skills will require all content-area teachers to integrate writing into their curriculum. However, writing's purpose and form differs in each subject, and generic content-area literacy strategies are insufficient to promote true disciplinary literacy (Warren, 2012). Disciplinary literacy experts note that while these general skills promote writing fluency across genres, genre-specific writing conventions must be learned to communicate effectively in subject areas (Shanahan & Shanahan, 2008). Within the field of science, students must be able to create sound connections between questions, evidence, and claims (Akkus et al, 2013) – skills that are not often addressed in English/language arts classes.

While developing scientific literacy through writing is important, many science teachers may not feel equipped to assess student writing, even when that writing is specific to science knowledge (Cutler & Graham, 2008). Research shows that to be truly effective instructors of writing, teachers need to be writers themselves (Colby & Stapleton, 2006), so they can develop their own self-efficacy for the skill. Poor self-efficacy in writing has been noted as one of the key barriers teachers feel unable to overcome (Dempsey, PytlikZillig & Bruning, 2009); how can they evaluate what they feel unable to produce? One possible solution is to provide science teachers with a tool to evaluate scientific writing and identify specific aspects of scientific rhetoric where students need instruction and support.

Teachers are not the only ones struggling to assess student writing – researchers have yet to develop a standardized measure of scientific writing. In many instances, instrument validation and reliability for researcher-created tools is not reported, limiting the opportunities to make cross-study comparisons (Miller, McTigue, & Scott, 2015). Additionally, while many studies have employed various tools to evaluate students' writing in science class, these measures tend to be prompt specific. For instance, Hand, Hohenshell, and Prain (2004a) created a rubric to give feedback and score 10th grade students' essays regarding the ethics surrounding DNA research. While useful for the purpose of the study, most of the criterion (such as explaining *Gene Expression* and identifying the DNA manipulation controversy) would not be applicable to other writing samples.

There is a clear need in both teaching and research fields for improvement in instrument development and assessment for content-area writing (Miller et al., 2015). The purpose of this study was to create and validate a rubric, named the *Rubric for Scientific Writing* (RSW), which can be used to support writing instruction in science classes and evaluate scientific writing. This rubric assesses both students' general writing skills (e.g., adherence to grammar conventions) and their ability to write appropriately within the scientific genre (e.g., use of evidence to support a scientific claim). Most existing rubrics have been developed to only serve one of these purposes, not both (e.g. Gunel, Hand, & McDermott, 2009) or have been designed for a specific prompt and lack generalizability (e.g. Hand, Wallance & Yang, 2004a). To address both needs as well as provide external validity, this rubric is aligned with the Common Core State Standards Writing Anchors (National Governors' Association, 2010) and Next Generation Science Standards (NGSS Lead States, 2013) for writing.

Theoretical Framework & Models

Theoretical Framework

The term *scientific literacy* evokes two different, but equally important, definitions. The *derived* sense focuses on the knowledge and information students must learn in order to think scientifically. This *derived* sense would include concepts such as understanding how a plant undergoes photosynthesis, and is a primary goal of science educators and the focus of most standardized exams. The other definition, termed the *fundamental* sense, refers to students' ability to read and write scientifically (Norris & Phillips, 2002). For the purpose of this study, writing in science class is viewed as a strategy to support both definitions of scientific literacy – a student who has mastered the *fundamental* sense will be able to use writing as a tool to grapple with the content essential to the *derived* sense of scientific literacy.

While not a traditional piece of the science teacher's content responsibilities, writing in science class has gained more attention in recent years (See Jagger & Yore, 2012; Wright, Franks, Kuo, McTigue, & Serrano, 2015). For instance, Lesley (2014) argued that writing allows students to participate in scientific rhetoric and thus deeply engage in the content. Therefore, this author recommended that teachers use informal writing platforms, like Twitter, to allow students to communicate with professional scientists, such as NASA physicists. Lesley demonstrates that by communicating via social media, students were provided a "scaffold for developing reading comprehension and writing skills required for reading and composing various genres of scientific text" (p. 378). In essence, students were able to learn both content and writing skills from the NASA scientists and thus took part in scientific conversation and rhetoric.

Writing in science class provides students a space to build "their own knowledge through questioning, reprocessing, reflecting, analyzing… and drawing conclusions" (Alev, 2010, p. 1343). Essentially, writing can allow students to *transform knowledge* (Bereiter & Scardamlaia, 1987), which involves engaging in a self-interaction that can build new understandings. Sociocognitive theory posits that knowledge is built through social interactions with others (such as the NASA scientists; see Vygotsky, 1980), however I posit that students can engage in a similar process of self-interaction while writing. In essence, writing allows students to accomplish learning tasks by helping them organize and reformulate existing knowledge into new ideas and concepts.

Published Standards

The expectations of scientific writing are echoed in the recent Common Core Writing Anchors (National Governors Association, 2010) and Next Generation Science Standards (NGSS Lead States, 2013), which are guiding documents for literacy and science education in the United States. According to the Next Generation Science Standards, "literacy skills are critical to building knowledge in science" (NGSS Lead States, 2013, p. 1). It is for this reason that the NGSS includes direct connections to Common Core writing anchors. The requirements of scientific writing will vary depending upon an author's purpose or audience, but in general, strong scientific writing makes a scientific claim, provides valid evidence to support that claim, and provides an analysis of that evidence to build a model of reality (Cavagnetto, 2010; Shanahan & Shanahan, 2009). In addition, an author must conform to the conventions of language to ensure his or her message is understood by the reader.

Existing Models of Assessing Scientific Writing

Models for evaluating students' writing in science class do exist, and many highlight the essential characteristics of an effective scientific writing measure. However, in addition to frequently lacking validation and reliability testing, many of these are content or context specific and therefore do not allow for cross-study comparisons. In this section I review some of the previously implemented models for assessing scientific writing in order to identify ideal characteristics of a measure.

Often times, what teachers and students require in a measure conflicts with the needs of researchers. For example, in addition to quantitative outcome measures, Gunel and colleagues (2009) created a standardized worksheet to evaluate and provide students feedback on their writing. Rather than giving students numerical scores, teachers answered questions such as "which parts of the explanation were especially easy to understand or helpful?" and "after reading this paper, what is your understanding of the goal and jobs of the circulatory and respiratory systems in humans" (Gunel et al., 2009, p. 365). This type of feedback provides the consequential validity lacking in many writing assessments, however it does not provide quantitative data sought after by educational researchers.

Rubrics have also been used to evaluate the quality of scientific writing. Providing clear descriptions of expectations allows students to identify characteristics of writing in this genre. For instance, Hand and colleague's (2004a) study utilized a rubric with 10 content-area categories, including providing definitions, explaining processes, and weighing the ethics of the topic. However, like the open-ended feedback sheet, this rubric is created specifically for one target writing project, and thus it would be difficult to compare scores on different assignments (Hand et al., 2004a).

Rubrics have also been used to measure specific genres within scientific writing. Christenson, Rundgren, and Hoglund (2012) developed a model to evaluate student writing about socioscientific issues (such as global warming or nuclear power use). These authors identified six socioscientific subject areas (sociology/culture, environment, economy, science, ethics/morality, and policy) which can be expressed using personal knowledge, values, and experiences. These variables were combined to make 18 categories (e.g., Environment/Value, Policy/Knowledge, etc) to holistically evaluate students' construction of scientific argument. This is known as the SEE-SEP model, standing for the six subject areas. These authors found evidence of all 18 categories in student writing, supporting this model as a valid measure of student writing (Christenson et al., 2012). However, this measure is specific to socioscientific writing and would not be an appropriate tool for other types of scientific writing.

One drawback to creating a more generalizable rubric is that it runs the risk of being so broad as to not capture the nature of scientific writing. Rivard (2004), for example, created a simple rubric to qualitatively evaluate students' science writing. Scores ranged from 0 to 4 and two research assistants on this project were able to reach relatively high interrater reliability (.89). Responses earning low scores were described as lacking "clarity" or being "difficult to understand", whereas high-scoring responses were "elaborate, complete", and demonstrated "a well-structured conceptual understanding" (Rivard, 2004, p. 429). This rubric could likely be used with other grades and content-areas; however the criterion of "clarity" and demonstrating understanding appear to measure general writing quality more so than the specific characteristics of quality science writing.

Principles of Effective Rubrics

While certainly not an exhaustive list of methods available for evaluating scientific writing, the aforementioned models help to illuminate what is necessary of an effective writing rubric. First, it must provide useful feedback regarding the strengths and weaknesses of the writing in order for students, teachers, and researchers to know how writing needs to be improved. Secondly, it must be broad enough to be able to evaluate scientific writing composed in a variety of contexts – while assignment-specific rubrics are useful for individual classroom activities or interventions, they require significant time to create and make cross-study comparisons difficult. Finally, while being broad enough to apply to a variety of contexts, an effective rubric must also be specific enough to address the writing expectations of the scientific community. If all three of these expectations are met, the resulting rubric can be used to teach scientific writing to students (by clearly detailing the expectations), evaluate assignments and provide constructive feedback to learners, as well as score writing produced during research interventions and compare their effectiveness across studies and populations.

The purpose of the current study is to develop and validate the *Rubric for Scientific Writing* (RSW), which meets the described characteristics of an effective measure of scientific writing. Ultimately, this study seeks to answer the following questions:

1) To what extent can the aspects of scientific writing emphasized by the Common Core and Next Generation Science Standards be measured by the *Rubric for Scientific Writing*?

2) How reliable are the scores produced by the *Rubric for Scientific Writing*?

3) How valid are the scores produced by the *Rubric for Scientific Writing*?

Methods

Instrument Development

The rubric for the present study draws from the Common Core State standards (National Governors Association, 2010) and the Next Generation Science Standards (NGSS, NGSS Lead States, 2013). While the majority of the NGSS focuses on scientific skills and competencies, they do explicitly state the importance of literacy skills in building students' scientific knowledge, and therefore include direct connections to the Common Core writing anchors. These writing components are detailed in Table 3.1. In addition, Common Core Writing Anchor #4 ("Produce clear and coherent writing in which the development, organization, and style are appropriate to task, purpose, and audience") would also measure quality scientific writing.

I used these standards to create descriptors for the RSW, which evaluates six aspects of students' ability to write in the scientific genre (See Appendix A). This approach has allowed me to create a tool that is both specific to scientific writing, like Christenson and colleagues' (2012) socioscientific rubric, while being broad enough to apply to a variety of classrooms and contexts, much like Rivard's (2004) tool. This rubric contains four performance levels, each with a designated number of points: poor (0 points), developing (1 point), approaching expectations (2 points); meets or exceeds expectations (3 points). Descriptions of the dimensions are detailed in Table 3.2.

Table 3.1

Essential Writing Anchors as identified by the Next Generation Science Standards

Common Core Writing Anchor	**Connection to Next Generation Science Standards**
CCR Writing Anchor #1: Write arguments to support claims in an analysis of substantive topics or texts using valid reasoning and relevant and sufficient evidence.	Central to the process of engaging in scientific thought or engineering practices is the notion that what will emerge is backed up by rigorous argument. Writing Standard 1 places argumentation at the heart of the CCSS for science and technology subjects, stressing the importance of logical reasoning, relevant evidence, and credible sources.
CCR Writing Anchor #2: Write informative/explanatory texts to examine and convey complex ideas and information clearly and accurately through the effective selection, organization, and analysis of content.	Building a theory or a model that explains the natural world requires close attention to how to weave together evidence from multiple sources. With a focus on clearly communicating complex ideas and information by critically choosing, arranging, and analyzing information, Writing Standard 2 requires students to develop theories with the end goal of explanation in mind. The demand for precision in expression is an essential requirement of scientists and engineers, and using the multiple means available to them is a crucial part of that expectation. With a focus on clearly communicating complex ideas and information by critically choosing, arranging, and analyzing information—particularly through the use of visual means—Writing Standard 2 requires students to develop their claims with the end goal of explanation in mind.
CCR Writing Anchor #7: Conduct short as well as more sustained research projects based on focused questions, demonstrating understanding of the subject under investigation.	Generating focused questions and well-honed scientific inquiries are key to conducting investigations and defining problems. The research practices reflected in Writing Standard 7 reflect the skills needed for successful completion of such research-based inquiries.
CCR Writing Anchor #9: Draw evidence from literary or informational texts to support analysis, reflection, and research.	The route towards constructing a rigorous explanatory account centers on garnering the necessary empirical evidence to support a theory or design. That same focus on generating evidence that can be analyzed is at the heart of Writing Standard 9.

Note. Common Core Writing Anchors from National Governor's Association (2010). Connection to Next Generation Science Standards from NGSS Lead States (2013).

Table 3.2

Dimensions of Writing measured by Rubric for Scientific Writing

Dimension	Common Core Writing Anchor(s)	*Meets or Exceeds Expectations* Description
Claim	1	The student addresses one claim in the paper. Claims might be explanatory, persuasive, or argumentative in nature but show concentrated focus of the topic of the paper.
Evidence/ Support	7 & 9	The student provides factual evidence and support from authentic sources (such as class experiments, course material, or professional, scholarly works). The evidence supports the claim made in the paper.
Analysis of Content	2	The student develops the claim further by synthesizing the evidence and formulating a clear explanation ***AND*** The information has been analyzed for the reader to provide precision to the model or theory the student is constructing.
Organization	4	Student supports the claim in an organized manner. Paragraph order is logical and each paragraph focuses on one main idea. The student shows evidence of choosing a text structure (such as problem-solution, compare-contrast, or chronological) that is appropriate to the audience and purpose.
Audience	4	Style and use of language is appropriate for the audience and purpose. Student provides enough background information for the audience to understand while employing appropriate scientific terms and language.
Presentation of Writing	4	The student's writing shows a mastery of writing conventions (such as grammar, semantics, syntax, and punctuation). Sentences flow appropriately and show variation in structure. The formatting is appropriate for scholarly scientific writing.

Writing Sample

The sample for this study consisted of 75 spontaneously written samples produced by students in grades six through 12. These samples were collected as part of a larger writing-to-learn intervention in science classes, and represented writings from the beginning and end of the intervention. Students were given approximately 20 minutes to complete each writing task, and the tasks had to be completed in class. While this approach, in one respect, may be seen as less-authentic as students did not have time to revise and edit, it did allow me to ensure that students produced the writing independently (i.e., without the assistance of parents). Additionally, this method mirrors the context of high-stakes writing exams, where students must complete a writing task in one sitting with limited resources (Behizadeh, 2014).

In order to properly validate the RSW, the writing samples needed to both represent authentic pieces of student work as well as demonstrate, to varying degrees of proficiency, scientific writing. Strong scientific writing requires that students create sound connections between questions, evidence, and claims (Akkus et al., 2013), and so I developed writing prompts that required the students to use their content knowledge as evidence to formulate an argument. These procedures also encouraged students to move beyond *knowledge telling* writing, where they simply repeat information from class, to more sophisticated *knowledge transforming* writing (Bereiter & Scardamalia, 1987).

Nearly all writing prompts followed a similar format – students were provided with a brief passage supposedly published in “The Journal of Ill-Informed Science”. Each passage related to a topic recently studied and included incorrect information. I worked with the classroom teachers to ensure that the inaccurate statement related to a “big idea” from that unit of study which teachers felt was essential to students’ overall understanding of the content. The students were instructed to write a letter-to-the-editor that accomplished three tasks that would help them construct a written argument: (1) identified what was incorrect (2) provided the correct information and (3) explained why their information was correct. Furthermore, students were reminded that “Good science writing has one claim that is supported with facts" and that "the writer also explains and connects all the information". Please see Figure 3.1 for a sample writing prompt. These writing samples were drawn from those produced by students in grades six through 11 during the intervention described in Chapter IV.

Name: ______________________________ Date: ______________

The following line appeared in the *Journal of Ill-informed Science*

Because faults exist below the lithosphere, landforms have nothing to do with plate tectonics

We know this isn't true! Write a letter to the editor that:

1) *Identifies* what is incorrect about this statement (your claim)
2) *Provides* the correct information (your facts)
3) *Explains* why your information is correct (your explanation and connection)

Remember good science writing has ***one claim*** *that is* ***supported with facts****. The writer also* ***explains and connects all the information****.*

To the Editor of the *Journal of Ill-informed Science*:

____

____

____

Figure 3.1. Sample writing prompt

Raters

Three certified English teachers and three certified science teachers (five Ph.D. students and one Ph.D. graduate) evaluated scientific writing samples composed by students in grades six through 11. I chose to have a group of raters with mixed content-area backgrounds to ensure that all aspects of writing (i.e., science content and English conventions) were addressed in the study. All raters also had classroom teaching experience, either at the K-12 or adult level, to ensure they had the real-world understanding of student work in addition to a research mindset (i.e., understand the importance of scoring fidelity). The raters are described further in Table 3.3.

Raters were told that they would be scoring all 75 writing samples twice, once without and once with the rubric, although they did not preview the rubric before the first round of scoring. The raters first read samples and provided a holistic score (0 to 3) and a brief explanation for their rating. A score of 0 was identified as "poor", 1 as "developing", 2 as "approaching standards", and 3 as "meets or exceeds standards". Raters were simply told to rate students based upon their "expectations for college readiness in your discipline (Science or English)". These categories were purposefully left vague to allow raters to use their professional judgment. Then, raters were asked to evaluate the writing a second time using the rubric. While they received no formal training before rubric scoring, raters were allowed to ask questions discuss the rubric with me at any point. During both rounds, raters were encouraged, though not required, to provide comments or feedback to explain their scoring. Raters were able to complete this process in their own time over the course of one month

Table 3.3

Descriptions of Raters

ID	Areas of Study: Bachelors	Masters	Ph.D	Certifications/ Licensures	K-12 Teaching Experience	Other Experience
Science Teacher J	Math & Science Education	Urban Education	Urban & Science Education	Math & science (grades 4-8); Supplementary English as a Second Language (ESL); Virtual Instructor Certificate	5 years science (grades 6-8)	NSTA New Teacher Fellow; Regional Science Teacher Mentor; University instructor for pre-service teachers (classroom management & technology)
Science Teacher K	History	Science Education	Science Education	Social Studies composite (grades 8-12); Generalist (grades 4-8)	4 years math and science (grades 4-5); 4 years social studies (grades 7 and high school); Home school community elementary science	University instructor for pre-service teachers (math and science methods)
Science Teacher L	Biology	Curriculum & Instruction	Science Education	Science (grades 8-12); Principal certification	4 years science (grades 9, 11, and 12)	University instructor for pre-service teachers (elementary science methods)
English Teacher S	Special Education (grades K-9)	Reading & Literacy; Reading specialist	Urban Education	Special Education (grades K-9); Reading Specialist (grades PreK-12)	7 years Special education English/Language Arts co-teacher (grades 7-8)	University instructor for pre-service teachers (Language arts methods); District writing scorer
English Teacher L	English; Spanish, Religion (minors)	Cross-cultural studies	English as a Second Language	Secondary English; Trainer for Teaching English as a Foreign Language teachers	English for elementary and high school students abroad (Latvia & Italy)	University instructor for pre-service teachers (Reading and ESL methods); English teacher for adults in China
English Teacher D	English & Government	English	Reading/ Language Arts	Grades 6-12 English Language Arts, Reading, & Government	8 years English (grades 9, 10, & 11); 2 years reading intervention (grades 7 & 8)	Regional English Language arts and Reading curriculum specialists & reading coach

. After completing all scorings, five of the six raters (three science and two English raters) participated in a focus group to discuss how they interpreted the rubric. Focus group interviews allow researchers to understand the relationship between a stimulus (the rubric) and an effect (the scores) (Merton & Kendall, 1946). I provided raters with their original scoring sheets and notes to help them recall and discuss their thoughts about scoring the writing samples. I selected writing samples the raters agreed upon and a few they disagreed upon to be the focus of the discussion. To begin the conversation, I gave each participant a copy of the rubric and a writing sample they generally agreed upon and asked them to describe why they made their decisions. I then introduced samples where the English and science raters agreed, followed by a few outliers where many raters disagreed. I continually prompted the participants to refer back to the rubric to identify necessary changes.

Analyses

Reliability. Psychometric reliability describes the consistency or correlation of a set of scores (Grissom & Kim, 2012). In order for this writing rubric to be effective in research we must establish that the scores it produces are reliable, meaning that raters can consistently assign scores both individually (i.e., each rater applies the rubric criterion consistently across different writing samples) and as a group (i.e., different rater's scoring of a single writing sample are consistent). These types of reliability are known as intra-rater and inter-rater estimates, respectively.

I calculated inter-rater reliability based upon the evaluators' scores of individual writing samples. This is considered a stability coefficient as it requires multiple assessments from different individuals. Cronbach's α, by contrast, is an internal consistency coefficient as it measures the consistency of scores from one administration (Reinhardt, 1996) and served as my intra-rater reliability estimate. This reliability estimate was calculated for each rubric dimension using the scores from all raters. Using both types of coefficients allowed me to examine the stability of this rubric across individuals (i.e., does each rater interpret the rubric in a similar manner?) and across writing samples (i.e., is the rubric specific enough for each rater to maintain consistency while scoring?).

Validity. I conceptualized validity through construct (using a factor analysis), criterion, and face validity for the rubric. First, I conducted a principal component factor analysis to examine the underlying constructs measured by the rubric (Thompson, 2004). In regards to criterion-related validity, due to limitations of published instruments, I was unable to make a direct correlation between my measures and a discipline specific standardized measure. However, I administered the more global *Test of Written Language* and compared students' scores on this measure with their writing sample scores. This test presents students with a picture and gives them twenty minutes to write a story based upon the visual. Raters then score the work for both contextual conventions and story composition, and these make up a composite writing score. This measure has been normed for children ages nine years to 17 years 11 months, and was thus an appropriate measure for this population (McCrimmon & Climie, 2011)

Face validity is achieved when an expert examiner comes to the conclusion that the items are measuring what they report to measure. Face validity is an important feature of an educational assessment as it describes the practical use of the measure (Nevo, 1985). To establish face validity, I compared the raters' rubric and holistic scores. While I did not expect there to be a perfect correlation, I hypothesized that the English raters' holistic scores would correlate more strongly with the writing conventions aspects of the rubric and the science raters' would correlate with the argument development aspects. Aspects where correlations deviated from these hypotheses were examined further during the focus group interview.

Focus Group. The focus group was video recorded and the conversation transcribed for analysis. To analyze the data, I first examined the quantitative data to identify findings which could be supported by the qualitative focus group. These themes became the first set of codes I used to analyze the transcription data. I then analyzed the transcription using constant comparative analysis (Glaser & Strauss, 1965), which allowed additional themes to emerge through the voices of the participants. These procedures allowed me to identify findings from the qualitative data that both converged and diverged from the quantitative findings.

Results

First I will present the quantitative results, followed by the analysis of the qualitative focus group data to provide converging evidence for the findings.

All six raters completed both the holistic scoring and rubric scoring of 75 writing samples. Their average scores and standard deviations are detailed in Table 3.4.

Table 3.4

Mean scores assigned by participants

ID	Holistic Score	Claim Score	Evidence Score	Analysis Score	Organization Score	Audience Score	Presentation Score
Science J	1.30 (0.92)	2.34 (0.84)	1.70 (0.90)	1.51 (0.91)	1.69 (0.76)	1.99 (0.72)	1.62 (0.62)
Science K	1.70 (0.95)	2.32 (0.66)	1.57 (0.93)	1.40 (0.97)	1.55 (0.59)	1.28 (0.80)	1.56 (0.67)
Science L	1.51 (0.97)	1.69 (1.01)	1.49 (0.98)	1.58 (0.98)	1.76 (0.97)	1.81 (0.89)	1.84 (0.88)
Science Rater Average	**1.46 (0.87)**	**2.21 (0.75)**	**1.53 (0.83)**	**1.44 (0.85)**	**1.62 (0.65)**	**1.64 (0.73)**	**1.65 (0.62)**
English S	1.62 (0.63)	1.76 (0.60)	1.54 (0.59)	1.25 (0.63)	1.68 (0.63)	1.76 (0.56)	1.94 (0.49)
English L	2.12 (0.61)	2.42 (0.73)	1.94 (0.73)	1.91 (0.89)	2.11 (0.78)	2.06 (0.78)	2.12 (0.67)
English D	2.08 (0.87)	2.09 (0.80)	1.99 (0.82)	1.73 (0.97)	1.41 (0.74)	1.44 (1.05)	2.19 (0.72)
English Rater Average	**1.91 (0.60)**	**2.11 (0.58)**	**1.81 (0.63)**	**1.64 (0.75)**	**1.71 (0.66)**	**1.74 (0.73)**	**2.09 (0.49)**

Note. Standard deviations are shown in parenthesis next to means

Reliability Analysis

I calculated the Cronbach's α for each teacher's scores to establish intra-rater reliability (See Table 3.5). The rater's reliability estimates ranged from .705 to .898, indicating that they were generally able to apply the rubric criteria consistently across writing samples.

Table 3.5

Intra-rater reliability estimates

	Science J	Science K	Science L	English S	English L	English D	Average
Cronbach's α	.898	.705	.894	.865	.810	.826	.833

I calculated inter-rater reliability based upon near-matched scores. As scoring writing is a rather subjective task, it is common practice to consider close agreement to be a match for the sake of research and scoring (Johnson, Penny, & Gordon, 2000). For the purpose of this study, two scores were considered a near-match if at least two of the three science and two of the three English raters gave the same score. The percent of samples which fit this criterion is detailed in Table 3.6. Agreement within content-areas ranged from 91% (English raters' ratings of evidence and organization) to 76% (Science raters' ratings of organization). In overall agreement (e.g., at least four raters, two English raters and two science raters agreed), the percentages ranged from 76% in evidence to 64% in audience.

Validity Analysis

Construct Validity -- Factor Analysis. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) for the data set was .865, and the Bartlett sphericity test was less than .000, indicating the data was sufficient to proceed with a factor analysis (Thompson, 2004). I first conducted an Exploratory Factor Analysis, allowing all the variables to correlate freely. All variables correlated onto one factor explaining 76.9% of the total variance. However, because I originally hypothesized that this rubric would measure two aspects of writing (general composition and scientific writing), I then repeated the analysis and forced the items into two factors. Using varimax rotation, the resulting model explained 86.9% of the variance, and the rubric items in each category clearly represented two theoretical factors: Scientific rhetoric and English composition (See Table 3.7).

Table 3.6

Percent of samples with near-matched scores

	Holistic	**Claim**	**Evidence**	**Analysis**	**Organization**	**Audience**	**Presentation**
English Rater Agreement	80%	81%	91%	88%	91%	83%	87%
Science Rater Agreement	91%	85%	85%	85%	76%	79%	85%
Overall Agreement	70%	67%	76%	73%	67%	64%	72%

Table 3.7

Factor Loadings

Rubric Category	Scientific Rhetoric	English Composition
Claim	.706	
Evidence	.909	
Analysis	.889	
Organization		.798
Audience		.750
Presentation		.903
Cronbach’s α	.925	.910
Total Variance Explained (%)	44.69	42.21

Criterion-Related Validity -- Correlations with TOWL-4 Scores. I administered the spontaneous writing subtest of the *Test of Written Language* (TOWL-4), form A, to all student participants. This test measures two aspects of student writing – their use of English conventions and ability to construct a story. I correlated these raw scores with each student's average scientific argumentation and English rhetoric score as derived from the rubric. These moderate correlations (ranging from .417 to .494), are all significant at the .01 level.

Face Validity -- Correlations with Holistic Scores. I calculated the correlations between each rater's holistic scoring and their scoring on the rubric. I then calculated the average correlation on each component of the rubric for the three English raters and three science raters. As the purpose of this analysis was to examine the similarities between individual rater's holistic and rubric scoring, inter-rater agreement was not considered as part of the procedure. All correlations were significant at the .05 level, and these results are detailed in Table 3.8.

Table 3.8

Correlations between raters' holistic and rubric scores

	Claim	**Evidence**	**Analysis**	**Organization**	**Audience**	**Presentation**
Eng. Holistic	.28	.36	.69	.57	.49	.31
Sci. Holistic	.48	.65	.74	.50	.42	.31

Focus Group

Summarizing the quantitative data presented above reveals three themes: (1) English raters' holistic scores were more highly correlated with analysis, organization, and audience scores; (2) Science raters' holistic scores tended to correlate more strongly with claim, evidence, analysis, and organization scores, and; (3) the factor analysis revealed two clear factors. These three themes were also expressed during the focus group. In the following sections, I will first provide evidence from the focus group that converges with these quantitative findings. Next, I will describe a unique theme that emerged from the qualitative data – the need for a rubric for scientific writing in both research and teaching.

Rubric Correlations with Holistic Scores. Science raters' and English raters' holistic scores correlated with different dimensions of the rubric, and it is clear that the raters had different expectations for the writing depending upon their academic discipline. In this section, I discuss each section of the rubric individually and present qualitative explanations for why these correlations exist and how raters interpreted each rubric dimension. For simplicity, I will simply refer to raters' *holistic correlations* when discussing how their holistic scores correlated with a specific dimension of the rubric.

Claim. Science raters' tended to have a higher holistic correlation with their claim scores than the English raters. One explanation may be that the Science raters tended to decide, very quickly, whether or not the information presented in the writing was accurate based upon the students' initial statements, and this appeared to greatly influence their holistic scoring. For instance, in a chemistry class that had discussed how ionized liquids (such as Gatorade) could help athletes recover more quickly from a workout, one student had misunderstood and wrote that people should drink salt water after exercising. While his writing had some merits, such as explaining how salt water contained more ions and would thus aid in muscle movement, the Science raters read the incorrect claim and gave a low score. Science Teacher J simply explained that she "scored low because their claim was 'drink salt water'", and this low score persisted when scoring the claim with the rubric. In other words, the science raters felt that accuracy was a necessary precondition to quality writing.

Evidence. The Science raters also had higher holistic correlations with their evidence scores. This may be partially due to the fact that Science raters considered visuals, graphics, and formulas as part of the students' use of evidence, whereas the English raters did not. For example, whereas Science Teacher J said that "some of the students drew the graph as their evidence and I was like 'oh, good, fair enough'", English Teacher L explained that she "just skipped that, because it wasn't words." In essence, the English raters tended to not view anything other than strict text as part of the writing and therefore did not consider visuals as contributing to the overall evidence and argument, whereas Science raters had broader definitions of the acceptable forms of writing

Additionally, the English raters viewed much evidence, if presented without context or analysis as distracting, and described it as reading more like a list of information than formal prose. However, the Science raters read long-lists of evidence as encouraging, believing this demonstrated that the students had a basic understanding of the concept. For instance, consider the following conversation about one student sample between Science Teacher J and English Teacher S:

> Science Teacher J: You could tell he was confident in his statements. Like, it was very clear he knew what he was talking about.
> English Teacher S: Yeah, but it was awfully “list-y”.

These two differing interpretations lead the English raters to provide low holistic scores to long lists of evidence, while Science raters provided much higher scores. Again, this related to the definitions of the appropriate forms of academic writing.

Analysis. All raters, regardless of their content-focus, had a strong holistic correlation with the rubric analysis scores. This indicates that analysis is viewed as an important part of writing in any subject area, and including this element in the rubric is essential for the measure to be valid. When discussing the importance of analysis in writing, the Science raters emphasized the fact that students who provided “justification” (Science Teacher L) for their statements would likely get a high holistic score. The English raters agreed, one (English Teacher L) even used a student example to describe a poor analysis: “So the student wrote ‘The temperature outside would have to be colder than the temperature of the scientist’s refrigerator’... okay… thank you? That adds nothing to your argument.”After reading this statement aloud, Science Teacher J quickly agreed, saying “no clear analysis there”. Thus, by this student simply providing a fact without justification or analysis, all the raters felt the entire piece was poorly written.

Organization. In the case of organization, both groups of raters had a relatively strong holistic correlation; however English raters’ tended to be slightly higher, indicating that this element of writing was more important in their professional views.

English Teacher L explained that organization was one of the main things she was looking for when conducting the holistic scoring. Science raters, by contrast, could appreciate strong organization but this would not make up for poor use of scientific information. In the case of one student describing the interaction between various body systems, the Science raters discussed how their low scores were likely a result of poor teaching of the material:

> Science Teacher J: I really feel like it was a matter of teaching, because I felt like his organization was good, and he tried to follow through on his evidence, but the textbook didn't say any of this information.
> Science Teacher L: That's why I said "science is wrong, attempts to justify"

Thus, although Science raters could identify and appreciate strong organization, in their holistic scoring this would not make up for poor science.

While strong organization could not make up for poor science, Science raters discussed how improving organization may lead to stronger scientific writing. One writing piece discussed during the focus group required the student to describe how different body systems interact. The student had provided a lot of information and had few grammar errors, but the overall organization was poor and there were few connections between facts. Science Teacher K explained that had the student perhaps focused on just one body system and "described how it supported the other systems" the student's organization would have been stronger and he would have made a better scientific argument. Therefore, while organization was not the main focus of the Science raters, organization was appreciated as an important part of scientific writing.

Audience. The Science raters tended to have very low holistic correlations with their audience scores, but this may be a result of not having a clear definition of audience as it pertains to writing. Because most of these prompts were designed to be written as letters, the Science raters considered using a formal opening, such as 'Dear Editor' as evidence of the student acknowledging the audience; however, the prompt had a similar greeting printed at the top (see Figure 2.1), so students would not necessarily re-write this introduction. Science Teacher K and Science Teacher L made this apparent in their conversation toward the end of the focus group.

> Science Teacher K: I thought it was hard to differentiate in audience
> Science Teacher L: Yeah
> Science Teacher K: Because they either acknowledged, or they didn't say anything at all.
> Science Teacher L: I started doing that as well, like saying "they didn't acknowledge it", so I started marking a bunch off. But then if you notice it's [the greeting] actually printed right here in the paragraph, so some were like doing it and being redundant"

In contrast, the English raters viewed the dimension of audience representing a composite of the tone and amount of information provided in the writing, which more closely aligned with the rubric description. For example, much of the students' writing reflected a tongue-and-check rudeness, often insulting the editor for printing incorrect information. English Teacher L expressed that this approach lowered the scores she provided for audience, arguing that "if you were writing to the editor of any journal, you would not be that rude… so I knocked them down on that for audience", whereas Science Teacher J stated that she "didn't even look at that".

Presentation. None of the raters had strong holistic correlations with presentation scores, most arguing that they were more concerned with the content of the student writing than whether or not the student made errors. English Teacher S made this clear in discussing how her background impacted how she scored the writing samples:

> English Teacher S: A general comment for probably all of mine, coming from a middle school special education background … if it is on the strong side, you are getting the benefit of my doubt. Because … I don't get hung up on tiny grammatical things and things like capitalization, because if I had in my past life, I would have gone insane. So … I didn't even consider things like capitalization, unless it was really obnoxious and distracting from the writing.

English Teacher L expressed similar sentiments, often describing errors as "just some minor *minor* spelling and grammatical things" rather than major issues she felt detracted from the overall writing.

While the Science raters made few comments about grammar or spelling mistakes, vocabulary usage occasionally impacted their presentation scores. The Science raters described instances where poor use of vocabulary and scientific language distracted them from the rest of the writing. For instance, in one prompt students were asked to explain Lewis Dot Structures, and frequently identified the marks in these figures as "dots", rather than electrons. Participant Science Teacher L was very frustrated, expressing that he "didn't like when they just kept saying… and the dots do this" as the "dots" did not actually do anything, they were simply representing an atomic relationship on paper. It appears that the Science raters included the appropriate use discipline-specific language as part of the overall presentation score.

Rubric Factor Analysis. The quantitative analysis revealed that the Rubric for Scientific Writing measured two distinct constructs. The presence of these two factors was instinctively noted by the raters, without any information about the development of the instrument. Science Teacher K tried to explain this during the focus group:

> Science Teacher K: I would say from a researcher's perspective, I'd also scale them separately, like you do the first three rubric items like "science proficiency", and second three is like … writing proficiency so that way

> you can get distinctive scores. So if I want to give somebody feedback I can say how comprehensive your science is, but your writing is weak.

The other raters agreed with this assessment, most saying that they felt stronger on one half of the rubric than the other. English Teacher L even stated that after reading the rubric a few times, she barely even looked at the half that focused on English rhetoric because she felt the expectations were so clearly aligned with her professional judgment as an English teacher.

Examining the raters' notes made during the holistic scoring provides further validity for this factor structure. The Science raters tended to make written comments such as "very detailed examples" or "logical and connected information" for writing with high holistic scores. By contrast, samples with low holistic scores tended to be paired with comments such as "conceptually confused" or "just listing information, no attempt to connect", indicating that Science raters were most concerned with the presentation of correct evidence and how well the student analyzed that information. The English raters made more written comments about structure and overall argument, describing strong samples as having "good structure" and weak ones as "lacking transitions" or "response is superficially developed".

Science Teachers' Need for RSW. One additional theme that emerged from the focus group was the need for a tool, such as the *Rubric for Scientific Writing*, to encourage writing in science classes. Consistent with Shanahan and Shanahan's work (2008), scientific writing was considered a distinct genre from that emphasized in most English/Language Arts classes. Early on in the focus group, Science Teacher L discussed the fact that, in his classroom experience, there were many students who would earn poor grades for their writing in English class, but would do quite well when asked to write for his science class. He attributed this to the fact that English teachers were more likely to emphasize and value creative prose, whereas he appreciated writing that was more “bare bones, facts” without “fluff”.

Additionally, the Science raters saw that having students write about their learning in science class illuminated areas of understanding. Science Teacher L commented that by reading the student writing, it was clear who did and who did not understand the scientific concepts. For example, on one writing sample where a student scored very high, Science Teacher K explained that the student used “language that’s not common in spoken language”, quoting a passage where the student explains how Lewis structures represent the relationship between electrons in compounds. Science Teacher L added that this made it clear that he “had listened in class” and that this student’s writing demonstrated he not only understood what Lewis structures were, but how they represented models of reality.

The Science raters also noted when students’ writing revealed gaps in understanding. One of the prompts, composed in a physics class, required students to explain how the temperature of a liquid would change as it went from a liquid to a solid. While many students were able to provide mathematical formulas to calculate the temperature of the liquid, they were largely unable to describe what was happening during this phase change. The Science raters noted with concern that the basics of this concept would have been taught at a much earlier age, and quickly engaged in a conversation about how they, as the classroom science teacher, would re-teach the information based upon the students' written responses.

Using the rubric also allowed all raters to better identify student strengths and weaknesses for the purpose of improving instruction. Science Teacher J discussed the fact that she preferred scoring the writing samples using the rubric because when doing the holistic scoring she felt tempted to give students scores such as "one plus" or "two minus". However, when using the rubric she was able to differentiate "the different components and give them the true three for their claim and the two for their evidence" which, from a teacher's perspective, she felt was a much more authentic evaluation of the students' writing.

Finally, the Science raters also spontaneously discussed how important writing is in science. When discussing one student's possible misconceptions, Science Teacher K noted that even if the student did know the information, "she can't write to communicate what's happening, so we've got to deal with that". The Science raters did not feel that writing was simply an extra thing to add to a science curriculum – they felt that it was an essential skill that had to be honed and practiced. Science Teacher K clearly detailed the relationship between writing and science in the following quote:

> …At the end of the day science is more than just investigating and knowing facts, it's about being able to communicate what you learned to a variety of different audiences. So in that regard it's clear that they learned the information but … they didn't finish the cycle. Like part of science is to inform the next generation of people who are going to investigate and kind of keep that cycle going.

At the end of the focus group, all three Science raters agreed that they saw a benefit to having a tool like the *Rubric for Scientific Writing* to support both the teaching of scientific writing to students and evaluating student writing in research settings.

Discussion

Multiple Common Core Writing Anchors (see numbers 1, 2, 7, and 9 in Table 3.2, specifically) have been identified by the Next Generation Science Standards (NGSS) as important for scientific communication (See Table 3.1). In addition, Common Core Writing Anchor number 4, which focuses on developing clear and coherent writing, is also an important aspect of quality writing. Therefore, to answer my first research question (*Can the aspects of scientific writing emphasized by the Common Core and Next Generation Science Standards be measured by the Rubric for Scientific Writing?*) I will examine each of these anchors individually.

Writing Anchor 1 focuses on the development of argumentation, which the NGSS identify as essential for scientific thought. Specifically, scientific argumentation needs to support claims made using relevant evidence and logical analysis (NGSS, 2013). Similarly, Writing Anchor 9 emphasizes the importance of providing evidence to support analysis in research. At the surface level, as written in the descriptors, the *Rubric for Scientific Writing* does measure students' claim, use of evidence, and analysis. At a deeper level, the fact that Science raters' holistic scores were highly correlated with these particular rubric dimension scores indicates that these dimensions are in fact measuring this essential aspect of scientific writing. In other words, the Science raters gave much weight to these aspects of writing.

Writing Anchor 2 also discusses the importance of effective analysis of content, and the NGSS argue that this anchor requires students to communicate "complex ideas and information by critically choosing, arranging and analyzing information… to develop their claims with the end goal of explanation in mind" (NGSS Lead States, 2013, p. 9), indicating that organization is an important aspect of scientific writing. As designed, the *Rubric for Scientific Writing* has a dimension dedicated to organization. Furthermore, in the focus group the raters noted that poor organization detracted from overall scientific writing quality.

While not mentioned by the NGSS as an element of scientific writing, Writing Anchor 4 examines students' ability to "produce clear and coherent writing in which the development, organization, and style are appropriate to task, purpose, and audience" (National Governors' Association, 2010, p. 18). This anchor can be viewed as a precursor to scientific writing, because if the writing is not clear it can be difficult, if not impossible, to evaluate the other aspects of scientific writing. The *Rubric for Scientific Writing* contains two dimensions, audience and presentation, which measure the elements described in this writing anchor. Audience strongly correlated with English raters' holistic scores, indicating that this aspect of writing can be measured by the tool, although further explanation may be necessary for audience to be interpreted similarly by Science raters. Evidence from the focus group did not indicate that Science raters did not value audience, only that they seemed unfamiliar with this element of writing. Future researchers should provide examples and training materials to help Science raters be more aware of audience and how it impacts the quality of scientific writing.

Although none of the raters' holistic scores correlated strongly with presentation, this did not mean they did not consider this element. Rather, as English Teacher S explained, as long as presentation did not interfere with overall understanding it tended to not impact holistic scoring. However, when using the rubric raters felt comfortable providing accurate scores for English conventions on the presentation dimension as they had the opportunity to credit students for their strengths in other aspects of writing.

The final Writing Anchor (number 7), focuses on research skills that would demonstrate students' understanding of subject matter. According to the NGSS, these skills are "key to conducting investigations and defining problems" (NGSS Lead States, 2013, p. 3), and are an essential piece of the scientific process. This aspect is not directly measured by the *Rubric for Scientific Writing*, as this anchor describes a long-term practice that may not always be evident in one piece of writing. However, the other elements of writing measured by this tool would help to identify quality scientific writing which could be the end goal of such scientific investigations.

Rubric Reliability

To answer my second research question (*How reliable are the scores produced by the Rubric for Scientific Writing?*) I examined the inter- and intra- rater reliability estimates and compared these to other similar tools. The intra-rater reliability scores indicated how well each rater was able to apply the rubric in a consistent manner across writing samples. These scores ranged from .705 to .898, with an average score of .833. This score is slightly lower than other measures, and the reasons why will be discussed further. This finding indicates that most of the variance (approximately 83%) within each rater's scoring is a result of true score variance rather than measurement error. Science Teacher K's intra-rater score of .705 was quite different from the rest of the group (1.77 standard deviations lower than the mean score), indicating that he may be an outlier in this area. However, Science Teacher K was the participant with the most content-diverse background (e.g., undergraduate degree in history, certified social studies teacher, and professional experience as a science teacher) and his comments during the focus group clearly indicate he was concerned with the quality of student writing as an essential piece of scientific communication. Therefore, unlike the other raters who focused on their specific discipline of training, he may have been considering the expectations of multiple disciplines in concert, which resulted in greater variation.

The inter-rater reliability estimates can be described by examining the near-matched scores on the various dimensions of the rubric. While the scores varied from 67% agreement to 91% agreement, the Science raters had a higher level of agreement on the dimensions of the rubric which compose the scientific rhetoric factor of the rubric (minimum 85% agreement). Likewise, the English raters' agreement was higher on the dimensions which compose the English rhetoric factor (minimum 83% agreement). This finding indicates that the raters were more comfortable in scoring the rubric dimensions that directly related to their discipline, and thus had less variation in their scoring.

To compare, Rivard (2004) used a five point rubric to score students' written responses in science class. Two graduate assistants were trained before scoring five different writing samples, and their exact agreement ranged from .85 to .95 with an average of .89 for the entire study. While the inter-rater reliability estimates for this study are indeed slightly lower than Rivard's sample, these scores were obtained without any training or calibration on the part of the raters. While resulting in lower reliability, this practice does mirror typical K-12 classroom settings where teachers receive minimal training on assessment materials such as rubrics. It can be presumed that future researchers, who can provide such introduction before the use of the rubric, will obtain scores that yield higher inter-rater estimates.

Rubric Validity

Answering my final research question (*How valid are the scores produced by the Rubric for Scientific Writing?*) required examining the converging findings of the qualitative and quantitative data. To establish criterion-related validity, I compared the scores on the RSW with student scores on a published, standardized writing exam, the TOWL-4. The students' factor scores produced moderate, but statistically significant, correlations with their scores on the TOWL-4. As these two tools are designed to measure very different aspects of writing, the moderate correlations indicate that there is some relation and, thus, scores from the RSW provide a valid measure of students' writing abilities.

Construct validity is typically established by an exploratory factor analysis procedure, which created a two-factor model explaining 86.9% of the overall score variance. This model measures two aspects of scientific writing – Scientific rhetoric and English Composition – and the rubric elements in each factor align with the expectations of the specific disciplines. This model also was supported by the comments made during the focus group. The participants inherently viewed the two factors in the rubric, and even suggested that the rubric could be broken into two composite scores aligning with the findings of the factor analysis. As this focus group was held before the factor analysis, I can also rule out any researcher-bias that may have led the raters to make these comments.

Face validity was also found in examining the correlations between raters' holistic and rubric scores. The Science raters had a high level of correlation on dimensions of the rubric most emphasized by the Next Generation Science Standards – that is, claim, evidence, and analysis. English raters' scores also correlated with the analysis dimension, as well as the organization and audience dimensions. When I examined the comments during the focus group, it was clear that Science raters were largely concerned with the presentation of scientific information and logical analysis. The English raters, by contrast, were less impacted by the quality of scientific information and thus scored students based upon their presentation of a logical, organized argument. These findings indicate that the *Rubric for Scientific Writing* produces scores with both construct and face validity.

Conclusion

Earlier in this manuscript, I argued that an effective tool for evaluating scientific writing must have three qualities: (1) it must be able to provide feedback to improve student writing; (2) it must be broad enough to be applied to multiple topics and contents within the field of science, and; (3) it must be specific to the genre of scientific writing. The *Rubric for Scientific Writing* accomplishes all three tasks. First of all, as noted by Science Teacher J in the focus group, the rubric allowed the raters to not just give a score, but note where students' strengths and weaknesses lay. This would allow teachers to offer direct instruction in areas of weakness, and provide researchers clear indications as to how interventions impact student achievement. Secondly, this rubric was consistently applied to score writing from multiple grade levels focused on an assortment of scientific topics, indicating that it is broad enough to be used in a variety of classroom and research settings. Finally, the alignment with Common Core Writing Anchors and Next Generation Science Standards, along with the construct and face validity established by this study, indicate that this tool is specific to the genre of scientific writing.

Overall, my findings demonstrate that the *Rubric for Scientific Writing* produces valid and reliable scores of students' scientific writing. The Science raters' holistic scores had moderate to strong correlations with their rubric scoring of students' ability to produce a scientific claim, support that claim with evidence, and provide scientific analysis. The factor analysis demonstrates that this rubric clearly evaluates the elements required of proper scientific argumentation and English rhetoric, which would allow future researchers and teachers to examine how instruction and interventions impact these two aspects of scientific writing.

Further research is still needed to truly describe the versatility of this tool. While the writing samples used in this study represented different grades and classes, they followed a similar format (see Figure 3.1). Future research must test this rubric with additional populations and authentic writing samples before these findings can be generalized. Additionally, the exploratory nature of the present study did not allow for the establishment of grading norms – that is, it is still unclear what a strong score for a student in each grade level would be. Finally, during the focus group participants requested anchor papers to model quality writing. Future lines of research may help develop these texts which would not only serve as models for teachers in the classroom, but also increase the reliability of the scores produced with the *Rubric for Scientific Writing*. As the Science and English raters also seemed to have some different interpretations of the rubric elements, when used in an interdisciplinary manner it may be best to have two discipline-specific raters scoring writing samples.

The goal of National Association for Research in Science Teaching (NARST) is "to help all learners to achieve science literacy", and the ability to engage in scientific writing is part of this task (NARST, 2015). Unfortunately, very little writing is currently being taught in middle and high school science classes, and much of what is assigned does not promote authentic science literacy skills (Drew et al., 2014). We must develop tools to support teachers who do not feel prepared to teach and assess scientific writing in order to foster strong scientific writing skills in future generations. The *Rubric for Science Writing* has the potential to aid both science teachers who may currently lack the self-efficacy to teach and assess writing as well as researchers who need a stable measure of students' scientific writing.

CHAPTER IV

WRITING-TO-LEARN INTERVENTION IN SECONDARY SCIENCE CLASSES: FOR WHOM IS IT EFFECTIVE?

Literacy skills are more than just covariates of students' science knowledge: prior test scores in reading can help predict racial and gender gaps in science achievement (Quinn & Cooc, 2015). Furthermore, research has demonstrated that reading and writing in science class can make students more active in the learning process; unlike traditional teaching practices, literacy activities are student-sensitive and promote active student involvement in "constructing their own knowledge through questioning, reprocessing, reflecting, analyzing the ideas ... and drawing conclusions and communicating their own ideas with peers and the teacher" (Alev, 2010, p. 1343). Engaging in scientific literacy is, therefore, fundamental to learning science (Norris & Phillips, 2002), and writing in science class allows students to mirror the work of field scientists (Warren, 2012).

While there have been many calls for content-area literacy instruction, exactly how to operationalize such approaches remains unclear (Miller, 2014). Most content-area literacy instruction involves generic strategies, such as the use of graphic organizers, which could be applied to any subject area (Warren, 2012). This practice is reflected in the types of writing most often assigned in science classes, including fill-in-the-blank worksheets and summarization tasks (Drew et al., 2014). However, even these strategies have historically been met with resistance from practitioners and there is little transfer from what preservice teachers learn in university coursework to what they practice as inservice teachers (O'Brien, Stewart & Moje, 1995). Recent literature argues that these generic literacy strategies are insufficient because students need more than just the skills to gain content knowledge – they must build "an understanding of how knowledge is produced in the discipline" (Moje, 2008, p. 97). This new approach to literacy in content area classes, known as *disciplinary literacy* (Shanahan & Shanahan, 2008), aligns with the Bereiter and Scardamalia's (1987) *knowledge transforming* writing, as both require students to use content to build new understandings. However, unlike Bereiter and Scardamalia, disciplinary literacy typically constrains the types of the writing to authentic forms which are practiced by experts within that discipline.

Using writing as a tool for learning is one method for authentically incorporating writing into discipline-specific coursework. Bangert-Drowns, Hurley, and Wilkinson's (2004) meta-analysis examined the effects of writing-to-learn interventions on academic achievements. When examining studies across grades and disciplines, these researchers found a modest, positive average effect (weighted Cohen's $d = .17$), but the effect varied greatly across studies. Still, these researchers argue that the positive results were too frequent to simply account to sampling error. However, when teasing out these findings further, Bangert-Drowns and colleagues found that students in grades six through eight demonstrated much lower effect sizes (weighted Cohen's $d = -.03$), with four out of the six included studies examining this population of students yielding negative effect sizes. However, the authors could only speculate as to *why* this effect was found.

Examining the interventions reported in the included studies, Bangert-Drowns et al. (2004) found that random assignment of participants and the involvement of the researcher in the classroom did not impact study effect sizes. Other findings from this meta-analysis included that certain study characteristics, specifically longer treatment lengths, two to three assignments per week, and in-class tasks of less than 10 minutes, were ideal. Finally, prompts that encourage metacognitive writing had both positive effect sizes and statistically significant results. In short, there are identified characteristics of writing interventions supporting student achievement in general, although this achievement seems to vary across grade levels.

While the studies reviewed by Bangert-Drowns and colleagues (2004) demonstrate that writing-to-learn can be an effective intervention strategy, most methods require a complete change in how content is taught and would require both teacher and administrator buy-in to be successful. Due to issues of feasibility and teacher resistance (O'Brien et al., 1995) the purpose of the present study is to design and implement a writing-to-learn intervention for science classes, modeled on the best practices identified by Bangert-Drowns et al. (2004), which could be authentically woven into existing classroom practices. Additionally, this study seeks to further examine the middle-school student disparity identified by Bangert-Drowns and colleagues (2004) by using similar intervention procedures as various grade levels to see whether age or another variable explains why some students benefit more from writing-to-learn in science class than others.

Existing Intervention Model: The Science Writing Heuristic

Many of the proposed intervention strategies are not compatible with current classroom practices. This is best exemplified by the Science Writing Heuristic (SWH), which currently dominates the work in the area of scientific literacy and holds great promise for school systems with the flexibility of making large changes away from traditional approaches to science instruction. Specifically, this model reframes science instruction to be more authentic, recursive, and to more closely mirror the work of professional scientists. Traditional laboratory reports contain five predetermined sections - purpose, methods, observations, results, and conclusions - all of which are essentially questions with right and wrong responses as the experiment generally replicates known findings. By contrast, “SWH emphasizes the collaborative and constructive nature of scientific activity... whereby learners are expected to engage in a continuous cycle of negotiating and clarifying meanings and explanations” (Akkus et al., 2013, p. 1748). In this model, students are not explicitly told how to do an experiment, instead they are expected to generate and answer questions. The semi-structured writing product encourages students to generate questions, design procedures to answer those questions, interpret data, propose claims with supportive evidence, and reflect upon how their views have changed (Choi et al., 2010).

Unfortunately, because SWH requires a major shift from traditional laboratory science instruction, it is difficult to determine which components of the model lead to concept knowledge gains. Hohenshell and Hand (2006) attempted to address this question through a quasi-experimental study. All participants experienced the same class content and laboratory experiences; however, the control group completed a traditional laboratory report and the SWH group used the alternative laboratory report template. After writing their respective laboratory reports, students were assessed for their content knowledge. Next, all students completed a writing activity in which they were expected to summarize all the information from the unit. Qualitative data indicates that students in the control group saw little difference between the laboratory reports and summarization activity, indicating that traditional laboratory reports did not elicit analysis and critical thinking skills. Additionally, while there were no significant differences in the groups' performance on the first post-test, a conceptual post-test administered after the summarization activity revealed that students who completed the SWH laboratory report retained more information than their control group peers. These findings suggest that the power of the SWH approach lies in its facility to encourage students to critically consider and use science content rather than memorize and report facts (Hohenshell & Hand, 2006).

However, the success of SWH is highly dependent upon the fidelity of implementation. Akkus and colleagues (2013) implemented SWH to discover the intervention's effects on different levels of student achievement. After providing teachers with two days of training, the authors conducted a series of observations to score teachers as either traditional or embracing the SWH approach. The student and teacher performance were compared against a control group of classrooms. Results demonstrated that high-implementation SWH classrooms performed better than other groups. However, in low-implementation SWH groups only high achieving students outperformed their peers.

When proper training and time for transition are provided, SWH has shown to improve student knowledge and content-area skills (Choi et al., 2010). However, it is clear that SWH requires full teacher buy-in, making small-scale interventions difficult. SWH implicitly redefines the role of the science teacher. In SWH classrooms, the writing product of interest is the lab report, which is achieved through student discovery. In this situation, the teacher's role is to facilitate the discovery rather than directly provide information. Therefore, the implementation of SWH may be faced with resistance, as it would require most schools to alter their entire instructional approach (Akkus, et al., 2013).

Purpose

While models such as SWH have shown to be effective, it is difficult for teachers to radically change their approach without leadership support, and it is equally difficult for principals to unilaterally implement new instructional strategies without teacher buy-in. O'Brien and colleagues (1995) argued that this is because many research based practices – such as SWH – do not account for the complex curriculum, pedagogy, and social climates of schools. Therefore, the field requires *feasible* intervention strategies (i.e. those that could be naturally integrated into existing science curriculum and classroom practices) to support scientific literacy. Additionally, in schools that are not within a crisis situation, enhancements to improve an instructional approach, rather than an overhaul, are most appropriate.

Therefore, the purpose of this study is to combine research-based best instructional practices (described below) into feasible writing-to-learn interventions in grades six through 11 science classes that can be easily replicated by classroom practitioners, therefore addressing major barriers to writing integration. Making an analogy to medicine, a patient suffering from obesity may know that overhauling diet and exercise habits will have a positive impact on their health, however these drastic changes often seem (and often are) impossible to implement and require a large investment of resources. However, if making a small change, such as taking the stairs rather than the elevator, proves to have a modest impact on overall health, the individual may be more likely to make this small change. Additionally, making similar step-wise changes to a curriculum allows researchers to understand the relative impact of each change.

The goal of this study is to identify whether a small change, the equivalent of taking the stairs, can have a modest positive impact on student achievement. Furthermore, as no two students are alike, this study also seeks to determine for whom this writing-to-learn is most effective as well as how affective factors can influence the impact of these instructional strategies. This study is guided by the following research questions:

1. How does the impact of writing-to-learn on scientific writing skills vary across grade levels?
2. For whom are writing-to-learn activities in science class most effective?
3. Can a small change, integrating writing into science classes, have a modest impact on students' scientific writing skills?

Theoretical Foundations of Present Intervention

Bangert-Drowns and colleagues' (2004) meta-analysis included only school-based studies, as the authors argue that "controlled research may yield findings of theoretical interest but not be generalizable to application in natural, complex learning environments such as classrooms" (p. 33). In a similar manner, the goal of this writing-to-learn intervention was to synthesize established best-practices for writing in science class in a manner that would not drastically alter the daily routines and objectives of the teachers. Therefore, when making decisions regarding design, aspects of external validity were given high priority for greater assurance that such findings could be replicated.

Bangert-Drowns and colleagues' (2004) findings indicate that random assignment of participants and researchers' participation in the intervention did not have an impact on effect sizes. Therefore, I worked within the confines of a natural classroom setting (e.g., where students have been previously assigned to classes and randomization is not possible) and integrated myself into the classroom routines in order to observe the intervention implementation. Bangert-Drowns and colleagues (2004) also discovered that effect sizes were directly related to treatment length, with longer interventions yielding stronger results. Finally, these researchers found that, regardless of overall length of treatment, two to three in-class assignments of less than ten minutes each per week were ideal.

A review of existing literature clearly demonstrates that, in the field of science, certain characteristics tend to yield more consistently positive outcomes. While not an exhaustive list, using evidence to form arguments (Klein & Rose, 2010), writing for authentic audiences (Choi et al., 2010), and having multiple opportunities to become proficient at writing-to-learn (Hand et al., 2004a), may support student achievement.

Furthermore, writing that promotes metacognitive thinking about learning may also increase students' knowledge acquisition. These principles, supported by theories underlying metacognitive and *knowledge transformation* writing, drive the current research and are the basis of the intervention studies which follow.

Multiple Opportunities for Writing

Strong effective writing takes time and practice, however as students progress through grades, less time is spent on writing instruction, with very little taking place within the context of science class. Additionally, the majority of writing middle and high school students complete involves very short responses, such as filling in blanks on a worksheet or taking notes (Graham & Harris, 2012). This sort of writing is *knowledge telling* and does not allow for *knowledge transforming* (Bereiter & Scardamalia, 1987). In order for writing-to-learn activities to be effective, students need to engage in writing activities on multiple occasions over a longer period of time.

Unfortunately, the limited amount of research in this area makes it difficult to discern exactly how much writing is enough for learning to take place. In fact, in Miller's (2014) review of content-area writing interventions, only two included studies explicitly investigated how a specific number of writing experiences impacted student learning. In the first of these studies, Hand and colleagues (2004a) examined how the number of writing assignments impacted student achievement. These researchers administered conceptual exams after the first and second writing task to create dependent measures of the students' content knowledge. The findings demonstrated that, regardless of treatment conditions, students developed stronger conceptual understandings after the second writing task. The researchers returned eight weeks after the conclusion of the instructional unit to re-administer the exam and found the same effects – two writing experiences were more effective at supporting student knowledge.

Knaggs and Schneider's (2012) work was also identified by Miller (2014) as investigating the impact of multiple writing experiences. These researchers used vee-maps (complex graphic organizers that separate content knowledge and evidence from arguments and conclusions) to help students evaluate the evidence and ideas in their scientific writing. Three high school classes participated in this study, each completing three writing assignments but using vee-maps as part of the process a different number of times. The authors then rated the students' writing products for understanding of relevant scientific concepts. Results demonstrated that the class that used vee-maps all three times had greater conceptual understandings than the group that used vee-maps once or the group that never used this tool. Considered in concert, these findings show that the more opportunities students have to engage in assignments that require them to think critically as they write, the better prepared they will be think like scientists.

Authentic Audiences

The vast majority of writing completed for science class is for a common, inauthentic audience: the teacher (Gunel et al., 2009). If the teacher is the main audience, students are likely to agree rather than explore opposing points of view (Newell et al., 2011). Furthermore, when students write for the teacher they generally believe they are expected to use advanced terminology and phrasing. This often results in a retelling of information without any deep processing – that is, students can compose an acceptable *knowledge telling* (Bereiter & Scardamalia, 1987) answer without understanding the content by recycling the verbiage of their teacher or the textbook.

However, when students are required to write for a different audience, they must undergo a translation process where information is converted from technical textbook language into something comprehensible to both author and audience. When that audience differs from the author, students undergo yet another round of information translation, which can increase content knowledge. For instance, Gunel and colleagues (2009) demonstrated that when students wrote for a younger audience, they developed deeper conceptual understandings than their peers who wrote for the teacher. Hand and colleagues' (2004a) findings support this claim and provide evidence that students recognize the benefit of writing for a younger audience. In a follow up interview, one student's description of writing for a younger student prior to composing an essay for the teacher illustrates this translation process: "First you dumb it down, learn that, and get the basics down really well, and then you can move into your large textbook definitions" (Hand et al., 2004a, p. 204). Clearly, this student underwent a process of translating information in order to first develop a strong conceptual understanding before composing in academic terms. Together, these studies demonstrate that writing for authentic audiences, especially a peer or younger audience, can support student learning in science classes.

Use of Evidence

Argument development is the core of scientific thought, as the field relies on argument "to establish or justify knowledge claims" (Akkus et al., 2013, p. 1747).

Unlike other fields, science is unique because it is collaborative, so scientists use arguments to work towards a common goal, "advancement of scientific knowledge" (Cavagnetto, 2010, p. 337). To proficiently write in the genre, students need to move beyond lab reports that summarize the replication of established findings (Norris & Phillips, 2002) and begin engaging in the rhetorical conversation of the field (Warren, 2012).

Studies have demonstrated that in scientific writing it is important to develop prompts that require students to use evidence to support arguments. For instance, Choi and colleagues (2010) show that one of the strongest predictors of total writing quality was the relationship between the claims made and evidence provided. These authors argue that, when done correctly, science "writing encourages students to hypothesize, interpret, organize, elaborate, synthesize and persuade others of the ideas that are central to the topic and to the information they collect" (p. 153). Therefore, it is essential for students to integrate evidence into science writing.

Metacognitive Prompts

Metacognitive writing, or writing about one's understandings, has been shown to improve college students' academic achievement, but the impact on students in high school and below has been inconsistent. Part of the struggle is that authors define metacognition differently (Glogger, Holzäpfel, Schwonke, Nückles, & Renkl, 2009). However, generally two types of metacognition appear in high school students' writing – negative and positive comprehension monitoring. Negative comprehension monitoring describes when students identify gaps in their knowledge, whereas positive monitoring involves noticing concepts that are well understood (Glogger, Schwonke, Holzäpfel, Nückles & Renkl, 2012).

Glogger and colleagues (2012) collected learning journals, defined as writing assigned to students to deepen understanding and retention, from high school math and science classrooms and coded the entries for four different learning strategies - three cognitive (rehearsal, organization, and elaboration), and metacognitive strategies. Codings were then correlated with student achievement, and through a cluster analysis the authors discovered trends in the use of learning strategies and achievement. In science, only rehearsal strategies (restating content information) were a strong predictor of student achievement. However, students with the highest levels of achievement in both courses used a combination of quality cognitive and metacognitive learning strategies. The authors argue that students at this level may have the metacognitive ability to identify gaps in knowledge in science, but that does not mean they have the maturity or motivation necessary for remediation. Thus, in this intervention I combine writing for authentic audiences with a metacognitive prompt to build in purpose, and hopefully motivation, for students to engage in metacognitive writing.

The goal of the current study was to combine these previously identified best-practices for writing-to-learn in science class in a multi-grade intervention. I hypothesized that using these four principles – using evidence to form arguments, writing for authentic audiences, providing multiple opportunities to write, and engaging in metacognitive writing – would encourage the success of the intervention, thus allowing me to compare student performance across grades and determine for whom writing-to-learn is an effective strategy. Table 4.1 further describes how each principle was operationalized in the intervention.

Table 4.1

Principles for successful writing-to-learn tasks in intervention

Principle	How Operationalized in Intervention
Using evidence to form arguments	Long writing assignments required students to use content-knowledge as evidence to support an argument.
Writing for Authentic Audiences	Short writing assignments required students to write about what they learned in class to a friend who was absent Long writing assignments were structured so students wrote a letter-to-the-editor correcting a misprint.
Multiple opportunities to write	Students had 10 short writing assignments and 3 long writing assignments over approximately eight weeks.
Metacognitive writing	Short writing assignments prompted students to identify what they had learned in class and explain the concepts in their own words.

Methods

This intervention took place in a small independent school in a rural-type setting in the southwestern United States. The school serves children from Pre-K through 12th grade and has been recognized for its strong college preparatory programs. As this is an independent school, the teachers were not constrained by the same state testing requirements as public school teachers, and therefore had more autonomy and flexibility in their curriculum and teaching practices. This site was specifically chosen for flexibility of teaching practices and progressive mindset, because this work, while research-based, was primarily exploratory.

Participants

The participants were 52 children in grades six through 11 and two science teachers who taught multiple grade levels. Mrs. James (all names are pseudonyms), taught grades six through eight as well as two sections of 10th grade chemistry. All her classes, except for the 8th grade, were able to participate in the intervention. Mr. Devin taught 11th and 12th grade physics, computer science, and robotics courses. However, due to scheduling constraints only his 11th grade physics class participated in the study.

Information and permission slips were sent home to all parents, who were provided opportunities to contact either the teacher or primary investigator with questions. From the five classes (sixth grade, seventh grade, two chemistry classes, and physics) a total of 54 students (representing nearly 86% of the possible participants) returned signed permission slips to participate in the intervention. Students were included in the final analysis if they were present for either the pre- or post- intervention testing, and also completed the first and last long writing assignments (described below). Two students were absent for either the first or last long writing task, resulting in 52 students included in the final analysis.

Intervention Methods

This intervention took place in five science classes representing students in both middle and high school. One of the conclusions reached by Bangert-Drowns and colleagues' (2004) meta-analysis was that writing-to-learn interventions tended to be more successful with older students. However, few researchers have repeated the same intervention at multiple grade levels. Testing the same intervention at multiple grade levels allowed me to closely examine how writing-to-learn impacts students at different grade levels.

Measures. All students completed similar pre- and post- measures to compare the impact of this intervention across groups and to find how variables mediate the effect of this intervention.

Affective Measure. Students completed the *Student Writing Affect Survey* (SWAS) (described in Chapter II) both before and after the intervention to examine the predictive validity of motivational constructs. Furthermore, as students' affect towards literacy can be impacted by interventions (see Wright et al., 2015), two administrations of the survey allowed me to monitor how the intervention impacted students' overall affect towards writing.

General Writing Skills. A lack of proficient writing skills will necessarily hinder the effectiveness of content-area composition. To assess students' overall writing ability, I administered the spontaneous writing subtests (form A) of the fourth edition of the *Test of Written Language* (TOWL) (Hammill & Larsen, 2009) before the intervention. This subtest present students with a picture and allocate twenty minutes for students to write a story based upon the visual. The writing samples were scored for the subscores of contextual conventions and story composition, as well as the spontaneous writing composite score. This test has been normed for ages 9 years to 17 years 11 months (McCrimmon & Climie, 2011).

Scientific Writing Skills. While the TOWL-4 provides information about the students' general writing, I also wanted to gather information about how the students' scientific writing skills change over the course of the intervention. Therefore, I collected copies of the long writing assignments (described below) from all participants and had two certified teachers use the rubric developed and described in Chapter III to evaluate the writing. Each of these assignments required students to write a "letter to the editor" of a science journal correcting a piece of inaccurate information. While the topics varied slightly between the pre- and post-test administration and across classes, the assignments were designed to allow for pre- and post- intervention as well as grade level comparisons.

General Science Knowledge. A child's background knowledge influences his or her ability to write about a topic (Graham, 2006). Therefore, I designed a measure of general science knowledge based upon multiple choice test questions from released versions of the Trends in International Mathematics and Science Study (TIMSS) (IEA International Association for the Evaluation of Educational Achievement, 2013) and the National Assessment of Educational Progress (National Assessment of Educational Progress, 2014). This method of selecting questions mirrors that of Rivard (2004), who created a 20-item multiple-choice test using items from "large-scale science assessments" (p. 428) in the public domain. This methodology of collecting information about students' general science knowledge follows that of Glogger and colleagues (2012).

The questions selected represent those requiring students to apply their knowledge or use scientific reasoning, rather than simply repeating content information. Additionally, questions were designed to evenly target the knowledge of 4^{th}, 8^{th}, and 12^{th}

grade students. As my participants were in grades six through 11, this variety of questions helped prevent ceiling and floor effects. Prior to administering these measures, two certified science educators (independent of the study teachers) reviewed the items to ensure the questions are all valid and pertinent to relevant scientific information.

Intervention Procedures. Approximately one week before beginning the intervention, I administered the pre-intervention assessments during students' study hall periods**.** I implemented the writing-to-learn intervention within six to eight weeks (depending upon class and school scheduling, see Appendix B for sample intervention schedule). During this time, students engaged in two types of writing activities two to three times per week, but in total, each student received the same amount of intervention time and writing tasks.

The majority of the writing activities (described as "short writing") asked the students to spend five to ten minutes at the end of class summarizing what they learned for a peer who was absent. The goal was that this authentic audience would encourage students to use to process the information rather than simply retell, thus cuing metacognition. Each class completed this activity 10 times over the course of the intervention.

The second type of writing activity (identified as "long writing") presented the students with text from a mock-journal. The text incorrectly stated a fact about the topic of study, and students will write a letter-to-the-editor identifying what was wrongly stated, providing the correct information, and explaining why their information is correct. The goal of this activity was to encourage students to use evidence while writing for an authentic audience. Each class completed one of these long writing assignments three times – once towards the beginning of the intervention, once near the middle, and once at the end – and students had a minimum of 20 minutes to complete this activity. Teachers held all students (both those participating in the study and those who were not) responsible for all writing activities as if they were any other classroom assignment – that is, student noncompliance would be met with the same consequences as not completing a teacher-generated assignment. This was not deemed coercive by the Institutional Review Board as it feel within normal classroom practices and all students were treated in the same manner.

Fidelity Measures. I worked closely with classroom teachers to schedule intervention times that would not greatly interfere with classroom instruction. I originally scheduled 12 short writing assignments, with the understanding that unexpected school events or changes in lesson plans may prohibit implementation. For all classes, at least 10 short writing assignments were completed during the intervention period.

I attended most class periods where writing took place. When I was not available, another graduate student or undergraduate research assistant went in my place to ensure fidelity. The researcher would assist the teacher in classroom activities (such as handing out papers, etc.) so the students quickly became used to the presence of the additional adult in the room. At the end of each class period, the researcher completed a fidelity form that recorded general classroom activities, student engagement during class and writing time, and amount of time students were allotted for writing (See Appendix C).

Additionally, the researcher took informal field notes and recorded any unusual disturbances or activities (see Table 4.2).

As the study progressed, teachers asked to make slight adjustments to the procedures to better adapt to their classroom process. As the goal of this study was to examine how this intervention would work in a real classroom setting, teachers were encouraged to "take charge" and modify the writing to fit the needs of their students. I briefly explained the principles supporting the writing tasks (see Table 4.1), and requested that the writing assignments must still adhere to those best practices identified through previous research (i.e., using evidence to form arguments; writing for authentic audiences; multiple opportunities to write; metacognitive writing). These modifications were noted by the researchers observing the class, and are summarized in Table 4.2.

Writing Scoring

Students' three long writing samples were first scored using the *Rubric for Scientific Writing* (RSW). This tool, based upon the expectations for writing established by the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013) and Common Core State Standards (National Governors Association, 2010), evaluates six different aspects of scientific writing on a scale of zero to three: Claim; Evidence and support; Analysis of content; Organization; Audience; and Presentation of writing. The first three elements correlate into an overall score called *Scientific Rhetoric* and the later three into an overall score called *English Composition.* Each sample was scored by at least one rater with an English/language arts education background and one rater with a science education background, and where raters disagreed I calculated an average score.

Additionally, I calculated the average scores by grade and school level (middle or high school).

Table 4.2

Observations and modifications noted on fidelity records

Class	**Mean minutes for short writing tasks**	**Short Writing Task Observations and Modifications**	**Mean minutes for long writing tasks**	**Long Writing Task Observations and Modifications**
6th Grade	9.30 (2.63)	#1 – Teacher provided candies during writing time	18.33 (5.51)	#3 – Day before school vacation; many students required redirection
7th Grade	9.44 (2.96)	#6 – Valentine's Day; students distracted #9 – Students could earn bonus points on a quiz for their writing	12.66 (6.43)	#1 – Integrated as part of a quiz #3 – Day before school vacation
Chem Class A	7.77 (1.72)	#1 – Substitute teacher #6 – Students could earn bonus points on a quiz for their writing	19 (18.25)	#2 – Students could earn bonus points on a lab assignment for their writing #3 – Teacher used assignment in place of a quiz, provided students feedback before submission
Chem Class B	9.20 (2.82)	#6 – Students could earn bonus points on a quiz for their writing	23.66 (14.36)	#3 – Teacher used assignment in place of a quiz, provided students feedback before submission
Physics	8.00 (2.00)	#6 – Substitute teacher	15.66 (9.02)	#2 – Valentine's Day; students distracted by activities (e.g., singing valentines in the classroom #3 – Teacher modified prompt to use as a quiz question

Note. Standard deviations displayed in parenthesis next to means.

However, as the purpose of this study was to examine how variables other than age impact the effect of writing-to-learn, I also examined the short writing samples to develop student profiles. While all students were provided the same prompt, how they responded to those prompts varied greatly. Therefore, I coded these short writing samples on a number of variables. First, I counted the words produced in each writing sample. Then, I counted the number of activities listed, key scientific vocabulary used, and number of scientific facts explained. I also noted whether students created some sort of visual, graphic, or formula to represent the information. Table 4.3 provides detailed descriptions and examples of these codes.

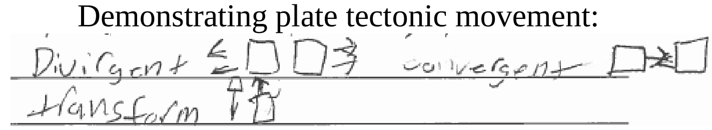
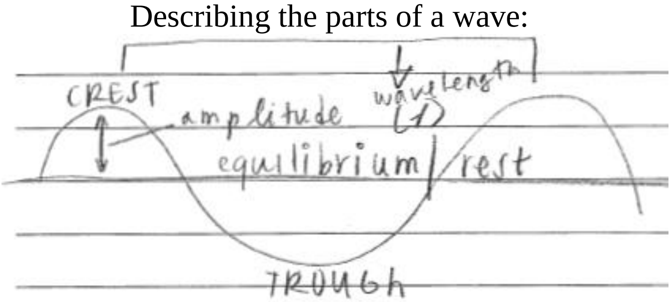
Data Analysis

Existing literature suggests that students in lower grades would not respond as well to writing-to-learn strategies as their peers, I first chose to examine the data by grade level. Therefore, I combined the students in grades six and seven (n = 26) to form a "middle school" group, and likewise combined the older students (n = 27) to form a "high school" group. I then conducted ANOVAs using the two groups' scores on the long writing tasks as dependent variables to examine the impact of grade level.

Cluster Analysis. As one aim of this study is to determine *for whom* writing-to-learn interventions in science are most effective, I also conducted a cluster analysis. Cluster analysis is a variable-centered methodology that connects participant scores to create profiles. The purpose of a cluster analysis is to organize "cases" (in this instance, students), by features so the resulting clusters exhibit high internal homogeneity. Cluster analysis is simply descriptive and atheoretical, however we can use group membership

Table 4.3

Short Writing Coding

Code	Description	Student Examples
Activity	Student summarizes an activity done in class. Generally includes an active verb.	*Today in science we did a project on Chapter 4* *We were paired into groups to start a lab* *Today we went over covalent bonds.*
Scientific Vocabulary	Words or short phrases used to describe the course content.	*I learned that density is the amount of matter per unit of volume* *Natural selection is a big part of an organism's life.* *We learned about ionic compounds, their formulas, and how their names changed when ionic bonding occurs.*
Scientific Facts	Course content paraphrased in students' words. Individual facts were identified if provided sufficient information to create a typical test question.	*Gregor Mendal used cross-pollination in plants and learned about dominant and recessive traits.* *The water cycle is also called the hydro-cycle, because hydro means water* *With a pendulum, it doesn't matter how high you hold the string before you let it swing, because it will swing at the same speed.*
Visuals	Tables, charts, graphics, or formulas used to represent course content. These were coded as either included (1) or not included (0)	Demonstrating plate tectonic movement:  Describing the parts of a wave: 

to describe the cases using statistical analyses such as ANOVA (Hair & Black, 2000). Cluster analysis has been used in previous literacy research for a variety of purposes, such as describing different approaches to shared book readings (e.g., Haden, Reese, & Fivush, 1996; Hammett, Van Kleeck, & Huberty, 2003), and identifying how writing traits can predict other variables (Glogger et al., 2012; Roid, 1994).

In cluster analysis, the variables should be measures that can be used to profile the cases (Hair & Black, 2000). For this reason, I used the pre-measures from the SWAS, General Science Knowledge test scores, and students' scaled scores on the TOWL as cluster variables. As I was also interested in how what students wrote during the intervention impacted their growth, I also included the average number of words, activities, scientific vocabulary, scientific facts, and visuals produced in the short writing assignments as variables in the cluster analysis (see Table 4.4).

Table 4.4.

Variables included in cluster analysis

Variable	Rationale for inclusion
Average SWAS Self-Efficacy Score	Student self-efficacy for writing will impact how he or she approaches writing tasks. As this score was highly correlated with the SWAS factor scores, only this measure was included (Meyers et al., 2013)
Average TOWL	Students' overall writing ability would impact what they did during the intervention
General Science Knowledge Scores (Pre-test)	It can be assumed that students with more background knowledge would feel more confident in scientific writing.
Number of activities*	While all students were provided similar writing tasks, how they responded to those tasks differed greatly. These measures help describe what students did during the intervention
Number of scientific vocabulary*	
Number scientific facts*	
Number of words*	
Percent of writing samples that include a visual	

* These items represent average counts across the 10 short writing samples

I examined four different models, respectively fitting the data into three, four, five, and six different profiles. For each model, I conducted ANOVAs to identify group differences in the outcome variables. The model finally selected contained five clusters and provided the most explanation for why some students achieved more than others.

Results

Student Writing Affect Survey

Students completed the Student Writing Affect Survey (*SWAS*) twice, approximately eight weeks apart (see Table 4.5). Examining both statistical significance and effect sizes of the pre- and post- intervention scores demonstrates that most factors measured by the SWAS did demonstrate growth. The only shift that was statistically significant was *attitude*, which also yielded a positive effect size (Cohen's d = .25). Additionally, only students' tendency to avoid writing yielded a negative effect size of -.22, indicating students may avoid writing more, but this finding was not statistically significant. These mixed results suggest that, overall, the intervention had a small, positive impact on students' overall self-efficacy for writing; consistent with previous research long-held beliefs take substantial time and evidence to change (Chinn & Brewer, 1993). For the remainder of the analyses I used a composite score averaging students' pre- and post- intervention means to measure the impact of writing self-efficacy on the effectiveness of the intervention. Utilizing an average score helped to ensure that this variable represented students' writing self-efficacy, rather than perhaps reflecting a student's good or bad mood on a particular day.

Table 4.5

SWAS Pre and Post intervention Scores

	Pre-Mean (SD)	**Post-Mean (SD)**	***p* value**	**Cohen's *d* effect size**
Attitude	2.39 (.81)	2.58 (.73)	.031	.25
Value	2.90 (.82)	3.03 (.61)	.233	.17
Social	2.44 (.79)	2.64 (.75)	.108	.25
Avoidance	2.33 (.80)	2.17 (.66)	.090	-.22
Confidence	2.95 (.83)	3.14 (.64)	.074	.26
Overall Self-Efficacy	2.59 (.72)	2.73 (.57)	.140	.20

Test of Written Language (TOWL)

For statistical analysis, I converted the students' spontaneous writing scores on the TOWL to scaled scores and percentile ranks. This allowed for direct comparison across grades as all students were measured on the same metric, relative to their age. I then ran *t*-tests to confirm that no statistically significant differences existed between the middle and high school students' writing abilities prior to the intervention. Average scaled scores and standard deviation are detailed in Table 4.6.

Long Writing Assignments

I calculated students' average scores for scientific rhetoric and English composition, as well as overall writing scores, for all three writing assignments. I first examined the students' growth from the first to the third writing tasks by calculating Cohen's *d* effect sizes (see Table 4.7). These results demonstrated moderate to strong growth by both middle and high school students in the area of scientific writing. Convergent with previous literature, the high school students did demonstrate more growth than the middle school students; however, the differences were not so drastic as to discount the effectiveness of the intervention for the younger population. Therefore, I decided to examine other student characteristics that might explain for whom writing-to-learn in science class is most effective.

Table 4.6

Scaled Scores and Percentile Ranks from TOWL

	Writing Conventions		**Story Composition**		**Spontaneous Writing Subtest**	
Grade	**Scaled Score**	**Percentile**	**Scaled Score**	**Percentile**	**Scaled Score**	**Percentile**
6	13.67 (2.68)	82.17 (23.32)	14.67 (4.37)	81.67 (31.02)	14.16 (3.34)	81.92 (25.11)
7	13.22 (2.76)	76.78 (23.77)	12.89 (4.31)	75 (31.12)	13.06 (3.46)	75.89 (25.66)
Middle School	13.52 (2.76)	80.37 (23.16)	14.07 (4.35)	79.44 (30.62)	13.8 (3.35)	79.91 (24.96)
10	11.92 (1.93)	70.42 (19.74)	12.25 (5.17)	63.75 (39.29)	12.08 (3.26)	67.08 (27.05)
11	12.67 (1.92)	77.67 (18.46)	14.42 (4.01)	83.08 (28.04)	13.54 (2.73)	80.38 (22.32)
High School	12.28 (1.88)	74.08 (18.65)	73.84 (34.14)	13.32 (4.56)	12.8 (2.97)	73.96 (24.68)

Note: Not statistically significant differences ($p > .05$) existed between any groups.

Table 4.7

Cohen's d Effect sizes for Long Writing Scores (First to third writing tasks)

	Middle School	**High School**
Scientific Rhetoric	.52	.68
English Composition	.11	.17
Overall Score	.32	.48

**Score change statistically significant at p < .05*

I first conducted an ANOVA to see if differences existed between the middle and high school students' scores on the long writing tasks (See Table 4.8). In all cases the assumption of homogeneity of variances was met (*p* ranged from .322 to .899). These results demonstrate no significant differences in writing achievement based upon grade level.

Table 4.8

ANOVA Results for Long Writing Scores by Grade Level

Long Writing #1							
				Middle School (*n* = 25)		**High School (*n* = 26)**	
	df	*F*	*Sig.*	*M*	*SD*	*M*	*SD*
Scientific Rhetoric	50	1.096	.300	1.48	.71	1.68	.67
English Composition	50	.827	.368	1.59	.70	1.75	.52
Overall Score	50	1.088	.302	1.54	.68	1.72	.56
Long Writing #2							
				Middle School (*n* = 26)		**High School (*n* = 20)**	
	df	*F*	*Sig.*	*M*	*SD*	*M*	*SD*
Scientific Rhetoric	45	1.599	.213	2.18	.60	2.40	.56
English Composition	45	1.259	.268	1.99	.53	2.18	.56
Overall Score	45	1.581	.215	2.09	.54	2.29	.53
Long Writing #3							
				Middle School (*n* = 23)		**High School (*n* = 27)**	
	df	*F*	*Sig.*	*M*	*SD*	*M*	*SD*
Scientific Rhetoric	49	3.952	.053	1.80	.51	2.09	.51
English Composition	49	1.620	.209	1.66	.54	1.83	.41
Overall Score	49	2.971	.091	1.73	.50	1.96	.44

Examining these results by grade resulted in an unexpected pattern – nearly all students' scores went up from long writing one to long writing two, and then dropped to lower than two but higher than one for long writing three. Some tentative explanations for this phenomenon were observed in the qualitative data – for instance all students completed the final writing task the week before a week-long vacation and were somewhat distracted – however, further analysis is needed to better interpret the impact across the intervention period. Therefore, I decided to conduct a cluster analysis to better describe for whom writing-to-learn is an effective intervention.

Cluster Analysis

I conducted a hierarchical cluster analysis to describe the similarities between the participants in the writing-to-learn intervention. All variables were converted to z-scores to ensure one variable did not have undue influence due to scale (Meyers et al., 2013). I decided to use the Ward method and squared Euclidian distance, because the Ward method has strong discriminating power (Hammett et al., 2003), and these methods have been used in other writing-to-learn intervention studies (Glogger et al, 2012). The squared Euclidian distances between cases ranged from 3.428 to 93.613.

Cluster Descriptions. General descriptive information about the participants in each cluster is detailed in Table 4.9.

Table 4.10 details the cluster z-scores on each of the input variables. All group differences were significant at the .01 level. I have labeled each cluster based upon the groups' pre-intervention measure scores as well as how they distinguished themselves during the intervention.

Table 4.9

Participant Descriptives by clusters.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
N	19	7	10	7	9
% Female	58	29	50	29	0
Mean Grade (SD)	6.89 (1.60)	8.71 (1.60)	9.20 (2.04)	10.71 (.76)	8.44 (2.35)
% ESL	0	0	10	14	22

Table 4.10

Cluster Analysis In-put Variables (z-scores)

	Cluster 1: Activities	Cluster 2: Poor Self-efficacy	Cluster 3: Strong Scientific Knowledge	Cluster 4: Visuals	Cluster 5: Avoiders
SWAS	.55 (.65)	-1.35 (.61)	.65 (.68)	-.05 (.46)	-.89 (.85)
TOWL	.56 (.56)	-.04 (1.07)	.24 (.80)	.25 (.34)	-1.38 (.69)
GSK Score	-.36 (1.10)	.40 (.38)	.76 (.56)	-.07 (.80)	-.35 (1.22)
Activities	.66 (.76)	-.45 (1.37)	-.24 (.84)	-.80 (.86)	-.06 (.68)
Scientific Vocabulary	-.15 (.75)	.46 (.70)	1.24 (1.15)	-.88 (.43)	-.91 (.67)
Scientific Facts	-.10 (.59)	.85 (.79)	1.15 (.71)	-.88 (.43)	-.91 (.67)
Words	.03 (.57)	.30 (1.06)	1.24 (.81)	-.84 (.51)	-.88 (.57)
Visuals	-.60 (.20)	-.62 (.14)	.39 (1.06)	1.90 (.52)	-.09 (.55)

Note. Standard deviations are displayed next to means in parenthesis

Cluster Outcomes. I conducted one-way ANOVAs to examine the differences between the five group means on the long writing assignments. These findings are summarized in Table 4.11. Except where indicated, the homogeneity of variances assumption was met in all cases, ($p > .01$). Tukey-Kramer post-hoc analysis revealed that for almost all first and second long writing scores, the *Avoiders* were statically different from all other clusters, except the *Visuals* cluster. The post-hoc analysis also demonstrated that, for the second writing scientific rhetoric score, the *Strong Scientific Knowledge* and *Avoiders* clusters were the only two with statistically significant differences (p = .009). However, on the third writing task, the post-hoc analysis showed no statistical differences between any of the clusters.

Because statistical significance is greatly influenced by sample size (Thompson, 2006), I calculated effect sizes for each groups' growth from the first to the third writing assignment (see Table 4.12)

Table 4.11

ANOVA Results for Long Writing Scores by Cluster

				Cluster 1: Activities		**Cluster 2: Poor Self-Efficacy**		**Cluster 3: Strong Scientific Knowledge**		**Cluster 4: Visuals**		**Cluster 5: Avoiders**	
Long Writing #1													
				(n = 18)		(n = 6)		(n = 10)		(n = 7)		(n = 8)	
	df	*F*	*Sig.*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Scientific Rhetoric	48	3.565	0.013	1.64	0.47	2.02	0.58	1.86	0.86	1.47	0.56	0.95	0.55
English Composition	48	3.786	0.01	1.8	0.34	1.88	0.51	1.86	0.79	1.75	0.44	1.03	0.57
Overall Score*	48	15.542	0.025	1.72	0.33	1.95	0.54	1.86	0.82	1.61	0.46	0.99	0.51
Long Writing #2													
				(n = 19)		(n = 6)		(n = 9)		(n = 7)		(n = 8)	
	df	*F*	*Sig.*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Scientific Rhetoric	44	3.263	0.021	2.25	0.62	2.42	0.4	2.61	0.42	2.29	0.76	1.67	0.29
English Composition	44	4.14	0.007	2.1	0.56	2.22	0.31	2.35	0.47	2.04	0.6	1.43	0.09
Overall Score	44	4.177	0.006	2.18	0.55	2.32	0.36	2.48	0.39	2.17	0.67	1.55	0.17
Long Writing #3													
				(n = 17)		(n = 6)		(n = 10)		(n = 7)		(n = 9)	
	df	*F*	*Sig.*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Scientific Rhetoric	48	0.829	0.514	1.84	0.54	2.23	0.43	2.02	0.43	2.06	0.64	1.83	0.61
English Composition	48	1.264	0.299	1.66	0.5	2.02	0.46	1.92	0.43	1.71	0.4	1.58	0.52
Overall Score	48	1.044	0.395	1.75	0.5	2.12	0.44	1.97	0.42	1.89	0.49	1.7	0.53

*Homogeneity of variance assumption not met (Levene's statistic p = .002). Thus, I conducted a Welch's ANOVA and Games-Howell post-hoc (Yiğit & Gökpinar, 2010).

Table 4.12

Cohen's d Effect sizes for Long Writing Scores by Cluster (First to third writing tasks)

	Cluster 1: Activities	**Cluster 2: Poor Self-efficacy**	**Cluster 3: Strong Scientific Knowledge**	**Cluster 4: Visuals**	**Cluster 5: Avoiders**
Scientific Rhetoric	0.39	0.41*	0.24	0.98	1.52*
English Composition	-0.33	0.29	0.09	-0.10	1.01
Overall Score	0.07	0.34	0.17	0.59	1.36*

**Score change statistically significant at $p < .05$*

Cluster Descriptions. In the following section, I summarize each cluster's pre-intervention skills, behavior during the intervention, and outcomes.

Cluster 1: Activities. Cluster 1 contained the most students (36%), and the majority of the participants in this cluster were middle school students. These students began the intervention with slightly higher than average scores on the SWAS and TOWL, indicating they had strong writing skills and positive feelings towards writing. During the intervention, the majority of their writing included descriptions of activities, with fewer instances of scientific vocabulary and facts than many of their peers. The teachers noticed this trend early on and encouraged students to focus on what they *learned* rather than what they *did* in their writing. In spite of this extra level of support, these students continued to describe classroom activities. Therefore, this cluster is called the *Activities* cluster.

Despite the concerns of the teacher, these students' writing scores did improve slightly over the course of the intervention. Their overall long writing scores demonstrated a very small, positive, effect size (*d* = 0.07). While their English composition scores did drop (*d* = -0.33), no support for this aspect of writing was provided during the intervention so significant growth was not expected in this area. By contrast, their scientific rhetoric scores increased, demonstrating an effect size of *d* = 0.39.

Cluster 2: Poor Self-Efficacy. The second cluster, comprised of four middle school students and three high school students, scored by far the lowest on the SWAS, indicating they had the most negative feelings towards writing. Their TOWL scores, by contrast, demonstrated average writing abilities, and they had the second highest scores on the General Science Knowledge assessment.

During the intervention, however, these students wrote slightly more words than average, and the group rated second in terms of number of scientific facts present in their writing. This group demonstrated growth in all aspects of writing on the long writing tasks. The strongest effect size was in scientific rhetoric (*d* = 0.41), indicating that the intervention supported their ability to express their scientific knowledge in writing.

Cluster 3: Strong Scientific Knowledge. The third cluster of 10 students, including one ESL student, demonstrated the highest score on the General Science Knowledge test and SWAS, indicating that these students valued writing and had significant background knowledge in science. This was especially evident during the intervention, when these students wrote the most words comprised largely of scientific vocabulary and facts. Together, these findings led this group to be known as the *Strong Scientific Knowledge* cluster.

Compared to the other clusters of students, this group's growth was the most limited. Their overall growth yielded an effect size of d = 0.17, with modest but promising growth in both scientific rhetoric and English composition. These results indicate that the intervention supported the learning of all students, including high achievers.

Cluster 4: Visuals. Seven of the participants, including one ESL student, had near average scores on all three pre-intervention measures. This group of students was comprised of all high school students enrolled in physics classes. While they wrote fewer words than their peers, they created the most visuals of the group, nearly two standard deviations above the mean. Therefore, this cluster is called *Visuals.*

This group's English composition scores did not change drastically during the intervention, which again is not surprising considering there was no direct instruction in this area and the students produced some of the fewest words. However, this group's scientific rhetoric scores demonstrated strong, positive growth (d = 0.98), indicating that the students improved in their ability to convey scientific concepts in writing.

Cluster 5: Avoiders. The nine students who make up Cluster 5 began the intervention with lower self-efficacy towards writing (as indicated by the SWAS) than most of their peers. Additionally, these students scored the lowest on the TOWL, indicating that writing may be an area of difficulty. This group, comprised of five middle school students and four high school students, wrote the least number of words during the intervention, and logically also used the fewest scientific vocabulary or facts in their writing. This group is thus called *Avoiders*.

Post-hoc analysis of these students' first and second writing task scores indicated that this group was statistically different from all other clusters, except the *Visuals* cluster. Despite this fact, and the concerning pre-intervention scores, this group of students made the most growth during the intervention, with all aspects of their writing yielding effect sizes of 1.0 or higher. Additionally, this is the only group whose scores from the first and third writing tasks demonstrated any statistically significant difference. As statistical significance is difficult to achieve with small sample sizes (Thompson, 2006), this finding is especially noteworthy.

Discussion

Previous findings have indicated that students in middle school were less likely than high school students to benefit from writing-to-learn interventions (Bangert-Drowns et al., 2004). However, few researchers have attempted to implement the same intervention in both middle and high school classes, thus making cross-study comparisons difficult. Furthermore, many of the most promising writing-to-learn strategies for science classes, such as the Science Writing Heuristic, require a fundamental shift in how science is taught and is not feasible for many teachers. Therefore, the purpose of the present study was to examine the effects of a minimally intrusive intervention on students' scientific writing skills at both the middle and high school levels.

My first research question considered: *How does the impact of writing-to-learn on scientific writing skills vary across grade levels*? Examining the results by grade level revealed that all students tended to do better on the second writing task than the first or the third, and that the third writing task tended to earn the second highest score. However, there were no statistically significant differences between the performance of the middle and high school students. This contrasts existing literature, which generally demonstrates that writing-to-learn interventions are less effective for middle grade students (Bangert-Drowns et al., 2004). In the present study, all five clusters demonstrated a similar pattern of growth, but also that the impact of the writing-to-learn intervention varied depending upon students' pre-intervention skills and feelings towards writing along with *what* they actually did during the intervention.

Answering my second research question (*For whom are writing-to-learn activities in science class most effective?*), I examined the results of the cluster analysis. All five clusters of students demonstrated growth in the area of scientific rhetoric (Cohen's *d* range .24 to 1.51). This is especially noteworthy for clusters 2 and 5, who both began the intervention with lower than average writing skills and self-efficacy for writing. This finding indicates that the intervention procedures supported all students' abilities to write and communicate scientifically.

By contrast, not all groups made growth in the area of English Composition; in fact, two groups (Cluster 1: Activities and Cluster 4: Visuals) demonstrated negative effect sizes in this area (Cohen's d = -0.32 and -0.08, respectively). However, this finding is not altogether surprising as there was no instruction in this area during the intervention. While research has demonstrated that students require multiple opportunities to write (Bangert-Drowns et al., 2004), students also require feedback and instruction in order for writing to improve. Additionally, certain aspects of writing considered valuable in science may hold less value within English composition, due to the influence of disciplinary expectations (Shanahan & Shanahan, 2008). For instance, literature demands the use of imagery through language, whereas science demands the use of actual images.

All clusters of students demonstrated some growth in their writing, but the students for whom the intervention was most effective were in the *Visuals* and *Avoiders* clusters. The *Visuals* cluster demonstrated strong growth in scientific rhetoric (d = 0.98), and the *Avoiders* cluster demonstrated effect sizes of at least 1.00 in all areas of writing. It should be noted that three out of the four ESL participants were in these two clusters; however, this sample size is insufficient to make broader claims about the impact of the intervention on second language students.

The *Visuals* students created the most graphics in their writing, which may have partially been an effect of the physics content covered in class during the intervention. The students were studying waves, and their teacher encouraged them to create mental pictures of the waves during class discussions. Additionally, graphical representations of information are common in K-12 science texts (Slough, McTigue, Kim, & Jennings, 2010) and research has demonstrated that older students are more likely to consider graphics when reading science materials (McTigue, 2009). Future research should investigate the relationship between students' use of graphics while reading science texts, the graphics they produce while writing scientifically, and the overall quality of their scientific writing.

The *Avoiders* group of students began with some of the lowest pre-intervention measures scores and, overall, produced the least during the intervention. The majority of their writing was simply descriptions of the activities from class, with some graphics to supplement the writing. However, these students made the most growth out of all the clusters, and they were the only group whose growth demonstrated both practical and statistical significance. Together, these findings do suggest that the intervention may be most effective for struggling and at-risk students.

One of the more interesting results is that while the students' scores on the first and second long writing assignments differed by clusters, the third long writing tasks demonstrated no statistically significant differences between the groups. Essentially, with minimal intrusion in the science class, students at all ability levels became more proficient scientific writers. While further research is necessary before these results can be generalized, this finding indicates that the intervention may help close achievement gaps in science literacy between struggling and achieving students in science classes.

Conclusion

The purpose of this study was to combine these research-based best instructional practices into feasible writing-to-learn interventions in grade six through 12 science classes that can be easily replicated by classroom practitioners, therefore addressing major barriers to writing integration. My results indicate that through exposure to relatively brief writing tasks, students' ability to engage in scientific rhetoric improved. These findings suggest that writing-to-learn strategies can be strategically implemented in middle and high school science classes to support scientific literacy development.

Furthermore, with minimal training the science teachers were able to take ownership of the intervention strategies and adapt them to fit their classroom practices. I described the principles behind the writing tasks (i.e., using evidence to form arguments, writing for authentic audiences, providing multiple opportunities to write, and using prompts that encourage metacognition) to the teachers, and therefore any modifications made did not interfere with the overall effectiveness of the intervention. Many practitioner publications that make recommendations for supporting science literacy are largely atheoretical (Wright et al., 2015). However, these findings indicate that when teachers are provided with theoretical foundations, they can make modifications that support the needs of their students while preserving the integrity of instructional practices.

Together, the findings of this study suggest that writing-to-learn can be an effective tool for supporting scientific literacy in both middle and high school science classes. Future researchers should expand this study to examine the effect on other populations, including second language learners. However, this study provides promising results that writing-to-learn can be efficiently and effectively implemented in current science curriculum practices.

CHAPTER V

CONCLUSIONS

In 2012, Fang argued that "Language enables experience to be transformed into meaning. It is through this transformation that people come to understand their experiential world, and the outcome of this transformation is what is called *knowledge*" (p. 20). Scientists are no exception to this phenomenon. To create scientific knowledge, researchers in the field must be able to turn their experiences into words that can be shared with others. In this view, scientific writing is an essential skill of a professional scientist.

Disciplinary literacy researchers have demonstrated that the genre of scientific writing differs from other subjects (Fang, 2012; Shanahan & Shanahan, 2008). Unfortunately, in middle and high school grades most writing instruction takes place in English class (Graham & Harris, 2012) and many science teachers report incorporating only minimal writing in their lessons (Drew et al., 2014). Therefore, if we hope to prepare students to one day work in scientific fields, we must develop methods for teaching writing in today's science classes.

This dissertation sought to address three major problems facing writing instruction in middle and high school science classes. First, without a measure that produced reliable and valid estimates of students' affect towards writing, it is impossible to determine the role of poor self-efficacy in student achievement, or to understand students' writing self-efficacy development. Secondly, researchers and teachers need to be able to quantify scientific writing achievement to both monitor the impact of interventions as well as provide feedback to support student growth. Finally, the field requires methods for using writing in science class that support scientific learning and are feasible in the reality of today's schools. By developing the Student Writing Affect Survey and Rubric for Scientific Writing, I have begun to address first two issues. Furthermore, the instructional strategies proposed in Chapter 4 suggest that a minimally intrusive intervention can help many students become stronger scientific writers.

Considered in concert, the conclusions from these three studies reveal three themes, which will be described in the remainder of this chapter. First, these studies demonstrate that by moving students from *knowledge telling* writing to a more complex *knowledge transforming* writing, we can engage them in authentic scientific rhetoric (Bereiter & Scardamalia, 1987). Secondly, while there are two distinct definitions of scientific literacy (that is, the fundamental sense and the derived sense; Norris & Phillips, 2002), supporting students' ability to read and write scientifically may help develop their general scientific knowledge. Finally, the literacy research field requires both measures and methods for implementing writing-to-learn in science classes.

Knowledge Telling Vs. Knowledge Transforming Writing

Most of the writing currently being completed in middle and high school science classes is composed of assignments such as note taking, fill-in-the-blank worksheets, or graphic organizers (Drew et al., 2014) – tasks that essentially require students to just tell what they have learned. This sort of *Knowledge Telling* writing, as defined by Bereiter and Scardamalia (1987) is insufficient for preparing students to participate in the scientific community. We need to be encouraging *Knowledge Transforming* writing, or writing that requires students to use their knowledge to build new ideas and arguments. This *Knowledge Telling* and *Knowledge Transforming* dichotomy highlights the fact that writing, in and of itself, does not necessarily require the author to create new ideas. Therefore, as educators and researchers we must build environments where students can learn to transform their knowledge through writing.

The first step in creating a space for *Knowledge Transforming* writing requires examining how the students feel about writing. Chapter 2 demonstrated that students with stronger self-efficacy for writing were more engaged in low-stakes writing tasks. If a child is avoiding writing altogether, he or she is unlikely to practice and refine his or her writing abilities. A teacher must be aware of which students have this tendency. Even for writers who have developed basic skills, Bandura (2001) tells us that "maintaining proficiency … demands continued investment of time, effort, and resources in self-renewal" (p.13). Therefore, we must measure students' self-efficacy towards writing to identify students who may struggle to become skilled *Knowledge Transforming* writers, as well as monitor how students' feelings may change during the course of instruction.

Currently, less than one third of science teachers in the United States report teaching students to write scientifically (Drew et al., 2014), yet these teachers are the most knowledgeable to do so in terms of discipline specificity. In order for students to become skilled at writing for a scientific audience, they must understand the expectations for this genre, and identify where they need to improve their writing. The

Rubric for Scientific Writing provides this support for students by outlining the expectations of the field. Furthermore, it offers a tool for teachers (who may not have received any training in writing pedagogy) to provide specific, relevant feedback to help students improve their writing. This sort of support will help students engage in scientific argumentation, a key component of scientific literacy (NGSS, 2013), and promote *Knowledge Transforming* writing.

Finally, the results of the intervention described in Chapter 3 demonstrate that a feasible intervention can help students engage in *Knowledge Transforming* writing. The participants in this study showed an average Cohen's *d* effect size of 0.60 in Scientific Rhetoric across all grades. This score reflects a growth in students' ability to develop scientific arguments in writing and use evidence to support their claims. In essence, students were better able to use the knowledge presented in class to build new ideas, which is the foundation of *Knowledge Transforming* writing.

Fundamental Science Literacy and Derived Science Literacy

The word *literacy* has come to take on two meanings – the ability to read and write, and being knowledgeable about some topic or subject. *Scientific literacy*, therefore, can describe both an individual's ability to read and write within the genre of science (the fundamental definition) as well as their state of being knowledgeable about science (the derived definition; Norris & Phillips, 2002). Too often, these two definitions have been at odds, with teachers and researchers debating which should receive preference in the classroom (e.g., see Gillis, 2014). I argue, however, that these two concepts need not be in competition, as supporting students' fundamental ability to read and write scientifically will enhance their general knowledge about science.

While writing holds great potential to develop both students' fundamental and derived science literacy, it is only effective if students choose to engage in the task. Understanding why students choose to or not to engage in literacy is the first step in understanding some of the barriers facing their engagement in fundamental scientific literacy. As shown in Chapter 2, students with strong self-efficacy for writing produced more scientific text. While motivation for literacy activities can vary depending upon the situation and topic at hand (Guthrie et al., 2006a), it is clear that a relationship exists between students' general feelings towards writing and the writing they create in science. Additionally, disciplinary literacy is conceptualized to be built directly upon the foundation of basic literacy skills (e.g., fluency), and broader generic skills (e.g., comprehension monitoring) which transcend content areas.

Contemporary research in the field of disciplinary literacy demonstrates that fields such as science, mathematics, and history, all have unique expectations for their specific writing genres (Shanahan & Shanahan, 2008; Warren, 2012). Therefore, science teachers cannot rely on the instructions students receive in English class to prepare them to write scientifically. Chapter 3 further supports this argument, as the Rubric for Scientific Writing measures two distinct factors: Scientific rhetoric and English composition. Furthermore, the focus group participants identified the portion of the Rubric for Scientific Writing which focused on the aspects of scientific writing. Therefore, this tool will help science teachers develop their students' fundamental scientific literacy skills by delineating expectations for the field. By engaging in this type of writing, students will analyze and evaluate the knowledge they have gained in science class, and thus become more scientifically literate in the derived sense (Krathwohl, 2002).

Implementing a writing-to-learn intervention can support students' fundamental science literacy, as demonstrated by the findings of Chapter 4. All student clusters improved in their use of Scientific Rhetoric, with some clusters demonstrating effect sizes over 1.0. However, what is most encouraging about these findings is that while the groups demonstrated statistically significant differences at the first writing task, these differences no longer existed at the last writing task. This finding indicates that an achievement gap may be narrowed through writing-to-learn strategies. Future research must examine whether this apparent effect on students fundamental scientific literacy will extend to their derived scientific literacy before generalized claims can be made. However, there is reason to believe that engaging students in the higher-order thought processes of analyzing, evaluating, and creating new knowledge through writing will also develop their derived science literacy.

Measures and Methods for Writing-to-Learn Interventions

As was demonstrated by Miller's (2014) review, writing-to-learn in science class is not a novel idea. However, the findings from existing research have come to a variety of conclusions. The cluster analysis in Chapter 4 indicates that many different variables influence the effectiveness of writing-to-learn interventions, including students' self-efficacy towards writing, and researchers require tools for quantifying their effects.

Additionally, the literacy research field requires a standard measure of scientific writing. Many published studies use prompt- or subject-specific rubrics to evaluate participant writing (e.g., Chrstenson et al., 2012; Hand et al., 2004). While these measures are appropriate for the researchers' goals, it makes meta-analytic comparisons nearly impossible as achievement is measured on different scales. The Rubric for Scientific Writing helps to alleviate this challenge by providing a tool that can be used with different populations and subject areas to quantify good scientific writing.

Finally, methods for implementing writing-to-learn into existing science classes are few and far between. The Science Writing Heuristic has the potential to naturally integrate authentic scientific writing into middle and high school science classes. However, the transformative nature of this approach requires both teacher and administrator commitment. Until policy makers remove external pressures, such as high-stakes testing, most stakeholders will not be able to take such risks and implement innovative approaches such as the Science Writing Heuristic. Therefore, this study developed an effective and feasible intervention strategy that can be implemented in the reality of today's classrooms. As detailed in Table 4.2, both teachers demonstrated a willingness to integrate the writing-to-learn approach into their daily practice. Thus, not only are the student-level results promising, the science teacher participants' actions suggest the buy-in potential for these strategies.

Conclusions

It has been nearly 30 years since James Britton argued that writing should be combined with other literacy skills to support learning (The National Institute of

Education, 1988), yet in science classrooms writing remains largely absent (Drew et al., 2014). While many educators and researchers would agree with Britton that writing can make a unique contribution to learning, how to harness that contribution is still contested. The findings from this dissertation contribute to this ongoing conversation by providing methods for integrating writing into existing science classes, and offering measures for quantifying the effects of these teaching strategies.

Large corporations in the United States spend 3.1 billion dollars annually providing professional development to help their employees become better writers (National Commission on Writing, 2004). As educators, it is imperative that we develop students' writing skills before they exit high school. Addressing this need will help to develop a generation of students who are more prepared to write scientifically and think like professionals in the field.

REFERENCES

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

Akkus, R., Gunel, M., & Hand, B. (2013). 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.

Alev, N. (2010). Perceived values of reading and writing in learning physics in secondary classrooms. *Scientific Research and Essays, 5*(11), 1333–1345.

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

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*, 1-26.

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.

Behizadeh, N. (2014). Mitigating the dangers of a single story: Creating large-scale writing assessments aligned with sociocultural theory. *Educational Researcher, 43*, 125-136. doi: 10.3102/0013189X14529604

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Brackett, M. A., Reyes, M. R., Rivers, S. E., Elbertson, N. A., & Salovey, P. (2012). Assessing teachers' beliefs about social and emotional learning. *Journal of*

Psychoeducational Assessment, 30(3), 219-236. doi: 10.1177/0734282911424879

Cavagnetto, A. R. (2010). Argument to foster scientific literacy: A review of argument interventions in K-12 science contexts. *Review of Educational Research, 80,* 336–371.

Chetty, R. (2012). Great teaching: Measuring its effects on students' future earnings. *Education Next*, 58-64.

Chetty, R. (2014). Measuring the impacts of teachers: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593-2632.

Chinn, C. & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1-49.

Choi, A., Notebaert, A., Diaz, J., & Hand, B. (2010). Examining arguments generated by year 5, 7, and 10 students in science classrooms. *Research in Science Education, 40,* 149–169.

Christenson, N., Rundgren, S.-N. C., & Hoglund, H.-O. (2012). Using the SEE-SEP model to analyze upper secondary students' use of supporting reasons in arguing socioscientific issues. *Journal of Science Education and Technology, 21,* 342–352.

Colby, S. A., & Stapleton, J. N. (2006). Preservice teachers teach writing: Implications for teacher educators. *Reading Research & Instruction, 45*(4), 354–376.

Conradi, K. Jang, B. G., & McKenna, M. C. (2014). Motivation terminology in reading research: A conceptual review. *Educational Psychology Review 26*, 127-164. doi: 10.1007/s10648-013-9245-z

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology, 100,* 907-919.

Dempsey, M. S., PytlikZillig, L. M., & Bruning, R. H. (2009). Helping preservice teachers learn to assess writing: Practice and feedback in a web-based environment. *Assessing Writing, 14,* 38-61.

Drew, S. V., Olinghouse, N., & Faggella-Luby, M. (December, 2014). Writing instruction in science class: A national survey. Presented at the Literacy Research Association Annual Conference, Marco Island, FL.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: The National Academies Press.

Dunn, J. E. (2000). Reading in the content areas: Unlocking the secrets and making them work. *Journal of Adolescent & Adult Literacy, 44*(2), 168–170.

Eccles, J. S., & Wigfield, A. (2002). Motivation, beliefs, values, and goals. *Annual Review of Psychology, 53,* 109-132.

Fang, Z. (2012). Language correlates of disciplinary literacy. *Topics in Language Disorders, 32*(1), 19-34.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior*. Reading, MA: Addison-Wesley Publishing Company.

Fordham, N. W., Wellman, D., & Sandman, A. (2002). Taming the text: Engaging and supporting students in social studies readings. *The Social Studies, 39(*4), 149–158.

Fulmer, S. M., & Frijters, J. C. (2011). Motivation during an excessively challenging reading task: The buffering role in relative topic interest. *The Journal of Experimental Education, 79*, 185–208.

Gillis, V. (2014). Disciplinary literacy: Adapt not adopt. *Journal of Adolescent & Adult Literacy*, *57*(8), 614-623. doi: 10.1002/jaal.301

Glasser, B. G., & Strauss, A. L. (1965). The constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436-445.

Gloggler, I., Holzäpfel, L., Schwonke, R., Nückles, M., & Renkl, A. (2009). Activation of learning strategies in writing learning journals: The specificity of prompts matters. *Zeitschrift für Pädagogische Psychologie, 23*(2), 95-104.

Glogger, I., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology*, *104*(2), 452–468.

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

Graham, S., & Harris, K. R. (2012). The role of strategies, knowledge, will, and skills in a 30 year program of research (with homage to Hayes, Fayol, and Boscolo). In V. Berninger (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 177–196). London: Psychology Press.

Graham, S., Harris, K. R., Fink, B., & MacArthur, C. A. (2001). Teacher efficacy in writing: A construct validation with primary grade teachers. *Scientific Studies of Reading*, *5*(2), 177–202.

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariage applications* (2nd ed.). New York, NY: Routledge.

Gunel, M., Hand, B., & McDermott, M. A. (2009). Writing for different audiences: Effects on high-school students' conceptual understanding of biology. *Learning and Instruction, 19*, 354–367.

Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, *32*, 282-313. doi: 10.1016/j.cedpsych.2006.05.004

Guthrie, J. T., Hoa, L. W., Wigfield, A., Tonks, S. M., & Perencevich, K. C. (2006a). From spark to fire: Can situational reading interest lead to long-term reading motivation?. *Reading Research and Instruction, 45*(2), 91-117.

Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006b). Influences of stimulating tasks on reading motivation and comprehension. *The Journal of Educational Research*, *99*(4), 232-245.

Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology, 92*(2), 331-41.

Haden, C. A., Reese, E., & Fivush, R. (1996). Mothers' extratextual comments during storybook reading: Stylistic differences over time and across texts. *Discourse Processes, 21,* 135-169.

Hair, J. F., & Black, W. C. (2000). *Reading and understanding more multivariate statistics*. Washington, D.C.: American Psychological Association.

Hammett, L. A., van Kleeck, A., & Huberty, C. J. (2003). Patterns of parents' extratextual interactions during book sharing with preschool children: A cluster analysis study. *Reading Research Quarterly, 38*(4), 442-438.

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

Hand, B., Hohenshell, L., & Prain, V. (2004a). Exploring students' responses to conceptual questions when engaged with planned writing experiences: A study with year 10 science students. *Journal of Research in Science Teaching, 41*(2), 186–210.

Hand, B., Wallance, C. W., & Yang, Y. –M. (2004b). 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. doi: 10.1080/0950069032000070252

Harris, K. R., & Graham, S. (2013). “An adjective is a word hanging down from a noun”: Learning to write and students with learning disabilities. *Annals of Dyslexia, 63*, 65-79. doi: 10.1007/s11881-011-0057-x

Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*(2), 75-110. doi: 10.3102/0162373708317689

Hodges, T. S., McTigue, E. M., Weber, N. D., Douglass, A. G., & Wright, K. L. (April 2015*).* Modeling the “write” teaching practices: Instructor influences on preservice teachers. American Educational Research Association, Chicago, IL.

Hohenshell, L. M., & Hand, B. (2006). Writing-to-learn strategies in secondary school cell biology: A mixed methods study. *International Journal of Science Education, 28*(2–3), 261–289.

IEA International Association for the Evaluation of Educational Achievement. (2013). *Trends in International Mathematics and Science Study*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Jagger, S. L. & Yore, L. D., (2012). Mind the gap: Looking for evidence-based practice of *science literacy for all* in science teaching journals. *Journal of Science Teacher Education*, *51*(5), 372-383.

Johnston, P. & Costello, P. (2005). Principles for literacy assessment. *Reading Research Quarterly, 40*, 256-267. doi:10.1598/RRQ.40.2.6

Johnson, R. L., Penny, J., & Gordon, B., (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education, 13*(2), 121-138.

Kear, D. J., Coffman, G. A., McKenna, M. C., & Ambrosio, A. L. (2000). Measuring attitude toward writing: A new tool for teachers. *The Reading Teacher, 54,* 10-23.

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, *11*(3), 203–270.

Klein, P. D., & Rose, M. A. (2010). Teaching argument and explanation to prepare junior students for writing to learn. *Reading Research Quarterly*, *45*(4), 433–461.

Knaggs, C. M., & Schneider, R. M. (2012). Thinking like a scientist: Using vee-maps to understand process and concepts in science. *Research in Science Education, 42*(2), 609–632.

Konopak, B. C., Martin, M. A., & & Martin, S. H. (1987). Reading and writing: Aids to learning in the content areas. *Journal of Reading, 31*(2), 109–115.

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory into Practice*, *41*(4), 212-218. doi: 10/1207/s15430421tip4104_2

Lesley, M. (2014). "Spacecraft reveals recent geological activity on the moon": Exploring the features of NASA Twitter posts and their potential to engage adolescents. *Journal of Adolecsent and Adult Literacy, 57*(5), p. 377-385. doi: 10.1002/jaal.258

Mason, L. H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. *Journal of Educational Psychology, 96*(2), 283-296.

McCrimmon, A. W., & Climie, E. A. (2011). Test review: Test of Written Language - Fourth Edition. *Journal of Psychoeducational Assessment, 29,* 592-596.

McKenna, M. C., & Kear, D. J. (1990). Measuring attitude towards reading: A new tool for teachers. *The Reading Teacher,* 43(9), 626-639.

McTigue, E. M. (2009). Does multimedia learning theory extend to middle-school students? *Contemporary Educational Psychology, 34*(2), 143-153. doi:10.1016/j.cedpsych.2008.12.003.

Merton, R. K., & Kendall, P. L. (1946). The focused interview. *American Journal of Sociology, 51*, 541-557.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation* (2nd ed.). Los Angeles, CA: SAGE publications.

Miller, D. M. (2014). Writing tasks in content-area instruction: A systematic review of the literature. Ph. D. Dissertation. Texas A&M University.

Miller, B., McTigue, E., & Scott, C. (in press, 2015). The quality of recent studies in content-area writing in secondary classrooms. *Literacy Research: Theory, Method, and Practice.*

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.

Mundai, L. (2011). Social desirability non-response bias and reliability in a long self-report measure: Illustrations from the MMPI-2 administered to Brunei student teachers. *Educational Psychology, 31*(2), 207-224. doi: 10.1080/01443410.2010.545049

National Assessment of Educational Progress. (2014). General information about The Nation's Report Card: Sample questions, science. National Center for Education Statistics. Retrieved October 2014 from http://nces.ed.gov/nationsreportcard/subject/field_pubs/sqb/pdf/2104_ sq_s_g4_g12.pdf

National Association for Research in Science Teaching (2015). *About NARST*. Retrieved July 2015 from https://www.narst.org/about/mission.cfm

National Commission on Writing. (2004). *Writing: A ticket to work, or a ticket out.* New York: College Board.

National Governors Association. (2010). *Common core state standards for English language arts & literacy in history/social studies, science, & technical studies.* Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School.

The National Institute of Education. (1988). *Writing and Reading in the Classroom* (J. Britton) (Tech. Rep. No. _425). Washington, DC: Center for the Study of Reading.

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement, 22*(4), 287–293.

Newell, G. E., Beach, R., Smith, J., VanDerHeide, J., Kuhn, D., & Andriessen, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly, 46*(3), 273--304.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states.* Washington, DC: The National Academies Press.

Norris, S., & Phillips, L. M. (2002). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224–240.

O'Brien, D. G., Stewart, R. A., & Moje, E. B. (1995). Why content literacy is difficult to infuse into the secondary school: Complexities of curriculum, pedagogy, and school culture. *Reading Research Quarterly, 30*, 442–463.

Quinn, D. M., & Cooc, N. (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school: Trends and predictors. *Educational Researcher, 44,* 336-346. doi: 10.3102/0013189X15598539

Reinhardt, B. (1996). Factors affecting coefficient alpha: A mini monte carlo study. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 3-20). Greenwich, CT: JAI Press.

Rivard, L. P. (2004). Are language-based activities in science effective for all students, including low achievers? *Science Education, 88*(3), 420–442.

Roid, G. H. (1994). Patterns of writing skills derived from cluster analysis of direct-writing assessments. *Applied Measurement in Education, 7*(2), 159-120.

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

Slough, S. W., McTigue, E. M., Kim, S., & Jennings, S. K. (2010). Science textbooks' use of graphical representation: A descriptive analysis of four sixth grade science texts. *Reading Psychology, 31*, 301-325. doi:10.1080/02702710903256502.

Steinmayer, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19*, 80–90.

Thompson, B. (2003). Understanding reliability and coefficient alpha, really. In B. Thompson, (Ed.). (2003). *Score reliability: Contemporary thinking on reliability issues*. Newbury Park, CA: Sage.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach.* New York: Guilford.

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing, 26*, 17–44.

Viswesvaran, C., & Ones, D. S. (2000). Measurement error in "Big Five factors" personality assessment: Reliability generalizations across studies and measures. *Educational and Psychological Measurement, 60*(2), 224-235. doi: 10.1177/00131640021970475

Vygotsky, L. S. (1980). *Mind in Society: The development of higher psychological processes.* Boston, MA: Harvard University Press.

Warren, J. E. (2012). Rhetorical reading as a gateway to disciplinary literacy. *Journal of Adolescent & Adult Literacy, 56,* 391–399.

Wigfield, A., & Guthrie, J.T. (1995) *Dimensions of children's motivations of reading: An initial study* (Research Rep. No. 34). Athens, GA: National Reading Research Center.

Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator, 109*(4), 21–29.

Wright, K. L., Franks, A. D., Kuo, L.-J., McTigue, E. M., & Serrano, J. (2015). Both theory and practice: Science literacy instruction and theories of reading. *International Journal of Science and Mathematics Education*. doi: 10.1007/s10763-015-9661-2

Wright, K. L., Hodges, T. S., & Franks, A. D. (2015). *Fighting both aliteracy and illiteracy: A meta-analysis of motivation during intervention research*. Manuscript submitted for publication.

Yiğit, E., & Gökpinar, F. (2010). A simulation study on tests for one-way ANOVA under the unequal variance assumption. *Communication Science University of Ankara Series, 59*(20), 15-34.

APPENDIX A

RUBRIC FOR SCIENTIFIC WRITING

	0 (poor)	1 (developing)	2 (approaching expectations)	3 (meets or exceeds expectations)
Claim* (CC #1) * Note: "Claim" can mean purpose, argument, or focus	The student's paper has no claim. ***OR*** The student's claim does not address the topic being explored. ***OR*** The claim cannot be explored, explained, or argued scientifically.	The student presents multiple claims. ***OR*** The claim made is not the central focus of the paper. ***OR*** The claim must be inferred by the reader.	The student gives one central claim, but the claim is underdeveloped/vague/broad. Thus, the paper requires inferences from the reader	The student addresses one claim in the paper. Claims might be explanatory, persuasive, or argumentative in nature but show concentrated focus of the topic of the paper.
Evidence/ Support (CC #7 and #9)	The student does not provide evidence to support the claim. ***OR*** The student's evidence is irrelevant to the claim.	The student provides insufficient evidence/support for their paper. ***OR*** Evidence is not factual or not scientific ***OR*** Evidence must be inferred by reader	The student's evidence is lacking in only a few areas. ***OR*** Some pieces of evidence do not contribute to the paper's claim or have questionable authenticity.	The student provides factual evidence and support from authentic sources (such as class experiments, course material, or professional, scholarly works). The evidence supports the claim made in the paper.
Analysis of Content (#2)	The student's analysis is thoughtless or irrelevant. ***OR*** The student fails to develop, justify, or clarify the claim with evidence and explanation. The student does not provide analyzed information to explain the model or theory the student is constructing	Student demonstrates an understanding of the information, ***however*** the writing is insufficient for the reader to understand the model or theory the student is constructing ***OR*** Student does not make connections between evidence and claim, or connections must be inferred. ***OR*** Rather than reviewing sources as a body of research, the author describes one piece of evidence at a time	The student provides an analysis of the evidence included, ***however*** the model or theory constructed is underdeveloped. ***OR*** The student's analysis is justified, ***however*** there is a disconnect in the student's commentary and explanation to the reader.	The student develops the claim further by synthesizing the evidence and formulating a clear explanation. ***AND*** The information has been analyzed for the reader to provide precision to the model or theory the student is constructing.

Organization (CC #4)	Writing is not organized. Paragraphs do not flow in a logical order and/or are not focused on a single idea.	The ideas are loosely connected; ***however*** the reader must infer connections among the ideas present. ***OR*** Multiple main ideas are found in each paragraph. ***OR*** The text structure is inappropriate to the audience and purpose.	Some organization is present, ***however…*** The entire paper is organized but the individual paragraphs are not. ***OR*** The paragraphs are organized but do not follow a cohesive, logical sequence.	Student supports the claim in an organized manner. Paragraph order is logical and each paragraph focuses on one main idea. The student shows evidence of choosing a text structure (such as problem-solution, compare-contrast, or chronological) that is appropriate to the audience and purpose.
Audience (CC #4)	Student's writing does not reflect an understanding of the audience. Language and style is inappropriate for the purpose. Student does not supply sufficient background information or provides too much extraneous information	Audience can be inferred, but writing may seem to target different audiences throughout paper, ***OR*** Language used is either too simplistic or too technical for the intended audience ***OR*** Student provides either too much or not enough background information.	The student uses style and language that is appropriate for the audience; ***however***, background evidence is used sparingly. The reader must make inferences to connect all of the ideas.	Style and use of language is appropriate for the audience and purpose. Student provides enough background information for the audience to understand while employing appropriate scientific terms and language.
Presentation of Writing (Conventions, Sentence Fluency, Style)	The student's writing contains a number of errors making the ideas difficult to comprehend. The formatting does not match that of scientific scholarly writing.	Student makes many errors, however a well informed reader will still comprehend. Sentences do not vary in structure or flow from one to another	Student's writing shows few errors, and those that exist do not impede reading comprehension. ***OR*** The student has more than 3 errors in grammar, semantics, syntax or punctuation. The writing might also include choppy sentences or a lack of fluency in sentences.	The student's writing shows a mastery of writing conventions (such as grammar, semantics, syntax, and punctuation). Sentences flow appropriately and show variation in structure. The formatting is appropriate for scholarly scientific writing.

APPENDIX B

SAMPLE INTERVENTION SCHEDULE

	Monday	Tuesday	Wednesday	Thursday	Friday
				Pre-intervention assessments	Researcher presentation to students about value of writing in science class (15 minutes)
Week 1	Short minute writing prompt at end of lesson		Short minute writing prompt at end of lesson		Long writing activity
Week 2	Short minute writing prompt at end of lesson		Short minute writing prompt at end of lesson		
Week 3	Short minute writing prompt at end of lesson		Short minute writing prompt at end of lesson		Long writing activity
Week 4	Short minute writing prompt at end of lesson		Short minute writing prompt at end of lesson		
Week 5	Short minute writing prompt at end of lesson		Short minute writing prompt at end of lesson		Long writing activity
Week 6	Post-intervention assessments				

APPENDIX C

OBSERVATIONS AND FIDELITY RECORD

Date: ________ Class/Teacher: ______________Type of writing activity: _____________
Number of students in room________ Number of present and absent participants:____
Time writing started_______ Time writing ended________ Observer________________

Rate the students' level of engagement during the lesson:

Off task/distracted for most of lesson	Off task/distracted for more than half of lesson	Equal parts on task and off task	On task for more than half of lesson	On task for all of lesson
1	2	3	4	5

Rate the students' level of engagement during the writing time:

Off task/distracted for most of writing time	Off task/distracted for more than half of writing time	Equal parts on task and off task	On task for more than half of writing time	On task for all of writing time
1	2	3	4	5

For the following possible observations, write "0" if not observed, "1" if less than half the students were engaged, "2" if about half the students were engaged, "3" if more than half the students were engaged, and "4" if all (or nearly all) students were engaged.

___ Referring to textbook
___ Referring to class materials (not textbook)
___ Asking adult for help with content
___ Asking adult for help with writing
___ Discussing content with peers
___ Discussing writing with peers
___ Use of electronic sources for content
___ Use of electronic sources for writing
___ Using dictionary
___ Using other source for writing support
___ Other: ____
___ Other: ____

Were there any unusual distractions or interruptions during class or writing (e.g. fire drill)? Please specify when the distraction occurred:

____

____

Additional observations during lesson or writing time

Please note if any individual students were particularly engaged or disengaged during the writing time, or any other observations that may indicate a student's overall motivation, self-efficacy, or attitude: