Spring 2014

The effects of the science writing heuristic (SWH) approach versus traditional instruction on yearly critical thinking gain scores in grade 5-8 classrooms

Ching-mei Tseng
University of Iowa

THE EFFECTS OF THE SCIENCE WRITING HEURISTIC (SWH) APPROACH VERSUS TRADITIONAL INSTRUCTION ON YEARLY CRITICAL THINKING GAIN SCORES IN GRADE 5-8 CLASSROOMS

by

Ching-mei Tseng

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Teaching and Learning
in the Graduate College of
The University of Iowa

May 2014

Thesis Supervisors: Professor Brian M. Hand
Professor Walter Vispoel

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Ching-mei Tseng

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Teaching and Learning at the May 2014 graduation.

Thesis Committee:

Brian Hand, Thesis Supervisor

Walter Vispoel, Thesis Supervisor

Cory Forbes

Soonhye Park

Amy Poremba

To My Grandparents and My Parents

ACKNOWLEDGMENTS

First and foremost, I would like to thank to my supervisors of this thesis project, Professor Brian Hand and Professor Walter Vispoel for their precious guidance and advice. They inspired me greatly to work in this project. Their willingness to motivate me contributed tremendously to this project. I also would like to thank my committee members, Professor Cory Forbes, Soonhye Park, and Amy Poremba for providing valuable comments to improve this project and also showing me unconditional support. Besides, I would like to thank the my colleagues and the college of education for providing me with a wholesome environment and facilities to complete this project. Also, I would like to take this opportunity to thank to many people I have known from the Iowa City and Boston for offering this subject, a lifelong project I have learned form all of you to pursue the purpose of life. It gave me an opportunity to participate and learn about teaching and learning. In addition, I would also like to thank Sue Cline at students service and Matt Volkman who provide me valuable information as the guidance and supports of completing my journey. Finally, an honorable mention goes to my families and friends for their understandings and love. You make me enjoy my teaching and learning and will continue doing so.

ABSTRACT

Critical Thinking has been identified in the Common Core State Standards (CCSS) and Next Generation Science Standards (NGSS) as skills needed to prepare students for advanced education and the future workforce. In science education, argument-based inquiry (ABI) has been proposed as one way to improve critical thinking. The purpose of the current study was to examine the possible effects of the Science Writing Heuristic (SWH) approach, an immersion argument-based inquiry approach to learning science, on students' critical thinking skills. Guided by a question-claims-evidence structure, students who participated in SWH approach were required to negotiate meaning and construct arguments using writing as a tool throughout the scientific investigation process. Students in the control groups learned science in traditional classroom settings. Data from five data sets that included 4417 students were analyzed cross-sectionally and longitudinally. Yearly critical thinking gain scores, as measured by Form X of Cornell Critical Thinking Test, were compared for students who experienced the SWH approach versus students who experienced traditional instruction in both elementary (5^{th} grade) and secondary schools (6^{th}-8^{th} grades). Analyses of yearly gain scores for data sets that represented a single year of implementation yielded statistically significant differences favoring SWH over traditional instruction in all instances and statistically significant interactions between gender and grade level in most instances. The interactions revealed that females had higher gain scores than males at lower grade levels but the reverse was true at higher grade levels. Analyses from data sets that included two years of implementation revealed higher overall gains for SWH instruction than for traditional instruction but most of those gains were achieved during the first year of implementation. Implications of these results for teaching critical thinking skills in science classrooms are discussed in detail.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

GENERAL OVERVIEW AND PURPOSE OF STUDY

Introduction

Learning science is about engaging students in the reasoning processes, rather than merely memorizing concepts and doing experiments following teachers' instructions. This inquiry process no longer just encompasses absorbing facts and concepts, but also negotiating and constructing knowledge through scientific ways of thinking, such as through generating their own questions, conducting scientific experiments, interpreting data and constructing sound evidence (Cavagneto & Hand, 2011; Hand, 2008, McNeill, 2008). The more opportunities students are given to immerse in this argumentative inquiry process, the more likely they are able to think scientifically. Scientific ways of thinking, such as hypothesizing, predicting, reasoning, and negotiating ideas, are essential for students to learn science and to develop critical thinking skills (Wilson et al, 2010). In turn, critical thinking skills provide students with tangible personal, academic, and professional benefits of successful problem-solving in both inside and outside of classrooms (Quitadamo et al, 2009).

Following a constructivist perspective, both of national and international science organizations have recently urged teachers to engage students in discovery, reflection and active learning, which in turn is expected to improve thinking skills that support scientific literacy. *Argument*, in particular, is viewed as a critical element of science instruction to enhance scientific literacy. Students utilize argument as the fundamental component to construct conceptual understanding (Cavagneto & Hand, 2011; Hand, 2008). In addition, *argumentation* (understood as the process of *argument*) is not only a central element but also a critical process and standard for proficient students in mathematics and language art, addressed in Common Core State Standards (CCSS). The Next Generation Science Standards (NGSS), the science educational counterpart of CCSS, also provides consistent, practical and researched-based standards to engage students in science instruction and prepare them to apply "critical thinking and creative problem-solving necessary to excel in the global society" (NRC, 2013). In general, a variety of national stakeholders, including parents, educators, schools, and public agencies, have advocated a long-term transformation of the K–12 educational system to prepare students better in science literacy, engage them actively and intelligently in global issues, and facilitate their problem solving and critical thinking skills.

Critical thinking has long been considered as one of the most important indicators of learning outcomes. Of special concern is the extent to which students have mastered critical thinking (or higher-order thinking skills), successfully enacted problem-solving, and applied such skills to the real-world situations (Bransford & Schwartz, 1999; Reece, 2002; Wilson et al, 2010). In its 2005 annual national report, the American Association of College and University (AACU) noted that analytical and critical thinking were seen to as an essential learning outcome by 93% of higher education faculty. On the other hand, 87% of undergraduate students indicated that education experiences, prior to college years, contributed to their ability to think analytically and critically (Geier et al, 2008; Quitadamo et al, 2009). However, only 6% of undergraduate seniors actually demonstrated critical thinking proficiency, measured by the Educational Testing Services (ETS) critical thinking assessment, reported from 2003 to 2004, and from 2006 to 2008. During these time frames, results from the American College Testing (ACT) Collegiate Assessment of Academic Proficiency test revealed a similar trend, with students improving their critical thinking less than one standard deviation from freshman to senior year. The above reports indicate a discrepancy between both educator expectations and students' perceptions of critical thinking and their abilities to demonstrate critical thinking proficiency using standardized assessments (Geier et al, 2008). Despite the collective call for enhanced critical thinking and problem-solving skills, instructional methods that measurably improve critical thinking skills have not been clearly identified

(Fisher, 2001; Quitadamo et al, 2009;Tsui, 2002). With the goal to improve K–12 student performance, research-supported education practices are essential to help students develop and master cognitive skills (Chiapetta, 2008; Zembal-Saul, 2009). In the recent years, developing corresponding pedagogy that supports the students' critical thinking skills has become a central focus of major educational initiative.

Since the 1980s, the science education reform movement has advocated that all Americans be not only scientifically literate but also active in scientific exploration (see, e.g. American Association for the Advancement of Science (AAAS) 1993; Minner et al, 2010; National Research Council (NRC), 1996, 2000, 2013). For example, the U.S. Department of Education released six "National Goals" for schools. Specifically, Goal Three specified that students be "able to reason, solve problems, and apply knowledge", and Goal Six specified that students be able to think critically. In addition, the National Science Education Standards (NSES, NRC, 1996) call for students to "do inquiry" in the process and arrive at the learning results of "know[ing] about inquiry". The NGSS (NRC, 2013) stated that inquiry-based learning not only helps student understand the reason for and importance of learning science content, but also enables them to develop scientific thinking skills across subjects through scientific exploration.

While the academic and personal learning benefits of critical thinking might not always be transparent to students and teachers, they are well established in the research literature. Students who are more likely to utilize reasoning in daily decisions and problem solving, are able to think critically perform better academically, and transfer what they learn to new learning situations (Bransford & Schwartz, 1999). The instructional efforts put forth to develop critical thinking skills can significantly increase student performance. This paradigm shift satisfies national calls for educational improvement by increasing students' abilities to solve problems and become engaged and productive citizens.

More recent reports further emphasize the need for improved science literacy as well as international competitiveness (Berland & Reiser, 2009; Bybee et al, 2006; Chiapetta, 2008; Duschl, 2008; Quitadamo et al, 2009). Students participating in learner-centered inquiry curricula not only develop scientific knowledge and skills through the inquiry activities, but also act like scientists to construct and interpret findings under authentic scientific circumstances (Driver et al, 2000). This concept highlights learning activities that students carry on from the time they engage in the essence of exploration and throughout the course of knowledge construction. In the real world, the results or answers to the scientific questions often cannot be known in advance. Therefore, the inquiry process involves not only exploration and discovery, but also invention and critique that underscores the nature of scientific exploration (Ford & Forman, 2006).

Current perspectives in science education emphasize the importance of Argument-based Inquiry (ABI) as a means to improve student science achievement (Cavagnetto, 2010; Ford & Forman, 2006). The operational definition of argumentation is aligned with the instructional meaning of critical thinking and the purpose of science education, which is to enable individuals to be able to define problems, generate research questions, conduct experiments based on claims proposed, negotiate claims based on evidence, and finally conclude and apply the information with necessary thinking skills (Berland & McNeill, 2010; Cavagnetto & Hand, 2011; Hand, 2012).

In ABI instruction, the teacher asks students to generate questions, produce inferences, and debate their findings. Argument-based instruction is a student-centered and teacher-guided instructional approach. In contrast to teacher-dominated instruction, ABI places students at the forefront of the learning process with teachers playing the role of coach, facilitator and modeler. ABI complements traditional instruction by providing a vehicle that engages students in investigating real world questions and to extend and applying their learning that connects with interests (Abbott & Wilks, 2005; Cavagnetto & Hand, 2011; Hand, 2012).

The ABI instructional approach adopted in the current study is called the Science Writing Heuristic (SWH) approach (Hand & Keys, 1997). This approach is grounded in a question-claims-evidence structure. Students are required to conduct inquiry investigations by posing their own questions about the concept under review, collecting data, constructing claims based on evidence, finding out what experts say, and reflecting upon their arguments to examine how their ideas have changed. Throughout the process, students are required to negotiate meaning utilizing a variety of writing forms, such as lab notes, reflection and summary writing.

Language is considered a means of promoting student scientific literacy that is bolstered through the use of writing (Keys et al, 1999; Hand, 2008; Hand & Prain, 2002; Rivard, 1994). Writing is also has been found to contribute to the development of critical thinking skills (Becker, 2006; Taylor & Sobota, 1998;). Galbraith et al (2005) suggested that writing improves thinking because it requires individuals to make their ideas explicit by activating a series of cognitive processes. The negotiation and reorganization of knowledge schemes (nodes) enable students to evaluate and choose among tools necessary for effective discourse, where they identify issues, formulate hypotheses and arguments. Writing provides an opportunity to think through arguments that serves as a cultivator and an enabler of higher-order thinking (Halliday & Martin, 1994). The act of writing requires students to make critical choices and argument through focusing and clarifying their thoughts, thereby taking them through utilizing higher-order thinking skills to respond to complex problems (Knudson, 1992; Prain & Hand, 1996; Quitadamo et al, 2009). As a result, writing converts students from passive to active learners. Writing can benefits students by restructuring their knowledge and transferring their knowledge across domains (Boscolo & Mason, 2001).

The SWH not only provides argument-based instructional strategies an instructor could employ, but also engages students in the scientific ways of thinking and argumentation. The argumentative process of SWH scaffolds students to build scientific explanations through reasoning, which helps them apply reasoning and reflective thinking to decide what to judge and act on (Ennis, 2005). Research on the SWH approach using standardized science tests, critical thinking tests, and writing tasks have supported its benefits over traditional instruction (Cavagnetto, 2010).

Calls for improving the critical thinking ability of students in both the science education and the general education field represent a consensus to align critical thinking with learning theory, epistemological orientations of science, and pedagogical practices. In this regard, immersion-based argument inquiry approaches, such as the SWH approach, are considered promising teaching practices that merit further exploration.

Purpose of the Study

The primary purpose of the current study was to examine the effect of the SWH approach on students critical thinking skills based on five data sets taken from science classrooms. Results for SWH versus traditional instruction were examined cross-sectionally and longitudinally using Cornell Critical Thinking Test, form X (CCTT-X) scores as the outcome measure. Students were sampled from grades five through eight with scores examined further with regard to gender, grade level, and year of SWH implementation.

Research Questions

Three research questions were addressed:

Research Question 1: To what extent do critical skills change from the beginning to the end of each academic year? (Data sets 1-5)

Research question 2: What are the main and interactive effects of instructional approach, gender, and grade level on CCTT-X gain scores across a single academic year? (Data sets 1-3)

Research Question 3: What are the main and interactive effects of year of implementation, instructional approach (SWH versus traditional), and gender on CCTT-X gain scores across two academic years? (Data sets 4-5)

Overview for Subsequent Chapters

In Chapter two, results of previous research are detailed to provide the conceptual framework for the study. Topics include the nature of Critical Thinking Argument-based Inquiry in which the study is grounded, the role of Critical Thinking in science education, the Science Writing Heuristic (SWH) as a pedagogical tool, and measurement issues and pedagogical strategies for effective instruction related to critical thinking.

In Chapter three, the rationale for the research design and analytical strategies are presented. The context of the study is described as well as the nature of participating schools, students and teachers; data collection instruments; procedures; and analyses.

In Chapter four, the results for the statistical analyses are discussed. Results are organized under each of the research question and summarized at the end of the chapter.

In Chapter five, possible explanations for the research findings are discussed in detail. Implications of the findings for critical thinking teaching instruction are then posed followed by suggestions for future research and caveats about the limitations of the research.

CHAPTER TWO
LITERATURE REVIEW

The purpose of this chapter is to establish the theoretical framework for the research questions central to this study. This chapter begins with an overview of the goal of recent educational reform that engages students in learning science as inquiry. During this scientific inquiry process, students not only engage in argumentative discourse, but also cultivate high-order skills, such as critical thinking. The researcher further provides a rationale of the current study in three parts. First, the importance of critical thinking and argumentation in learning in science classes is examined. This is followed by a discussion of measurement, empirical studies and sources related to the effects of students' critical thinking and argumentation in science classrooms. This discussion focuses on students' learning, as well as teachers' instructions drawn from historical and social parallels between educational movements and empirical research. Finally, the concept of critical thinking within educational contexts is examined in relation to teaching and transfer of skills. Information provided in this section will lend support for the use of the Science Writing Heuristic (SWH) approach, which was the focus of this study.

The Importance of Critical Thinking in Science Learning

In past decades, a number of national reports indicated a concern about the effectiveness of science education and teaching practices, particularly in responding to U.S students' relatively low scores on international standardized tests, and math and science competitions. Compared to other industrialized countries, the performance of the U.S. students in science and math is decreasing (Geier et al, 2008). Following constructivist perspectives, students are now expected to actively engage in laboratory activities, including making observations, finding problems, and generating questions (Hand et al, 2008; McNeill et al, 2006). After posing claims from data they collect, students get further involved in comparing, justifying and communicating explanations from the results with their peers and teachers (Cavagnetto & Hand, 2010). Learning science, consequently, involves reasoning processes of justifying experiment findings and constructing meaningful knowledge and skills, rather than simply doing experiments and memorizing concepts. During this process, students advance their learning experiences from fact-focused exploration and experimentation to evidence-based explanation and argumentation. As the result, students become more scientifically literate, more effective at problem-solving and better critical thinkers (Mangiante, 2013).

Beginning in the 1980s, several new national standards movements for the science education curriculum were launched from organizations including the American Association for the Advancement of Science (AAAS), Science for All Americans (1990), Benchmark for Science Literacy (AAAS, 1993), the National Science Education Standards ((NSES), National Research Council (NRC), 1996), a Framework for K-12 Science Education (NRC, 2012), and most recently--the Next Generation Science Standards ((NGSS), NRC, 2013). These initiatives have all advocated that science instruction focus on inquiry-based learning that requires students to get involved in communication and argumentation, and actively participate in their learning (Minner et al, 2010; Norris et al, 2008).

An important goal of "Science for All Americans" is to prepare all Americans to be not only scientific literate, but also active in scientific exploration (AAAS, 1993). Through this process, students shift from "knowing about inquiry" to "doing inquiry" (NSES, NRC, 2000). Consequently, learning shifts away from a traditional knowledge-transmission environment to a knowledge-construction one. That is, these standards emphasize a new pedagogy and methodology that requires students to develop skills of argument such as making claims, using evidence, and evaluating claims based on evidence with other students (Hand, 2012; Jimenez-Aleixandre & Erduran, 2008; Newton et al, 1999). Since the advent of the Framework for K-12 Science Education and NGSS

(NRC, 2013), contemporary conceptions of learning science encompass foundations of scientific literacy, inquiry, argumentation, and the nature of Science.

Literacy is a means of advancing students' conceptual understanding, and critical reasoning skills, and metacognitive processes. With the focus of CCSS and NGSS on science, technology, engineering, and mathematics (STEM), science teachers and educators are encouraged to use communication and argumentation instructions to develop students' competencies, and ultimately promote students' scientific literacy. Through both inquiry-based and argument-based inquiry instruction, students are also expected to become literate in global society. By engaging in argument and argumentative discourse, students are developing reasoning and thinking ability in determining the best explanations and decisions about issues that intrigue them both in the classroom and the real world (NGSS, 2013). Improving students' scientific literacy and thinking skills is thus a desired goal of education to prepare them with knowledge to contribute to the competitive workforce of the 21st Century (Kuhn, 2010; Ku, 2009).

Critical Thinking has also been considered one of the most important "generic abilities" or "core concepts" in many goals of education (Knight, 2006). Furthermore, the CCSS and NGSS (NRC, 2013) both highlight the importance of communicative capacity and critical reasoning as cross-cutting skills essential for post-secondary and the workplace. Being able to understand sources and evaluate information is central to constructing knowledge. Developed capabilities for argumentation and critique often lead to further exploration and experiments that can result in extending and refining proposed models, explanations, or designs (Duschl & Osborne, 2002). These activities involves a bundle of skills, such as analyzing arguments, making inferences (using inductive or deductive reasoning), critiquing, judging or evaluating, and making decisions or solving problems, that are commonly recognized as critical thinking (Ennis, 1985, 1995; Lai, 2011; McNeill, 2011; Siegel, 1992). Specifically, the National Science Educational Standards (NRC, 2000) state: "Think crucially and logically to make the relationships between evidence and explanation ” (NRC, 2000. p. 58), and “Scientific explanations emphasize evidence, have logically consistent arguments, and use scientific principles, models and theories (NRC, 2013, p.66)”. Like the Common Core standards, their counterparts in English language arts and math mirror NGSS’s shift on focus to the merging of argumentation and other practices, such as explanation, critical reasoning and higher-order thinking. Through argumentative discourse, students are better positioned to be constructors, evaluators and critics of knowledge (Jimenez-Aleixandre & Erduran, 2008; McNeill & Pimentel, 2010). Students are also encouraged to “write arguments to support claims using valid reasoning and relevant and sufficiency evidence. That is, students are encouraged to engage in argumentation in “interpreting data, using mathematics and computational thinking, engaging in argument from evidence-based curriculum” (NRC, 2013,p.42), as well as in developing “ reasoning to support or refute an explanation or a solutions with respect to how well they meet the evidence” (NRC, 2013, p.62). NGSS also suggests, “classroom instruction must include critical skills” (p.392), so that students are able to “critically read scientific literature...”(p.411) and apprehend “complex ideas and information by critically choosing data (p.500).”

In this regard, when students think critically, they are better at determining sound claims and evidence and construct decisions and arguments open to debate (Abbott & Ailks, 2005; Bailin, 2002; Sigel, 1992). Learning to handle argumentative discourse is an essential critical thinking skill across different age levels and subject areas that can extend students’ learning from the particular content to a general domain of knowledge learned (Kuhn, 1993; Yeh, 2001). Critical thinking from this perspective is related to developing the capacity of argumentative discourses, which then contributes to the adjustment and production of new learning contexts.

Definitions of Critical Thinking

In this section, literature on the definition of critical thinking is reviewed from different theoretical orientations and academic disciplines including both thinking as abilities/skills or dispositions, and as domain-specific or domain-general.

Similar to many other constructs, definitions of critical thinking vary without a clear consensus. Despite the widespread recognition of its importance, critical thinking has multiple and diverse definitions depending on theoretical orientations that include philosophical, cognitive psychological, and educational perspectives (Lai, 2011). The philosophical perspective focuses on hypothetical qualities and characteristics of the individual, rather than his/her actual behaviors to perform what he/she think (Bailin et al., 1999; Facione, 1990). The cognitive psychological perspective focuses on how people think or how they could think under hypothetical ideal conditions, and encompasses behaviors actually demonstrated during thinking (Halpern, 1998; Sternberg, 1986; Willingham, 2007). The educational perspective stresses instructional applications of critical thinking in classroom practices as well as the real-world interactions (Kennedy et al, 1991). For example, the highest levels of Bloom's hierarchical taxonomy, analysis, synthesis, and evaluation, are commonly used to define what critical thinking is and how it functions in learning environments.

Other definitions of critical thinking have been offered by researchers, such as Paul (1993), Ennis (1985) and Sternberg (1986). Paul (1993) defined critical thinking as a "unique kind of purposeful thinking in which the thinker systematically and habitually imposes criteria and intellectual standards upon the thinking; taking charge of the construction of thinking; guiding the construction of thinking according to the standards; and assessing the effectiveness of the thinking according to the purpose, the criteria, and the standards" (p. 25). According to Ennis, critical thinking is "reasonable reflective thinking focused on deciding what to believe in or do, and provides a set of criteria for assessing it" (Ennis, 1995, p. 45). Sternberg (1986), drawing from both cognitive psychology and educational approaches, put forth a list of purposeful and meaningful aspects of critical thinking.

Other assertions also have been made about the nature of critical thinking. If understood as searching for evidence, critical thinking would be closely related to developing rational criteria, a position also maintained by cognitive psychologists (Lai, 2011). For example, Kuhn explains the development of scientific reasoning as coordination of theory and evidence (Kuhn, 1999, 2000; Kuhn & Dean, 2004). Referring to an educational ideal, Siegel (1992) emphasizes assessment components of critical thinking and the disposition of critical thinkers to seek evidence for their beliefs. While he sees rationality of science as being grounded in a commitment to evidence, Siegel (1992) conceives of critical thinking as the "educational cognate of rationality, involving consistency, impartiality and fairness." (p. 22) Finally, Facione (1990) advocates an educational approach in which critical thinking can be taught within domain-specific areas, or from "events in everyday life" (p.10). The benefit of the educational approach is that it is based on years of classroom experience and observations of students learning environments.

Table 1 includes a summary of definitions of Critical Thinking from different points of view cited here (Adapted from Ennis, 1995, and Lai, 2011)

Table 1. Recap of Critical Thinking (Adapted from Ennis, 1985 and Lai, 2011)

Approach	Definition of "Critical Thinking"
Philosophical Perspective	• "the skill and propensity to engage in an activity with reflective skepticism" (McPeck, 1981, p.8); "the analysis of good reasons for belief, understanding the various kinds of reason involves understanding complex meanings of field-dependent concepts and evidence" (McPeck, 1981, p.24); • "disciplined thinking that is clear, rational, open-minded, and informed by evidence" (Halpern, 1998) • "aimed at forming a judgment," where the thinking itself meets standards of adequacy and accuracy (Bailin et al, 1999, p.287)
Cognitive Psychological Perspective	• "the mental processes, strategies, and representations people use to solve problem, make decisions, and learn new concepts" (Sternberg, 1986, p.3) • "reasonable reflective thinking focused on deciding what to believe in or do, and provides a set of criteria for assessing it " (Ennis, 2002) • "the use of those cognitive skills or strategies that increase the probability of a desirable outcome" (Halpern, 1998, p.450) • "seeing both sides of an issue, being open to new evidence that disconfirms your ideas, reasoning dispassionately, demanding that claims be backed by evidence, deducing and inferring conclusions from available facts, solving problem, and so forth" (Willingham, 2007, p.8)
Educational Perspective	• "the ways of analyzing, synthesizing and evaluating information during the learning process" (Anderson et al, 2001, p. 24) • "the learning process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and evaluating information to reach an answer or conclusion" (Kennedy et al., 1991, p. 13). • "Critical thinking is not *hard* thinking nor is it directed at solving problems (other than improving one's own thinking). Critical thinking is inward-directed with the intent of maximizing the rationality of the thinker." (Glaser, 1941) • "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or conceptual considerations upon which that judgment is based" (Facione, 1990, p.3)

Dispositions or Skills

Despite differences among definitions and theoretical approaches, critical thinking is generally viewed as having both cognitive/dispositional and abilities/ skills components (Facione, 1990, Lai, 2011). Critical thinking dispositions refer to open-mindedness, inquisitiveness, a propensity to seek reason, a desire to solve problem, consideration of other people's perspectives, and a flexibility for and willingness to process diverse viewpoints (Dam & Volman, 2004; McPeck, 1990; Paul, 1992, 1995). To successfully exemplify these skills, a learner needs to possess appropriate dispositions (Reece, 2002).

Compared to the "disposition perspective," proponents of the "ability/skills perspective" stress abilities (activities and behaviors) of the critical thinkers more than their thoughts. Critical thinking abilities most commonly proposed include components of analyzing data, proposing arguments, judging credibility of sources, making inferences using inductive or deductive reasoning to solve problems and make decisions (Ennis, 1985; Norris & Ennis, 1989).

Although the ability to think critically is distinct from possessing the disposition to enact the behaviors, critical thinkers typically possess both cognitive dispositions and skills (Ennis, 1985). Demonstrations of the critical thinking ability are to some extent building on the idea of dispositions. For example, Paul (1995) asserted that critical thinking is a disposition as well as a set of micro-logical skills. McPeck (1990) further stated that critical thinking is the ability to suspend decision and continue to seek for clarification, until there is sufficient evidence to establish a proposition or an action that is valid (Ennis, 1985).

Accordingly, critical thinkers are able to determine and maintain focus, seek and offer reasons, remain open to situations, look for alternatives and to strive to clarify and challenge questions in light of insufficient evidence and reasoning (Kennedy et al, 1991).

In essence, true critical thought is exhibited only by individuals who possess both the ability and the disposition to think critically.

Domain-Specific or Domain-General

There are both domain-specific and domain-general aspects of critical thinking. Critical thinking skills can be generalized across different contexts and domains and thus be taught in a generic way. However, other researchers argue that general critical thinking skills that transcend subjects do not exist; critical thinking skills can only be taught within the context of a specific domain (Lai, 2011).

McPeck (1981, p. 24) defined critical thinking in part as "the analysis of good reasons for belief." While acknowledging that a limited number of general thinking skills exist, he argued that different subjects or domains have different epistemologies that alter the meaning of critical thinking from subject to subject. More specifically, critical thinking involves thinking or solving events mostly within a specific context so that a learner can focus on particular resourceful conditions, instead of overly diluting their thinking power. McPeck (1990) also noted that domain-specific thinking skills are the most useful ones. The more general the thinking skills, the less applicable they are. Similarly, Bailin (2002) argued that when the concept of critical thinking is common and general, it is so generic in criteria and practice that it is not useful. Bailin (2002) adds that domain-specific knowledge is fundamental because what constitutes valid evidence, arguments, and standards vary across domains. The way people learn is rooted in various kinds of content knowledge and understanding that are necessary for successful completion of a particular task. Therefore, general instruction in critical thinking skills is unlikely to be successful since critical thinking skills are inherently domain-specific in that it is easier to learn to think critically within a specific domain (Willingham (2007). In addition, criteria vary across disciplines. For example, deductive proof is the gold standard for reasoning in math, whereas in the social sciences, statistical inference is more prevalent. When it comes to art, subjectivity often plays a greater role than in these other areas (Ennis, 1989).

However, Lipman (1988) argued that while criteria may differ across domains, the fundamental meaning of critical thinking remains the same. Moreover, if critical thinking is domain-specific, the transfer of critical thinking skills across domains would be almost impossible, unless learners are explicitly taught to transfer, and provided with adequate opportunities to practice these skills in a variety of domains (Ennis, 1989). It is worthwhile to note that, while these debates remain, a majority of existing instruments used to access critical thinking are designed as general tests of critical thinking rather than ones embedded within the context of a specific domain. Through evaluating and comparing the general and the infusion approaches to teaching critical thinking, it becomes apparent whether critical thinking skills are domain-specific or general. In turn, the clarifications of features of domain-specific or general inform the design of instruments to assess those skills (Wilen, 1995).

Relationships to Other Concepts

In the process of defining the concept of critical thinking, researchers have drawn connections with other concepts, including reasoning, metacognition, motivation, and creativity. In the context of the current study, argumentation and metacognition are primarily stressed.

Argumentation

Scientific practices defined in the Next Generation Science Standards (NGSS) include communication and argumentation as necessary skills not only for arriving at students' learning outcomes during K-12 education, but also for successful postsecondary accomplishments. Ultimately, students are expected to prepare to be competent citizens in a world fueled with innovations in science and technology (NRC, 2003). In practice, the concepts of critical thinking and argumentation are inextricably linked and developed in parallel. Both ought to be integrated into learning and emphasized in science classrooms. Research on learning in science presents argumentation as a fundamental aspect of the discipline in constructing knowledge (Newton et al, 1999; Kuhn, 1993; NRC, 2013). The construction of new knowledge and theory occurs through argumentation in which learners debate and justify claims using evidence (Driver et al, 2000; Kuhn, 2010). Teaching science, thus, requires engaging students in disciplined focus to construct reliable claims that inform explanations and sound decision making (Ford & Forman, 2006; McNeill et al, 2006; McNeill, 2011). Critical thinking, also described as "critique", is an essential component of argumentation and an important practice in scientific discourse (Kuhn, 1999). However, traditional classrooms often prioritize the final product instead of providing students with sufficient opportunities to reflect and communicate about their learning (Sampson & Clark, 2008).

Argumentation represents an attempt to establish valid claims that are expected to be supported by data, warrants, backings, or be adjusted by qualifiers and rebuttal (e.g., Berland & Reiser, 2009; Drduran et al, 2004). This process can be interpreted in terms of individual-structural or social-dialogic meanings (Jiménez-Aleixandre & Erduran, 2008; McNeill, 2011). Many science educators adapt Toulmin's (1958) model of argumentation where a claim or explanation is justified using various supports such as evidence, logic, warrants, and reasoning. Within the dialogic perspective, argumentation stresses the interaction and communication between learners in which they attempt to negotiate or convince each other of the validity of claims. (e.g., Cavagnetto, 2010; Erduran et al, 2004; Sampson & Clark, 2008). Argumentation, as a learning tool, supports students learning and understanding in developing different ways of thinking and reasoning (Kuhn, 1993; Norris et al, 2008). Argumentative discourse helps students to articulate explanations, elucidate their thinking, and thereby, provide a valuable vehicle for advanced reflection and assessment (Berland and Reiser, 2009; Geier, 2010, NRC, 2008).

This approach affords students opportunities to learn not only science content but also ways of negotiating science, which ultimately enhance understanding of language and promote literacy (Driver et al., 2000; Norton-Meier,Hand, Hockenberry & Wise, 2008).

Critical thinking and argumentation are associated in a variety of ways:
For example, Erduran et al. (2006) proposed five contributions of argumentation and corresponding perspectives to frame these contributions. They noted that argumentation should improve the development of communicative competence and critical thinking in particular. This perspective draws from the theory of communicative action and the socio-cultural influences. Other contributions are to support the development of reasoning, particularly the choice of theories or positions based on rational criteria, the philosophy of science, and developmental psychology. Consequently, engaging in argumentation encompasses the construction, critique and use of questions and the evaluation of claim and evidence for multiple explanations.

The role of argument-based inquiry (ABI) and best ways to support it in K-12 classrooms has been the focus of many recent studies (Cavagnetto & Hand, 2011). A variety of classroom practices derived from this research touch upon the use of curriculum materials (Berland & Reiser, 2009; McNeill et al., 2006), teaching strategies (McNeill, 2009; Berland & McNeill, 2010), and technology tools (Clark & Sampson). In his review, Cavagenetto (2010) identified three main types of interventions differentiated by their orientations towards designing learning environments: (1) immersion in practice; (2) explicit instruction in argument; and (3) understanding of socio-scientific issues. With the onset of the national standards, these approaches are expected to be implemented more frequently and scrutinized for possible benefits and drawbacks. The present study was focused on the Scientific Writing Heuristic (SWH, Hand & Keys, 1997)—an immersion argument-based inquiry approach aligned with the premises and practice of NGSS and students' scientific explorations informed by argumentation. Students identify big questions through individual or group concept mapping and follow them up with pre-laboratory activities that include informal writing, making observations, posing questions and conducting experiments. After conducting an experiment, students share and compare claims and evidence through argumentation and negotiate appropriate concluding statements in small or large groups. Through this process students think critically in generating and evaluating hypotheses.

Metacognition

Metacognition can be simply defined as "thinking about thinking," which is different from cognitive reflection alone. Flavell (1979) believed that critical thinking plays an essential role in metacognition in the sense that wise and thoughtful life decisions follow from "critical appraisal of message source, quality of appeal, and probable consequences needed to cope with these inputs sensibly" (p. 910). The cognitive and metacognitive strategies used for task analysis, decision-making, and problem solving involve a mixture of executive functioning skills and metacognitive ability (Paul & Elder, 2002; Schraw et al, 2006). When categorizing people's learning styles, Perkins (1983) described advanced learning as a process of reasoning, evaluation or self-reflective presentation of arguments (critical reasoning), and methodological reflection. In addition, Finocchiaro (2005) suggested that learning is a series of acts in critiquing and reasoning that involve students in making information relevant and meaningful by linking prior knowledge and new knowledge in a meaningful format.

Halonen (1995; also see Halpern,1998) identified metacognition as an ability to monitor the quality of critical thinking in decision making, analyzing data and synthesizing evidence. Wilen (1995) emphasized the role of metacognition in promoting critical thinking in classroom. He endorsed an infusion approach in which teachers model critical thinking skills within a specific subject matter through think-aloud strategies.

In keeping with Paul's (1992) definition of critical thinking, as "disciplined, self-directed thinking that exemplifies the perfections of thinking appropriate to a particular mode or domain of thought (p.12)," self-regulation could be viewed as a phenomenon that connects critical thinking and metacognition. Facione (1990) viewed self-regulation as one component skill of critical thinking, whereas Schraw et al. (2006) compressed it into three components: cognition, metacognition, and motivation, the first of which includes critical thinking. In this regard, metacognition is an active monitoring state to regulate individual's cognitive processes and an application of a set of heuristics to help learners organize their methods and control their responses during learning activities (Martinez, 2006).

Some researchers, such as McPeck (1990), disagreed with the above links between critical thinking and metacognition. He claimed that the ability to recognize whether a particular skill is relevant and to implement which on a task is not properly exemplifying critical thinking but rather general intelligence. Nevertheless, metacognition can at least be seen as a supporting condition for critical thinking because monitoring the quality and executive function of the thinking process leads to deeper and longer engagement in high-quality thinking (Lai, 2011; Sternberg, 1986; Perkins, 1988).

Measurement of Critical Thinking

In this section, premises and challenges in assessing critical thinking are reviewed, and specific recommendations from the literature for measuring critical thinking are made. After general ideas of assessing critical thinking are introduced, three widely-used instruments are documented as examples.

Assessing Critical Thinking

In addition to understanding how critical thinking is defined, it is also important to understand how to operationalize and measure it. One challenge is to define constructs in ways that they can be observed and interpreted (Baron, 1987; Knight, 2005; Reece, 2002). When it comes to teaching and assessing higher-order thinking skills, Bloom's hierarchical taxonomy of learning has served as one of the most widely cited sources for educational practitioners (Anderson et al, 2001; Ennis, 1987, 1995). This framework also has also been used to design critical thinking inventories. Researchers also have developed detailed rubrics to measure learners' conceptions and use of evidence and their role to generating alternative perspectives (Berland & Mcneill, 2010; Sampson & Clark, 2008). Most measures of critical thinking are focused on aspects of skill rather than dispositions, attitude, habits-of-mind or the virtue of thoughtfulness (Ennis, 1985; Facione, 2000). Ennis's (2009) annotated list of critical thinking tests includes three types of measures: (a) Multiple-Aspects Critical Thinking Tests that cover more than one aspect of critical thinking (and thus are comprehensive to some degree), (b) General Critical Thinking Tests that cover only one aspect of critical thinking, and (c) Subject-Specific Critical Thinking Tests.

Psychologists and educators have questioned and debated the validity of results from these instruments based on how they are structured, developed and administered (Norris, 1989). For example, some researchers raise questions about whether paper-and-pencil tests reflect actual reasoning performance in the real world. In addition, discrepancies exist between what teachers are attempting to teach (abstract thinking) and what a test may measure (skill at deductive reasoning; Lai, 2011). For example, Facione (1990) noted that several instruments that aim to measure abstract thinking or higher-order thinking actually consist of well-structured problems that are better designed to check for factual memorization rather than the nature of executive functioning.

A better approach would be to use open-ended tasks, real-world or authentic problem contexts, and performance-based problems that require students to go beyond recalling or restating previously learned information (Ennis, 1985; Fisher, 1996; Lai, 2011). These assessments allow for more than one defensible solution and require the activation of logical reasoning and executive functioning. When working on these tests, students must provide logical evidence and sound arguments in support of the claims and assertions they make.

Another challenge to assessment of critical thinking is employing a sound research design for valid interpretation of results. For example, the sample used may not be representative of the target population (e.g., voluntary and self-selection) and /or subjects may not be assigned at random to treatment conditions (Cohen, Manion & Morrison, 2000). Erduran, 2008 and others (see, e.g., Hohenshell & Hand, 2006) recommend that mixed-methods models be used in which qualitative research is conducted after a well-reasoned quantitative experiment to understand the nuances of treatment effects (Erduran, 2008; Hohenshell & Hand, 2006).

Critical Thinking Assessments

Although the majority of critical tests are aimed at college student and adults, the present discussion is focused only on critical thinking tests designed for elementary and middle school grades. Approaches to creating such tests reflect the diversity in definitions and breadth of critical thinking skills already discussed. Assessments used include multiple-choice tests, interviews, naturalistic observations, essay tests, and performance-based tasks (Ennis, 1987, 1995). The three most widely-distributed assessments for elementary and secondary students are: the Cornell Critical Thinking Test Form X and Z (CCTT-X and CCTT-Z; Ennis, Millman & Tomko, 2005), the Critical Thinking

Assessment Test (CAT; Education Testing Service, 1998) and the Watson-Glaser Critical Thinking Appraisal (WGCTA; Watson and Glaser, 1996).

Cornell Critical Thinking Tests

(CCTT, X & Z, Ennis, Millman & Tomko, 2005)

Ennis, Millman and Tomko, (1985, 2005) developed the Cornell Critical Thinking Tests, form X and Z (CCTT-X and CCTT-Z) based on the premise that critical thinking is a reflective process in deciding what to believe in or do. The two forms of the CCTT are designed for different ranges of grade levels with subscales intended to measure dispositions and specific abilities corresponding to developmental levels. These tests include multiple-choice items that measure domain-general critical thinking ability related to classroom learning and real-world experiences. Form CCTT-X is designed primarily for students ranging from upper elementary grades to beginning college levels. The CCTT-Z is catered more to students in gifted high school programs, advanced college students, or other adults. Both levels of the test share a common test manual. The manual (Ennis, Millman & Tomko, 2005) includes information about how the subscales were created, evidence of validity and reliability, administrative procedures, and instructions for score use. The instruments each yield five subtest scores that are summed to provide an overall estimate of critical thinking skill.

Critical Thinking Assessment Test

(CAT, Education Testing Service, 1998)

The Critical Thinking Assessment Test (CAT) was developed by Educational Testing Service (ETS, 1998) to assess problem solving skills based on real world scenarios intended to engage and interest students. The CAT unique in that it allows for personalized responses. This approach reflects a principle of “dynamic assessment” in which examinees write short essays to link their learning experiences with complex problems that mimic aspects of real world problem solving. Weaknesses in this approach include longer administration time per item, subjectivity in scoring, and possible inconsistencies in scoring among raters. To address the latter two issues, the manual includes detailed scoring rubrics and training instructions for raters. The test is also evaluated annually to inform better use and facilitate revisions.

Watson-Glaser Critical Thinking Appraisal

(WGCTA, Watson and Glaser, 1996)

The Watson-Glaser Critical Thinking Appraisal focuses on decision-making and judgment skills. Items are targeted to aspects of inference, recognition of assumptions, deduction, interpretation and evaluation of arguments. Each question takes about five minutes to answer and involves reading passages that include problems, statements, arguments, and interpretations. The original versions of its two alternative forms (WGCTA-A and WGCTA-B) each have 80 items that can be completed in 60 minutes. Watson and Glaser later developed a shorter version (Form S) comprised of 40 items that can be completed in 30-45 minutes. The WGCTA was developed and normed as a paper-and-pencil measure. In Table 2, characteristics of the three instruments are summarized.

Table 2. Characteristics of the CCTT, CAT, and WGCTA

	Description	Authors/Source
Cornell Critical Thinking Test, Form X (CCTT-X)	• Content: Measure students reasoning ability related to their classroom learning experiences and the real-world corresponding experiences • Format and Distribution: A multiple choice task, inducing 71 items with 5 scenarios, suggested to be taken within 45 minutes to complete • Subscales: observation, induction, credibility, deduction, definition, and assumption identification • Target: Students from grade five to grade 14	Ennis, Millman, and Tomko (2005); P.O. Box 1610, Seaside, CA 93955
Cornell Critical Thinking Test, Form Z (CCTT-Z)	• Content: Content: Measure students reasoning ability related to their classroom learning experiences and the real-world corresponding experiences • Format and Distribution: A multiple choice task, inducing 75 items with 6 scenarios, suggested to be taken within 50 minutes to complete • Subscales: induction, credibility, prediction and experimental planning, fallacies(especially equivocation), deduction, definition, and assumption identification • Target: Students in advanced placement program or gifted education	Ennis, Millman, and Tomko (2005); P.O. Box 1610, Seaside, CA 93955

Critical Thinking Assessment Test (CAT)	• Content: Measure one of the broad academic areas of humanities, social sciences, or natural sciences that require skills in inquiry, analysis and communication • Format and Distribution: A performance-based assessment inducing 45 items, suggested to be taken within 90 minutes to complete • Subscales: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. • Target: Upper level elementary and secondary level school students	Fisher (2001); Educational Testing Service, Princeton, NJ 08541
Watson-Glaser Critical Thinking Appraisal, Form A & From B (WGCTA-A and -B)	• Content: Measure general abilities related to critical thinking • Format and Distribution: Two alternative forms, each including 80 items with 32 scenarios, suggested to be taken within 60 minutes. • Subscales: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. • Target: upper level elementary and secondary school students	Watson and Glaser (1996); Harcourt Brace Educational Measurement (Psychological Corporation) 555 Academic Court, San Antonio, Texas 78204
Watson-Glaser Critical Thinking Appraisal, Form S (WGCTA-S)	• Content: Measure general abilities related to critical thinking • Format and Distribution: 40 items with 16 scenarios, suggested to be taken within 30-45 minutes. • Subscales: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. • Target: upper level elementary and secondary school students	Watson and Glaser (1997); Harcourt Brace Educational Measurement (Psychological Corporation) 555 Academic Court, San Antonio, Texas 78204

Studies about Critical Thinking

In this section, the literature on critical thinking pertinent to education is reviewed. Topics include: developmental changes, educational experiences, and gifted and special education programs.

Critical Issues Related to Studies on Critical Thinking

Developmental Changes

According to Kuhn (1993)'s developmental model, mental representations of reality can be right or wrong and critical thinking is a vehicle by which reality and assertions can be compared and assessed. Research following a Piagetian tradition tended to depict the cognitive processes of young children as deficient in relation to those of adults (Kennedy et al., 1991). Piaget's stages of development imply, for example, that children are incapable of essential features of thought such as formal operations or abstract reasoning. However, recent empirical research has shown that young children can use the same cognitive processes as adults (Dam & Volman, 2004; Kuhn, 1993; Kuhn & Udell, 2003). Although critical thinking ability is improved with age, even students at the primary grade level can benefit from critical thinking instruction (Kennedy, et al.,1991).

Children begin developing critical thinking competencies at a very young age (Dam & Volman, 2004; Kennedy et al, 1991; Kuhn & Udell, 2003; Willingham, 2007). Toward the end of the preschool years, young children can recognize that mental representations are products of the mind. When those representations do not mirror external reality, they become susceptible to falsification (Kuhn, 1999; Heyman & Legare, 2005). Although in theory all people can be taught to think critically (Dam & Volman, 2004; Kuhn, 1993; Kennedy et al, 1991), national reports and empirical data indicate that many college students and adults are deficient in such skills (Geier et al, 2008).

Koenig and Harris (2005), when observing children's interactions with surroundings, found that children are able to differentiate the credibility of various sources of information as early as three and four years of age. Kuhn (1999), in her developmental model of critical thinking, asserted that children's reasoning and thinking show an obvious improvement at pre-school age. Similarly, Heyman and Legare (2005) found that, compared to younger children, seven year-old children become increasingly aware of others' motives to distort the truth.

Educational Experiences

Researchers have examined how critical thinking skills relate to traditional, school related experiences (Baron, 1987) as well as the effects of interventions intended to improve critical thinking (Abrami et al, 2008; Anderson et al, 2001; Bataineh & Zghoul, 2006). Included in this research are studies focused on how school experiences, skills, and attitudes relate to critical thinking in science (Becker, 2006; Dam & Volman, 2004; Erduran et al, 2004; Halonen, 1995). The majority of studies on critical thinking have involved either young children or college student and adults. Historically, researchers believed that studies of more mature individuals would provide better insights into the nature of critical thinking. Later, research was targeted at very young children in part to dispel the notion than such children are not equipped to think critically (Dam & Volman, 2004; Kuhn, 1993; Kuhn & Udell, 2003). In comparison, research on individuals between pre-school and post-secondary school years is more limited. Most researchers working in the area of critical thinking agree that background knowledge is essential for students to demonstrate critical thinking (Kennedy et al., 1991; Willingham, 2007). In general, the more education people receive, the higher their critical thinking abilities.

Paul and Elder (2006) used the Cornell Critical Thinking Test Form Z to study the differences among gifted students. They found that gifted eighth and ninth grade students had higher intellectual reasoning than gifted sixth grade students and attributed these differences in part to the increasing intellectual demands of higher level coursework.

Kalman (2002) investigated graduate and undergraduate students' overall critical thinking ability using the Watson-Glaser Critical Thinking Appraisal (WGCTA) subscales for Inference, Recognition of Assumption, Deduction, and Total Critical Thinking. Overall, they found no gender differences but that graduate students outperformed undergraduates on assumption and deduction. From these results, they emphasized the need to better mentor teachers in improving critical thinking skills.

King, Wood, and Mines (1990) compared undergraduate students to graduate students in the social sciences and math and found a significant effect for educational level on two of critical thinking assessments. The total scores from the WGCTA increased with educational levels for both male and female social science majors. However, for mathematics majors, scores decreased for female graduate students compared to female undergraduate students. The researchers concluded that higher critical thinking abilities are not necessarily an inherent outcome of higher education.

Rogoff (2003) studied critical thinking abilities of 116 undergraduate students enrolled in a teacher preparation program using the Cornell Critical Thinking Test Form Z (CCTT-Z). They concluded that critical thinking was a meaningful contributor to success in the program. Male students outperformed female students in total scores and all six subscales, except observation. Similarly, Downs (2008) found that doctoral students outperformed masters-level students in both academic achievement and critical thinking.

Higgins et al (2004) tested 155 pre-service teachers' critical thinking ability using the Watson-Glaser Critical Thinking Appraisal (WGCTA) and Critical Thinking Assessment (CTA). They found that total critical thinking scores on the WGCTA for this group was below national norms at the time of testing but that females outperformed males. The researchers speculated that teachers with weak critical thinking skills might be unable to teach or apply such skills in the classroom.

Taken as a whole, the research reviewed here highlights the importance of emphasizing critical thinking at all age levels, and particularly in teacher training programs. The research further reveals that critical thinking can improve with experience in school but such growth cannot be presumed across all situations or for particular gender groups.

Gifted and Special Education Programs

Historically, demonstration of critical and creative thinking has been considered a key criterion in identifying gifted students (Knight, 2005; Paul & Elder, 2006). Experiences in teaching critical thinking within gifted programs then led to beliefs that such skills can be taught not only to gifted students but to general education students as well. Critical thinking is also considered particularly important for instructors in special education. For example, in designing an appropriate individualized education program (IEP) to guide educational opportunities of students with disabilities, special educators must be able to evaluate information gathered from observations and research-based interventions. Smith and Simpson (1989) concluded that without understanding and possessing critical-thinking abilities, special educators cannot perform competently. They found that the more experienced and educated special education graduate students were, the higher levels of critical thinking on the WGCTA they achieved, when compared to the pre-service special education students.

McInerny and McInerny (2002) concluded that there is a reciprocal relationship between critical thinking and research skills of students studying special education. In particular, students' critical thinking skills increased as their research skills improved, and such thinking skills were positively correlated with research competent and ability to implement research-based interventions. The use of research-based interventions that emphasized understanding application, evaluation, and synthesis of information were considered critical components in facilitating critical thinking within the curriculum.

Kuhn (1999) argues that, while everyone may not reach the highest stages of critical thinking, most people have the potential to become critical thinkers. Empirical research also suggests that students of all intellectual levels can benefit from critical thinking instruction (Kennedy et al, 1991; Reece). Critical thinking skills, in turn, can increase the possibility of succeeding in all academic disciplines. (Abrami et al, 2008; Bransfor & Schwartz, 1999; Downs, 2008; Nickerson; 1988),

Instructional Implications

In this section, instructional implications related to the teaching and transferability of critical thinking skills are explored. Specific instructional recommendations and development of effective learning environments for fostering the development of critical thinking will be emphasized.

Teaching of Critical Thinking

Critical thinking has been viewed as "consistent internal motivations to act toward or respond to persons, events, or circumstances in habitual, yet potentially malleable ways" (Facione, 2000, p. 64). A key question is whether it is possible to teach critical thinking skills. Programs and curriculum intended to promote students' critical thinking have been launched, mostly at the college level, to address this question. Primary concerns within these programs are what and how to teach.

Bailin et al. (1999) proposed that critical thinking instruction should include teaching students to value reason and truth; respect others during discussion; be open-minded; be willing to see things from another's perspective; and perceive the difference between definitions and empirical statements. Facione (1990) urged that critical thinking be taught in domain-specific contexts, or with content drawn from events in everyday life and real-world experiences.

Ennis (1989) proposed four overarching approaches to teaching critical thinking: general, infusion, immersion, and mixed-model, depending on the degree to which critical thinking is taught explicitly or integrated into general instruction. The general approach involves teaching general critical thinking skills in a separate course. The infusion approach embeds general critical thinking principles within a specific subject matter course. The immersion method does not incorporate explicit instruction in critical thinking, but assumes that students can gain the subject-specific critical thinking skills through taking that specific course. Both the infusion and the immersion approach in particular assume that students will acquire critical thinking skills through engagement with the subject matter. The mixed approach would combine the general critical thinking course approach with either the infusion or immersion subject-specific approaches (Ennis, 1989).

Halpern (1998) also advocates explicit instruction in critical thinking. Her model for teaching critical thinking includes instruction in dispositions and skills related to critical thought and structured training in distinguishing and retrieving of information that lead to enhancing reasoning and problem-solving ability, as well as metacognitive monitoring, such as evaluating the appropriateness of goals and progress toward them.

The effectiveness of teaching general learning strategies is unclear (Chiapetta, 2008; Fisher, 2003; Halpern, 1998). Kennedy et al. (1991) reviewed studies from 1960 to 1990, and concluded that instructional interventions have generally shown positive impact but that no single particular approach is superior to others. Overall, they recommend using the mixed approach.

Use of argument is an example of a general cognitive learning practice that can be applied across a variety of domains, such as language arts, mathematics, debate, economics and science (Driver, et al., 2000; Kuhn & Udell, 2003). The structure of the argument has some similarities across these domains, but specific content and context are also important. However, Fisher (2001, 2003) emphasizes that simply telling students about being critical is not substitute for more elaborative and heuristic instruction in how to think critically (Fisher, 2001, 2003).

Gellin (2003) reviewed different types of explicit instruction and asserted that heuristic instruction (i.e., providing detailed instruction in critical thinking) best enhances student learning. Ennis (1987), who studied seventh and eighth graders' scientific thinking dispositions and academic achievement in middle school settings, found a correlation between improved critical thinking and improved project-based learning. Halpern's (2001) research supported a mixed approach involving both domain-general and domain-specific instruction.

Belenky (1986) compared the effectiveness of embedded instruction (n = 65) and immersion (n = 60) approaches to teaching critical-thinking skills to fifth-grade students using the Ennis-Wier Critical Thinking Essay Test. His findings revealed no statistically significant difference between the two teaching approaches. Gokhale (1995) studied pre-service teachers using a systematic intervention module consisting of ten two-hour sessions with the purpose of increasing total Critical Thinking Assessment (CTA) scores of 107 pre-service teachers. The post-test results revealed superior performance for individuals exposed to the intervention module compared to those in a control group.

Sternberg (1986) launched a thinking training program in conjunction with other services to promote critical and creative thinking. His findings revealed that the program did promote overall development of scientific critical thinking. The effects of critical thinking on academic achievement were not necessarily immediate but tended to be long-lasting (Abrami et al, 2008; Sternberg, 1993; Tindal & Nolet, 1995).

Taken as a whole, most studies reviewed here reveal that critical thinking is teachable using a wide range of instructional approaches, but not how to best integrate these efforts in a global curriculum (Thayer-Bacon, 2000). Nevertheless, these findings might encourage teachers to view scientific critical thinking as an ability that can be enhanced through various means in science classrooms (Bataineh & Zghoul, 2006; Ku, 2009).

The Transferability of Critical Thinking

Learning transfer is a phenomenon in which new and unfamiliar learning tasks are approached through the application or integration of previous experiences. The more integration of previous dispositions, knowledge, and skills into curriculum and instruction, the more transfer becomes accessible (Bransford & Schwartz, 1999; Brown, 1990; Halpern, 1998; Thayer-Bacon, 2000). The extent to which critical thinking skills can be transferred to new contexts is closely related to domain-specificity versus domain-generality.

Cognitive perspectives generally depict higher-order thinking as dominantly domain-specific with spontaneous transfer to new contexts being rare (Ennis, 1989; Kennedy et al., 1991). That is, students' abilities have little room to transfer from one domain to another. Willingham (2007) reviewed the literature on critical thinking instruction and concluded that teaching general competence in thinking is difficult because a student's critical thinking skill in one context, or domain, may not stretch across multiple contexts or domains.

In opposition to the conclusions above, findings from several studies has affirmed that student's learning from one discipline can be transferred across other areas including extracurricular, community, or workplace activities (Bransford & Schwartz, 1999; Brown, 1990). Halpern (2001), for example, followed 120 college students for six months to determine whether they would transfer critical thinking skills acquired in the context of a specific discipline to an entirely new curriculum. Six months after the course was over, most students in the study still possessed and were to apply the reasoning techniques they had previously learned to a non-academic topic (Halpern, 2001).

Nickerson (1988), after reviewing instructional practices to improve students thinking, concluded that the success of transferability depends on the content of what is being taught and how it is being taught. Hendricks (2001) studied the distinction between traditional schooling that decontextualizes knowledge versus situated learning. A total of 220 seventh graders were studied and assigned at random to experimental or control groups. The students were then taught about causality, an important component of critical thinking. They were given a “transfer task” to complete two weeks after the instruction and another six weeks after the instruction. Interviews were also conducted after six weeks. Because transfer was very poor for both groups of students, Hendricks concluded that more direct transfer training that corresponds to learners’ internal values is likely needed.

Boscolo and Mason (2001) studied writing as a learning tool to express ideas, descriptions, and explanations. They designed a writing curriculum to examine the extent to which learning to write in history could be transferred to science. They found that students who used writing as a learning tool in history also used it in their science class and that students gained deeper conceptual understandings than those who were not taught to use writing as a learning tool. Their two-year, multi-institutional research projects are among the most important in enhancing our understanding of writing as a leaning tool and transfer mechanism. The first-year writing curriculum provides students with essential knowledge and skills in composition that enable them to transfer what they learn to other course work. The second year’s curriculum is intended to facilitate transfer of writing skill to post-graduate, professional and other endeavors. The curriculum includes a series of “enabling practices” that promote writing transfer, including introducing rhetorically-based concepts, engaging students in metacognition and metacognitive awareness, and implementing explicit instruction and modeling transfer-focused thinking. Their findings suggest that, with explicit rhetorical education, students are more likely to transform rhetorical awareness into performance. Transfer of rhetorical knowledge and strategies between self-sponsored and academic writing also are encouraged by designing writing tasks with authentic audiences and purposes that prompt metacognitive reflection and engage students in the active writing process.

Success of transfer also depends on the distance between or similarity of tasks (Bailin, 2002). Transfer skills to a new but similar task versus to an entirely new discipline vary. In addition, individual difference variables and educational experiences need to be taken into account, and varying degrees of scaffolding may be necessary to help a diverse body of students develop critical thinking.

Hernstein et al. (1986) examined the relationship between critical thinking and academic aptitude within a year-long critical thinking course involving 400 grade-four and grade-five students. Four tests, including General Ability Test, Target Ability Test, IQ test and Watson-Glaser Critical Thinking Appraisal (WGCTA) were given before, during, and after the course to both experimental and control groups. These tests were supplemented with an oral argument examination of randomly selected individuals from the experimental and control groups. Their results revealed that the treatment group performed higher on the General Abilities Test and Target Abilities Test than did the control group.

Royalty (1995) studied 180 college students to determine the extent to which critical thinking skills could be applied to novel domains. The Cornell Critical Thinking Test, an IQ test, and a Statistical Reasoning test, were administered. Both IQ and critical thinking test scores were correlated with reasoning skill, with critical thinking scores accounting for a unique proportion of the variability. He concluded that critical thinking skills can be transferred to novel subject areas, but the role of aptitude and its overlap with critical thinking needs to be further clarified.

Learning Environments that Promote Critical Thinking

Designing learning environments to support critical thinking in science and other classrooms is not easy. Conventional school education systems can undermine development due to time constrains, absence of appropriate learning materials, and emphasis on passive assimilation of scientific knowledge and theories (Tsui, 2002, Van Gelder, 2005). Explicit instruction, collaborative or cooperative learning, constructivist techniques and inquiry based learning approaches all have been advocated to encourage better development of critical thinking skill. Gokhale (1995) and others (see, e.g., Dan & Volman, 2004; Dillenbourg et al.;1996; Heyman, 2008; Nelson, 1994; Paul, 1992; Wertsch, 2008). Gokhale (1995) believes that students' critical thinking is facilitated in environments where collaborative learning is used, regardless of the task domain, by emphasizing the value of social interactions for promoting cognitive development. Erduran et al. (2006) urged teachers to support critical thinking through the practice of argumentation.

Inquiry-based learning environments also have been advocated as effective means to promote critical thinking. In such learner-centered environment, students actively engage in the learning activities, take charge of problem solving, design scientific experiments, interpret data, and compile evidence to support or refute particular hypotheses (Johnson & Johnson, 1993). To promote critical thinking, teachers have been encouraged to utilize models, graphic organizers and concrete examples to illustrate abstract concepts (Heyman, 2008; Van Gelder, 2005). The goal here is to provide tools to structure as well as to evaluate arguments. Other techniques, emerging from the theory of argumentation and critical thinking pedagogy, include concept maps, argument diagrams, argumentative writings and writing framework (Cavagneto, 2010; Erduran et al., 2006; Erduran, Simmon, & Osborne, 2004; Hand, 2008, 2012; Jimenez-Aleixandre & Erduran, 2008; McNeill, 2011; McNeill, Lizotte, Krajcik, & Marx, 2006; Newton et al, 1999; Tusi, 2002).

Another method used to facilitate critical thinking is called Argument-Based Inquiry (ABI). With ABI instruction, teachers ask students leading questions to draw from them information, inferences, and predictions. Empirical evidence regarding the effect of ABI on students' critical thinking ability is limited, but seems to indicate that long term exposure to ABI is needed to foster critical thinking ability (Cavagnetoo and Hand, 2011). The role of possible influences of age, gender, academic achievement, and educational background on critical thinking skills in ABI learning environments is largely unknown.

A common purpose in creating learning environments to promote critical thinking is to engage students in active learning. However, when students engage in the construction of knowledge, a sense of uncertainty may block the process of scientific exploration and critical reasoning. Psychosocial stressors of being questioned or rejected can threaten students and undermine the development of cognitive ability and critical thinking (Willingham, 2007). Thus, from an affective point of view, it is essential for critical thinking to take place within a non-threatening learning environment. That is, students need to work within an emotionally supportive environment so that they can comfortably discuss, critique, and examine their spontaneous ideas (Abbott & Ailks, 2005; Berland & McNeill, 2010; Yor & Treagust, 2006). To encourage students to willingly explore and express ideas, teachers need to accept and empower learners and accommodate their inner needs. In doing so, students gradually take responsibility for their learning, better reflect on their thinking and learning, engage actively in a classroom discourse (i.e., debates and critiques), and ultimately become better thinkers (Koenig & Harris, 2005; Paul, 1995; Sternberg, 1986).

To test some of these ideas, Yeh (1998) studied an eighth grade classroom in which students used argumentative writing as a learning tool for a science-course unit.

When students’ learning was authentic and meaningful, their resistance to documenting what they learned and communicating their thoughts diminished. This finding supports the view that a learning environment should allow students to feel affirmed and free to participate personally in the construction of their knowledge (Yore & Treagust, 2006).

Wertsch (2008) investigated 58 medical school students thinking styles under three instructional conditions. Participants were given informational instructions, conditioning instructions, or no instructions. With informational instruction, students were provided with resources to enhance ownership, whereas with conditioning instruction learning was more direct and constrained. Results indicated that enjoyment, critical thinking, and quality of performance was higher the informational group that for the conditioning group.

Overall, the findings summarized here suggest that behaviors, language, and perceptions of teachers can have important effects on student learning and critical thinking (Bee & Boyd, 2004; Crawford, 2000). Teachers who empower active learning and maintain positive relationships and respectful accept student cognitive knowledge and skill levels are generally more effective in promoting students’ critical thinking. In the chapters to follow, results of a comprehensive study of the effects of instruction using the Scientific Writing Heuristic (SWH, Hand & Keys, 1997) in science classrooms are reported. This approach incorporates many to the themes noted here to facilitate critical thinking skills.

CHAPTER THREE
METHOD

The purpose of this chapter is to establish the methodological framework for the study as well as identify and justify instruments, data collection and data analysis procedures. The chapter will begin with a discussion of the rationale for the research design and the use of gain scores. Next the context of the study and the sample will be identified and characterized, followed by the data collection procedures. Finally, instruments and analytical strategies will be discussed.

Research Design

The present study involved a secondary analysis of Cornell Critical Thinking Test form X (CCTT-X) scores for five data sets using a quasi-experimental design. The main independent variables across analyses were instructional method (Science Writing Heuristic--SWH approach, versus traditional), gender, grade level, and year of implementation. A quasi-experimental design was used because it was not always feasible to randomly assign students to instructional conditions. Consequently, the data were interpreted as exploratory and examined for general trends with the hope of conducting purer experimental studies in the future.

Research Context

The key instructional approach examined in this study was an immersion argument-based inquiry developed by Hand and Keys (1999) called the Science Writing Heuristic (SWH) approach. This approach is structured around questions, claims, and evidence in which students are required to conduct inquiry investigations by posing their own questions about the concept under review, collecting data, constructing claims based on evidence, finding out what experts say, and reflecting upon their arguments to examine how their ideas have changed (Cavagnetto, 2010; Cavagnetto & Hand, 2011). Students consistently negotiate meaning individually, in small groups, and at the whole class level where importance is placed on public and private construction and critique of knowledge.

The SWH approach differs from conventional approaches of teaching argument structure in that it requires students to collect data from their own investigations, make decisions about which data points will be used as evidence, and then construct a logically connected explanation using these data points. Students are required to make decisions about what data are appropriate and provide reasoning on how the data points form evidence to support their claims. Data are viewed as being explicitly different from evidence, with evidence being seen as data plus reasoning. Each of these procedural steps and epistemological viewpoints are necessary to facilitate efficient science learning. As a result, the SWH approach is considered a tool to promote and scaffold scientific argument within science classrooms (Cavagnetto, 2010, Hand, 2004, Gunel & Hand, 2009, Martin & Hand, 2010).

The SWH approach consists of a framework to guide activities as well as a metacognitive support system to prompt students' reasoning about data. The SWH approach consists of two templates: one addresses teacher activities (Figure 1) and the other directs student writing (Figure 2). In the teacher template, the teacher uses a series of strategies, including reading, and writing, from small group to whole class discussion, to support students' engagement with the activity. Teachers are encouraged to provide students with multiple opportunities to negotiate meaning from their experience. In essence, the teacher template illustrates the necessary pedagogy to support student learning. The student template serves to scaffold student understanding of scientific concepts while writing the laboratory report by relating claims to evidence. The teacher template highlights important phases of suggested activities to enhance learning by promoting negotiation of meaning among students and/or among students and teachers in both small- and large-group activities. Moreover, the teacher template provides strong pedagogical focus for implementing and conducting scientific investigation as a means to learn scientific methods and procedures (Figure 1 and Figure 2).

- **Exploration of pre-instruction understanding**
- **Pre-laboratory activities**
- **Laboratory activity**
- **Negotiation- individual writing**
- **Negotiation- group discussion**
- **Negotiation- textbook and other resources**
- **Negotiation- individual writing**
- **Exploration of post-instruction understanding**

Figure 1. A template for teacher-designed activities to promote science learning (adapted from Hand et al, 2008)

- **Beginning questions or ideas**: What are my questions about this experiment?
- **Tests and Procedures**: What will I do to help answer my questions?
- **Observations**: What did I see when I completed my tests and procedure?
- **Claims**: What can I claim?
- **Evidence**: What evidence do I have to support my claim? How do I know? Why am I making these claims?
- **Reading**: How do my ideas compare with others?
- **Reflection**: How have my ideas changed?

Figure 2. The SWH approach, a template for students' thinking

Participants

All data for the current study were collected at either elementary and/or secondary school science classrooms in public schools of the Midwest in the USA. The longitudinal data reported here were collected over two years. All data sets are part of larger SWH projects. The participants in each school reflect similar proportions of ethnic, students with individual education programs, and social economic status distributions (see Table 3). The unique features of each data set follow.

Data Set 1

Data Set 1 was part of the largest cluster involved in the SWH grant project, with 282 teachers and over 4500 students coming from 48 buildings. All were elementary schools located approximately within twenty miles of large urban cities in the central, southeastern, northeastern and west regions of the same state. The school districts consisted of approximately 3% minority students and 97% white students. Data reported here are for the first years of a three-year grant. Specifically, in Data Set 1, participants were 1138 5^{th} graders (594 and 544 in treatment and control group, respectively) and 462 6^{th} graders (265 and 197 in treatment and control group, respectively). Gender distributions were nearly identical across the SWH treatment and control groups with approximately 57% female. Ethnic distributions were also highly similar across the treatment and control groups, and largely Caucasian (96% or more).

Data Set 2

Data Set 2 included 29 teachers and over 1,950 students from 7 buildings in northeastern region of the state. All were middle/junior high schools classified as middle to upper middle class in socio-economic status with about 5% minority students. The participants consisted of 160 6^{th} graders (77 and 83 in treatment and control group, respectively), 239 7^{th} graders (149 and 90 in treatment and control group, respectively), and 500 8^{th} graders (304 and 196 in treatment and control group, respectively). The gender distributions were nearly identical across the SWH treatment and control groups with a 49/51% split for males versus females. The ethnic distribution was also highly similar across the treatment and control groups, and largely Caucasian (>97%).

Data Set 3

Participants within Data Set 3 were recruited for a particular research purpose with the majority of participants involved in the SWH treatment group. This school district was classified as being middle to upper middle class socio-economic status. Specifically, participants are 260 6th graders (152 and 108 in treatment and control group, respectively), 264 7th graders (149 and 115 in treatment and control group, respectively), and 450 8th graders (254 and 196 in treatment and control group, respectively). The gender distributions were nearly identical across the SWH treatment and control groups, and approximately 52% female. The ethnic distribution was also highly similar across the treatment and control groups, and largely Caucasian (>97%).

Data Set 4

Data Set 4 was a subset of 433 students from Data Set I who were followed for two consecutive years encompassing 5th and 6thgrades. The gender distributions again were nearly identical across the SWH treatment and control groups, and close to 53% female. The ethnic distribution was highly similar as well across the treatment and control groups and largely Caucasian (>95%).

Data Set 5

Data Set 5 was a subset of Data Set 2 with 511 students followed for two successive years in 6th and 7th grades. The gender and ethnic distributions were nearly identical across the SWH treatment and control groups with about 51% female and more than 97% Caucasian.

Table 3. Demographic Information of Participants

Data Set		Grade	Total (N)	SWH Subtotal (# of Students)	SWH Gender(%) Male	SWH Gender(%) Female	SWH Ethnicity (%) White	SWH Ethnicity (%) Others	SWH IEP(%)	Traditional Subtotal (# of Students)	Traditional Gender(%) Male	Traditional Gender(%) Female	Traditional Ethnicity (%) White	Traditional Ethnicity (%) Others	Traditional IEP(%)
Cross-Sectional	1	5	1138	594	45	55	98	2	15.4	544	45	55	98	2	12.4
		6	462	265	42	58	96	4	14.7	197	42	58	97	3	11.3
	2	6	160	77	50	50	96	4	14.0	83	49	51	96	4	11.3
		7	239	149	46	54	95	5	13.3	90	46	54	95	5	13.3
		8	500	304	43	57	95	5	12.6	196	43	57	94	6	13.7
	3	6	260	152	40	60	93	7	16.9	108	40	60	93	7	13.1
		7	264	149	50	50	92	8	11.2	115	47	53	92	8	15.5
		8	450	254	44	56	95	5	15.5	196	44	56	95	5	12.7
Longitudinal	4	5/6	433	227	45	55	98	2	13.4	206	45	55	98	2	13.7
	5	7/8	511	259	46	54	95	5	17.7	252	46	54	95	5	15.7

Note: N/A=no data collected

Instruments, Data Collection, and Analysis

Critical Thinking Assessment

The Cornell Critical Thinking Test (Level X, CCTT-X, Ennis & Millman, 2005) measures critical thinking ability of students in grades 4 through 14. The test consists of four sections with a total of 71 multiple choice items. All questions are designed around the main theme of the exploration of a new planet. The questions assess critical thinking skills in general without specific content, and there is no requirement of science content to complete this test. The test contains five subscales of critical thinking: induction, deductions, credibility, observations, and assumptions. These subscales are closely aligned with the thinking skills necessary to engage in the SWH approach.

Reliability information reported in the manual includes both KR-21 and Spearman Brown, with KR-21 coefficients ranging from .55-.83 and split half coefficients ranging from .71-.90 (Ennis, Millma & Tomka, 2005). CCTT-X-scores are also significantly correlated with scores from other critical thinking tests (e.g., Watson-Glaser (Zm), Logical Reasoning Test, Part II, Form A) with coefficients ranging from 0.31 to 0.62.

In the CCTT-X manual, Ennis, Millman, & Tomko (2005) report descriptive statistics (e.g., Ns, means, standard deviations) for 29 samples studied from the 1960s through 1985. Groups sampled included average and gifted students in grades four-twelve, average college freshmen, and average professional public school teachers. These individuals were sampled from public school districts and higher education institutions in rural, suburban, and urban areas within the states of California, Colorado, Florida, lllinois, New York, Ohio, and Washington. Means and standard deviations of CCTT-X total number correct scores for groups of possible interest in the present study are provided in Table 4.

Table 4. CCTT-X Total Score Means and Standard Deviations from Previous Research Related to the Current Study Reported in the Administration Manual for the Cornell Critical Thinking Tests (Ennis. et al, 2005)

Description of School	Grade	N	Mean	SD
1. Students in an integrated, predominantly middle-class school district in a median-sized city in downstate Illinois.	4,5,6	64	36.8	8.0
2. Students in an agricultural community in central Wyoming	4,5	165	32.0	7.8
3. Students in a Catholic elementary school in a large city in Nebraska	5	49	37.9	7.6
4. Social studies students in a private school for gifted in a Midwestern stat	4,5	35	38.4	5.1
5.Science Students in for gifted program in Illinois state-found project	6,7,8	39	38.4	5.1
6. Students in a whole-school middle school in Illinois	8,9,11,12	135	40.4	7.1
7. Students in a Catholic middle school in a large city in Nebraska	9,10,11	55	45.4	8.6

Data Analysis

Students yearly CCTT-X gain scores (posttest-pretest) were used as the main dependent variable. Posttest and pretest scores were derived by counting the total number of correct responses.

Study Variables

For all reported analyses, student participants were divided into SWH and traditional instruction groups, and this classification served as the main independent variable of interest in all analyses. Several co-variables also were included in the analyses to increase statistical accuracy and precision, and to isolate more specifically the effects of treatment on critical thinking performance. These included gender and grade level in the cross-sectional analyses (Data Sets 1-3) and gender, grade level and year of implementation in the longitudinal analyses (Data Sets 4-5). Descriptive statistics for CCTT-X pretest, posttest, and yearly gain scores are provided for all data sets and research conditions in Chapter 4.

Analysis of CCTT-X Gain Scores

Quasi-experimental designs commonly used in educational research include the nonequivalent control group design (Campbell & Stanley, 1966) and untreated control group design with pre-test and post-test (Cook & Campbell, 1979). In both designs, groups (treatment and control) are given a pretest and posttest, but assignment of students to each group is not random. The phrase "nonequivalent group" refers to: (1) lack of randomization of students to each group, which is considered a crucial element to the validity of causal conclusions, and; (2) the lack of convincing evidence that the groups are essentially the same (Campbell & Stanley, 1966). In contrast, the untreated control group design with pre-test and post-test refers merely to the lack of randomization in assigning students to groups.

Computation of Gain Scores

Gain Scores, standing for changes in CCTT-X total raw scores between pretest and posttest, were used as composite measures of critical thinking over a given school year. The change from pretest to posttest was computed by subtracting each person's pretest score from his or her posttest score –

$$G = X - Y, \quad \text{(Equation 1)}$$

where G is the gain score, X is the posttest score, and Y is the pretest score.

Reliability of Gain Scores

Determining of reliability of yearly gain scores was important. The reliability of gain scores was calculated using the following equation (adapted from Stanley, 1971, p. 385, formula 24):

$$r_{DD'} = \frac{r_{xx'}\hat{\sigma}_x^2 + r_{yy'}\hat{\sigma}_y^2 - 2r_{xy}\hat{\sigma}_x\hat{\sigma}_y}{\hat{\sigma}_x^2 + \hat{\sigma}_y^2 - 2r_{xy}\hat{\sigma}_x\hat{\sigma}_y}, \quad \text{................................ (Equation 2)}$$

where $r_{DD'}$ is the estimated reliability coefficient for the gain scores, $r_{xx'}$ is the estimated internal consistency of pre-test scores, $r_{yy'}$ is the estimated internal consistency of post-test scores, $\hat{\sigma}_x$ is the estimated standard deviation of pre-test scores, $\hat{\sigma}_y$ the estimated standard deviation of post-test scores, and r_{xy} is the sample correlation coefficient between pretest and posttest scores.

A split-half approach was used to calculate the internal consistency of pre- and post-test scores for each administration of the CCTT-X. The following method was adapted from Crocker and Algina (1986, p.136-137) to derive the differences scores used in calculating internal consistencies.

1. Items were rank-ordered in difficulty (p-values), based on responses of the examinees, creating 35 pairs plus one additional item.
2. One item within each pair was assigned at random to half 1 and the other to half 2. The remaining item was then assigned at random to one of the sets.
3. The reliability of the full length pre- and post-test scores was then computed using the formula below adapted from Raju (1977) in which test halves can have unequal numbers of items.

$$r_{xx'(Raju)} = \frac{r_{x_1x_2}\hat{\sigma}_{x_1}\hat{\sigma}_{x_2}}{\left(\frac{k_1}{k_1+k_2}\right)\left(\frac{k_2}{k_1+k_2}\right)\hat{\sigma}_x^2}, \quad \text{.. (Equation 3)}$$

where $r_{xx'(Raju)}$ is the estimated reliability of the full length test, $r_{x_1x_2}$ is the sample correlation coefficient between half-test scores, $\hat{\sigma}_{x_1}$ is the estimated standard deviation of half-test 1 scores, $\hat{\sigma}_{x_2}$ is the estimated standard deviation of half-test 2 scores, k_1 is the number of items in half-test 1, k_2 is the number of items in half-test 2, and $\hat{\sigma}_x^2$ is the estimated variance of the full-test scores.

Research Questions and Associated Data Sets

Research Question 1: To what extent do critical skills change from the beginning to the end of each academic year. (Data sets 1-5)

Research question 2: What are the main and interactive effects of instructional approach, gender, and grade level on CCTT-X gain scores across a single academic year. (Data sets 1-3)

Research Question 3: What are the main and interactive effects of year of implementation, instructional approach, and gender on CCTT-X gain scores across two academic years. (Data sets 4-5)

Statistical Analyses

Research Question 1 (To what extent do critical skills change from the beginning to the end of each academic year?) was addressed using paired samples t-tests. For each academic year, CCTT-X mean scores at the beginning of the year were compared to CCTT-X mean scores at the end of the year. Changes in mean scores were examined across instructional methods as well as separately for students receiving SWH and traditional instruction for each of the five data sets.

Research question 2 (What are the main and interactive effects of instructional approach, gender, and grade level on CCTT-X gain scores across a single academic year?) was addressed using three-way ANOVAs with Instructional Approach (SWH versus traditional) Gender (male versus female), and Grade Level (5^{th} versus 6^{th} in Data Set 1; 6^{th}, 7^{th}, and 8^{th} in Data Sets 2 and 3) as independent variables, and yearly CCTT-X gain scores as the dependent variable.

Research Question 3 (What are the main and interactive effects of year of implementation, instructional approach, and gender on CCTT-X gain scores across two academic years?) was addressed using split-plot ANOVAs with one within-subject factor (Year of Implementation) and two between-subjects factors (Instruction Method and Gender). Analyses were limited to Data Sets 4 and 5 in which gain scores were available for two successive years. Yearly CCTT-X gain scores served as the dependent variable in these analyses.

Effect-size indexes were derived to facilitate interpretation of statistically significant in all analyses. These indices consisted primarily of standardized mean differences (i.e., d-values). Such indices were computed by dividing an observed mean difference by the pooled within-group estimate of the population standard deviation of the dependent variables. According to Cohen (1988), d-values between .20 and .49 are categorized as small, those between .50 to .79 as medium, and those greater than or equal to .80 as large. A summary of research questions and associated analyses appears in Table 5.

Table 5. Summary of Research Questions and the Corresponding Analyses

Research Question	Data set	Independent	Dependent	Analysis
1. To what extent do critical skills change from the beginning to the end of each academic year	1-5		Within-year Pre- and Post- CCTT-X Scores	Paired Sample T-test
2. What are the main and interactive effects of instructional approach, gender, and grade level on CCTT-X gain scores.	1	Instructional approach, Gender, Grade (5, 6)	CCTT-X Gain Score for First Year of Implementation	Three-way ANOVA
	2	Instructional approach, Gender, Grade (6, 7, 8)	CCTT-X Gain Score for First Year of Implementation	Three-way ANOVA
	3	Instructional approach, Gender, Grade (6, 7, 8)	CCTT-X Gain Score for First Year of Implementation	Three-way ANOVA
3. What are the main and interactive effects Year of Implementation, Instruction Method and Gender on Yearly CCTT-X gain scores.	4	Instructional approach, Gender Year of implementation	Yearly CCTT-X gain score	3-way Split-Plot ANOVA
	5	Instructional approach, Gender Year of implementation	Yearly CCTT-X gain score	3-way Split-Plot ANOVA

CHAPTER FOUR

RESULTS

Overview

The purpose of this chapter is to report results for the data analyses. The chapter begins with reporting of reliability information for critical thinking skill test scores, followed by sections corresponding to the three research questions addressed in the study. These sections focus on: (a) changes in critical thinking scores, as measured by the Cornell Critical Thinking Test-X (CCTT-X) within each academic year examined, (b) cross-sectional analyses into the effects of instructional method, gender, and grade level on yearly critical thinking skill gain scores, and (c) longitudinal analyses into effects of year of implementation, instructional method, and gender and on yearly critical thinking gain scores. Gain scores were calculated by subtracting critical thinking test scores for the start of the school year (October) from critical thinking scores at the end of the school year (May). These scores were used primarily to control for incoming differences in critical thinking scores that may have existed between the groups who received Science Writing Heuristic (SWH) approach versus traditional instruction.

Reliability of Critical Thinking Scores

Reliability of critical thinking scores for each administration of the CCTT-X was calculated using a split-half approach. Items were rank-ordered in difficulty (p-values), creating 35 pairs plus one additional item. One item within each pair was assigned at random to half A and the other to half B. The remaining item then was assigned at random to one of the sets. The reliability of the full length test was computed using the approach described by Raju (1977) in which test halves can have unequal numbers of items. Table 4 shows the means and standard deviations for

CCTT-X pretest and posttest scores for half and full tests across all data sets. Table shows the correlations between test halves, split-half reliability coefficients for pretest and posttest scores, correlations between pretest and posttest scores, and reliability coefficients for yearly gain scores. Split-half reliability coefficients ranged from 0.89 to 0.97 (mdn = 0.945) and gain score reliability coefficients from 0.47 to 0.81 (mdn = 0.70).

Table 6. Means and Standard Deviations for CCTT-X Pretest and Posttest Scores for Half and Full Tests Across All Data Sets

Pretest

Data set	N	Half 1		Half 2		Full	
		Mean	SD	Mean	SD	Mean	SD
Data set 1	1600	18.75	6.43	17.90	6.70	36.65	8.17
Data set 2	899	20.98	5.68	20.16	5.76	41.14	8.19
Data set 3	974	20.94	6.09	20.01	6.33	40.95	8.68
Data set 4							
Year I	433	19.41	6.74	18.79	6.24	38.20	8.74
Year II	433	20.50	6.77	19.69	5.98	40.19	8.76
Data set 5							
Year I	511	19.41	5.74	18.93	5.88	38.34	8.48
Year II	511	21.74	5.32	20.92	5.89	42.66	8.89

Posttest

Data set	N	Half 1		Half 2		Full	
		Mean	SD	Mean	SD	Mean	SD
Data set 1	1600	20.60	6.09	19.72	6.27	40.32	8.44
Data set 2	899	22.90	6.78	22.05	5.74	44.92	8.25
Data set 3	974	22.76	5.69	21.77	5.76	44.46	8.33
Data set 4							
Year I	433	21.29	6.74	20.55	6.76	41.70	8.33
Year II	433	21.77	5.75	20.34	5.77	42.09	8.75
Data set 5							
Year I	511	21.25	6.43	20.20	6.72	41.29	8.44
Year II	511	22.42	6.09	22.23	5.79	44.63	8.85

Table 7. Correlations Between Test Halves, Split-Half Reliability Coefficients for Pretest and Posttest Scores, Correlations between Pretest and Posttest Scores, and Reliability Coefficients for Yearly Gain Scores

Data set	N	Pretest			Posttest			r between halves	Reliability of gain scores
		Mean	SD	r_{Raju}	Mean	SD	r_{Raju}		
Data set 1	1600	36.65	8.17	0.97	40.32	8.44	0.96	0.82	0.81
Data set 2	899	41.14	8.19	0.96	44.95	8.25	0.96	0.84	0.75
Data set 3	974	40.95	8.68	0.96	44.53	8.33	0.95	0.80	0.77
Data set 4									
Year I	433	38.20	8.74	0.95	41.84	8.33	0.94	0.82	0.70
Year II	433	40.19	8.76	0.90	42.11	8.75	0.92	0.83	0.47
Data set 5									
Year I	511	38.34	8.48	0.94	41.45	8.44	0.93	0.80	0.67
Year II	511	42.66	8.89	0.89	44.65	8.85	0.92	0.78	0.57

Changes in CCTT scores from the beginning to the end of each school year

Research Question 1 was focused on the extent to which critical skills improved from the beginning to the end of each academic year. Potential improvements were examined across instructional methods as well as separately for students receiving SWH approach and traditional instruction for each of the five data sets. Data Sets 1-3 included students' pretest and posttest CCTT-X scores for a single year of implementation, whereas Data Sets 4-5 included two successive years of implementation. Means and standard deviations for pretest, posttest, and gain scores and results for dependent sample t-tests comparing pretest and posttest mean scores are presented in Table , Table and Table for the total, SWH approach and traditional groups, respectively. Statistically significant improvements in CCTT-X scores were found for all groups and data sets. Effect size indexes (standardized posttest-pretest mean differences, i.e., d-values) varied from 0.22 - 0.46 (mdn = 0.43) for the total group, from 0.22 - 0.61 for the SWH group (mdn = 0.53) and from 0.13 - 0.33 (mdn = 0.24) for the traditional group. Median d-values for the SWH and traditional groups would reflect medium and small effects respectively based on guidelines suggested by Cohen (1988).

Table 8. CCTT-X Pretest, Posttest, Yearly Gain Scores, Dependent-Sample t-tests, and d-values for the Total Sample in All Data Sets

Data set	N	Pre-test		Post-test		Gain Score		t	d
		Mean	SD	Mean	SD	Mean	SD		
Data set 1	1600	36.65	8.17	40.32	8.44	3.67	3.58	28.04***	0.44
Data set 2	899	41.14	8.19	44.95	8.25	3.81	3.91	29.23***	0.46
Data set 3	595	40.95	8.68	44.53	8.33	3.58	3.99	33.43***	0.45
Data set 4									
Year I	433	38.20	8.74	41.84	8.33	3.64	3.58	21.14***	0.43
Year II	433	40.19	8.76	42.11	8.75	1.92	2.92	13.71***	0.22
Data set 5									
Year I	511	38.34	8.48	41.45	8.44	3.11	2.47	28.49***	0.37
Year II	511	42.66	8.89	44.65	8.85	1.99	2.43	18.52***	0.22

Table 9. CCTT-X Pretest, Posttest, Yearly Gain Scores, Dependent-Sample t-tests, and d-values for the SWH Approach Group in All Data Sets

Data set	N	Pre-test		Post-test		Gain Score		t	d
		Mean	SD	Mean	SD	Mean	SD		
Data set 1	741	36.92	7.99	41.55	7.36	4.63	3.69	24.95***	0.60
Data set 2	369	41.87	7.28	46.62	7.22	4.75	4.19	26.11***	0.66
Data set 3	555	41.35	8.02	45.87	8.24	4.52	4.27	25.22***	0.56
Data set 4									
Year I	227	36.92	8.55	41.12	8.00	4.20	2.43	22.67***	0.51
Year II	227	43.4	8.39	45.34	8.24	1.94	2.21	10.43***	0.23
Data set 5									
Year I	259	37.78	8.8	42.82	8.35	5.04	3.35	27.784***	0.59
Year II	259	43.38	8.63	45.27	8.85	1.89	2.74	14.098***	0.22

Table 10. CCTT-X Pretest, Posttest, Yearly Gain Scores, Dependent-Sample t-tests and d-values for the Traditional Instruction Group in All Data Sets

Dataset	N	Pre-test		Post-test		Gain Score		t	d
		Mean	SD	Mean	SD	Mean	SD		
Dataset 1	859	37.25	8.09	39.80	7.97	2.55	3.10	15.02***	0.32
Dataset 2	369	41.11	7.98	43.56	7.98	2.44	2.96	15.84***	0.31
Dataset 3	419	40.88	7.99	43.23	7.76	2.35	3.20	17.33***	0.59
Dataset 4									
Year I	206	38.06	9.06	40.04	8.39	1.98	1.94	9.46***	0.23
Year II	206	40.17	8.39	42.21	8.97	2.04	2.64	10.93***	0.23
Dataset 5									
Year I	252	38.18	8.87	40.29	8.67	2.11	3.20	16.23***	0.24
Year II	252	42.64	9.03	44.59	9.05	1.95	3.11	12.31***	0.22

Effects of instructional method, gender, and grade level on CCTT-X gain scores

Research question 2 was focused on possible main and interactive effects of instructional approach, gender, and grade level on CCTT-X gain scores. Yearly CCTT-X gain scores served as the dependent variable in a set of three-way ANOVAs with Instructional Approach (SWH versus traditional) Gender (male versus female), and Grade Level (5^{th} versus 6^{th} in Data Set 1; 6^{th}, 7^{th}, and 8^{th} in Data Sets 2 and 3) as independent variables.

Means, standard deviations and sample sizes for CCTT-X gain scores for Data Set 1 are given by instructional approach, gender group, and grade level in Table . The ANOVA results, which appear in Table 10, reveal a statistically significant effect for Instructional Method ($F(1, 1592)=115.3$, $p< .001$). On average, students who received the SWH approach ($\underline{M}$ = 4.63) improved more in critical thinking over the school year than did students who received traditional instruction ($\underline{M}$ = 2.55). The standardized mean difference between the SWH and traditional group equaled 0.61, which would be fall between a medium and large effect according to guidelines suggested by Cohen (1988).

Table 11. Means, Standard Deviations and Sample Sizes for Yearly Changes in CCTT-X Scores by Instructional Approach, Gender, and Grader Level for Data Set 1

Grade		SWH			Control			Total
		Male	Female	Combined	Male	Female	Combined	
5	Mean	4.70	4.66	4.68	2.60	2.51	2.55	3.66
	SD	3.30	3.78	3.55	3.16	3.04	3.10	3.50
	N	286	308	594	270	274	544	1138
6	Mean	4.40	4.66	4.52	2.68	2.42	2.54	3.68
	SD	4.17	3.76	3.97	3.14	3.12	3.12	3.76
	N	136	129	265	93	104	197	462
Total	Mean	4.60	4.66	4.63	2.62	2.48	2.55	3.67
	SD	3.60	3.77	3.69	3.15	3.06	3.10	3.58
	N	422	437	859	363	378	741	1600

Table 12. Three-Way ANOVA Summary Table for Data Set 1

Source	Sum of Squares	df	Mean Square	F	p	□2
Instructional Approach	1358.65	1	1358.65	115.3***	0.000	0.068
Gender	0.317	1	0.32	0.03	0.870	0.000
Grade Level	1.929	1	1.93	0.16	0.686	0.000
Instructional Approach x Gender	6.524	1	6.52	0.55	0.457	0.000
Instructional Approach x Grade Level	1.763	1	1.76	0.15	0.699	0.000
Gender x Grade Level	0.398	1	0.40	0.03	0.854	0.000
Instructional Approach x Gender x Grade Level	4.338	1	4.34	0.37	0.544	0.000
Error	18759.9	1592	11.78			

Descriptive statistics and ANOVA results for Data Set 2 appear in Table 11 and Table 12, respectively. The ANOVA results again reveal a statistically significant main effect for Instructional Method ($F(1, 887)=63.42$, $p < .001$), as well as a significant Gender by Grade Level interaction ($F(2, 887)=3.3$, $p< .05$). Consistent with Data Set 1, students who received SWH instruction (M = 4.76) improved more on average than did students who received traditional instruction (M = 2.44), and this effect would again fall between medium and large (i.e., d = 0.65) according to Cohen (1988). The Gender by Grade level interaction is depicted in Figure 3. Simple effect tests for gender differences at each grade level were statistically significant in all cases with females achieving higher average gains than males at the 6th grade level ($F(1, 887)=44.42$, $p < .001$, $d = 0.46$), and males achieving higher average gains than females at 7th ($F(1, 887)=24.42$, $p < .001$, $d = 0.29$) and 8th grade levels ($F(1, 887)=59.42$, $p < .001$, $d = 0.55$).

Table 13. Means, Standard Deviations and Sample Sizes for Yearly Changes in CCTT-X Scores by Instructional Approach, Gender, and Grader Level for Data Set 2

Grade		SWH			Control			Total
		Male	Female	Mean	Male	Female	mean	
6	Mean	4.12	5.11	4.58	1.58	1.98	1.81	3.14
	SD	3.12	3.01	3.09	4.24	2.86	3.50	3.58
	N	41	36	77	36	47	83	160
7	Mean	5.03	4.08	4.57	3.15	2.67	2.92	3.95
	SD	3.32	4.73	4.08	1.88	2.93	2.43	3.63
	N	77	72	149	47	43	90	239
8	Mean	5.69	4.06	4.89	2.85	2.34	2.49	3.95
	SD	4.6	4.58	4.49	3.19	2.77	2.90	4.11
	N	154	150	304	60	136	196	500
Total	Mean	5.27	4.21	4.75	2.63	2.33	2.44	3.81
	SD	3.88	4.44	4.19	3.19	2.81	2.96	3.91
	N	272	258	530	143	226	369	899

Table 14. Three-Way ANOVA Summary Table for Data Set 2

Source	Sum of Squares	df	Mean Square	F	p	□2
Instructional Approach	871.671	1	871.67	63.42***	0.000	0.041
Gender	22.634	1	22.63	1.65	0.200	0.000
Grade Level	36.591	2	18.30	1.33	0.265	0.002
Instructional Approach x Gender	4.741	1	4.74	0.35	0.557	0.000
Instructional Approach x Grade Level	34.45	2	17.02	1.24	0.290	0.001
Gender x Grade Level	90.578	2	45.29	3.30*	0.038	0.000
Instructional Approach x Gender x Grade Level	21.677	2	10.84	0.79	0.455	0.001
Error	12191.7	887	13.75			

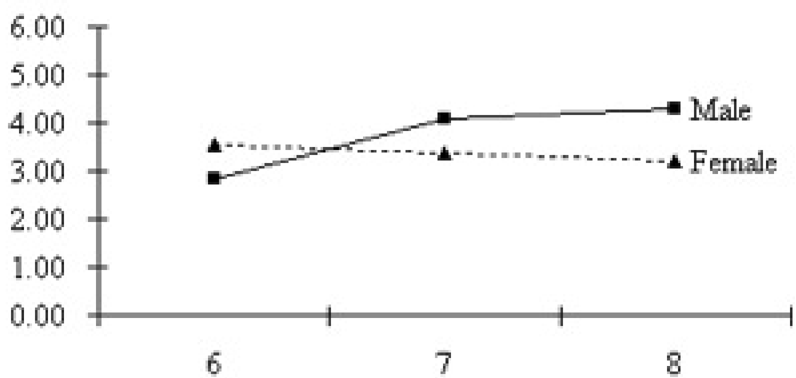

Figure 3. Gender by Grade Level Interaction for Data Set 2 Yearly Gain Scores

Means, standard deviations, sample sizes for Data Set 3 by experimental condition appear in Table , and corresponding ANOVA results appear in Table . The ANOVA results reveal statistically significant effects for Instructional Method, Grade Level and the Gender by Grade level interaction. Once again, students who received SWH instruction improved more in critical thinking over the school year than did students who received traditional instruction, and this difference was more than half of a standard deviation higher for the SWH group (d = .58). The Gender by Grade Level interaction, whose interpretation takes precedent over the Grade Level main effect, is shown in Figure 4. Simple effect test for gender differences at each grade level were statistically significant with females achieving higher average gains than males at the 6th grade level ($F(1, 962)=39.40$, $p < .001$, $d = 0.40$), and males achieving higher average gains than females at 7th ($F(1, 962)=9.48$, $p < .001$, $d = 0.17$) and 8th grade levels ($F(1, 962)=63.08$, $p < .001$, $d = 0.63$).

Table 15. Means, Standard Deviations and Sample Sizes for Yearly Changes in CCTT-X Scores by Instructional Approach, Gender, and Grader Level for Data Set 3

Grade		SWH			Control			Total
		Male	Female	Mean	Male	Female	mean	
6	Mean	4.2	4.88	4.39	1.74	1.98	1.84	3.33
	SD	3.98	3.45	3.84	4.05	2.86	3.57	3.93
	N	111	41	152	61	47	108	260
7	Mean	4.39	4.08	4.24	2.22	2.26	2.23	3.37
	SD	3.42	4.73	4.09	3.18	3.32	3.22	3.86
	N	77	72	149	72	43	115	264
8	Mean	5.67	4.11	4.75	3.30	2.42	2.69	3.85
	SD	4.32	4.68	4.59	3.17	2.79	2.94	4.08
	N	104	150	254	60	136	196	450
Total	Mean	4.78	4.22	4.52	2.40	2.30	2.35	3.58
	SD	4.01	4.52	4.27	3.52	2.90	3.20	3.99
	N	22	263	555	193	226	419	974

Table 16. Three-Way ANOVA Summary Table for Data Set 3

Source	Sum of Squares	df	Mean Square	F	p	□2
Instructional Approach	1029.89	1	1029.89	70.41***	0.000	0.068
Gender	18.688	1	18.69	1.28	0.259	0.001
Grade Level	93.003	2	46.50	3.18*	0.042	0.007
Instructional Approach x Gender	2.038	1	2.04	0.14	0.709	0.000
Instructional Approach x Grade Level	1.751	2	9.38	0.64	0.527	0.001
Gender x Grade Level	110.603	2	55.30	3.78*	0.023	0.008
Instructional Approach x Gender x Grade Level	10.897	2	5.45	0.37	0.689	0.001
Error	14072.1	962	14.63			

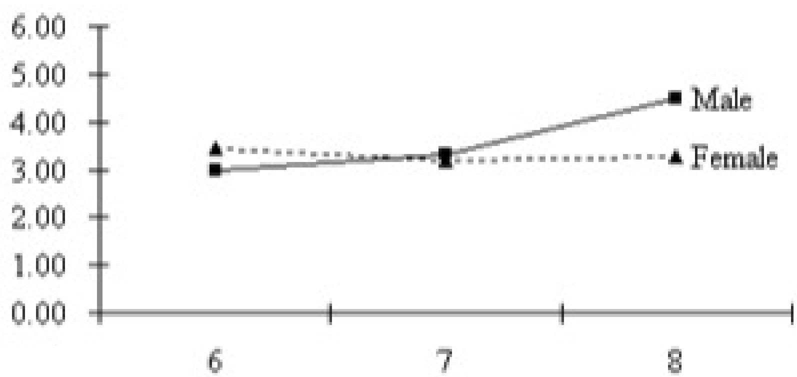

Figure 4. Gender by Grade Level Interaction for Yearly Gain Scores for Data Set 3

Effects of year of implementation, instructional method and gender on CCTT-X gain scores

Research Question 3 was focused mainly on comparisons of gain scores over years of implementation. Analyses were limited to Data Sets 4 and 5 in which gain scores were available for two successive years. Yearly CCTT-X gain scores were analyzed for each data set using a split-plot ANOVA design with one within-subject factor (Year of Implementation) and two between-subjects factors (Instruction Method and Gender).

Means, standard deviations, and sample sizes for CCTT-X scores for Data Set 4 at the beginning and end of each year of implementation and corresponding yearly gain scores are provided in Table 15. A plot of CCTT-X mean scores for the beginning and end of each year of implementation by instructional method is shown in Figure 5. The figure reveals that mean scores are essentially unchanged between the end of the first year of implementation and the start of the second year of implementation for the control group but increased over that period for the SWH group. The split-plot ANOVA results for yearly CCTT-X gain scores in Table reveal statistically significant effects for Instructional Approach ($F(1, 429)=35.84$, $p< .001$), Year of Implementation ($F(1, 429)=84.59$, $p< .001$) and the Instructional Approach by Year of Implementation interaction $F(1, 429)=69.41$, $p< .001$). The interaction, whose interpretation takes precedent over the main effects, is shown in Figure 6. Simple-effect tests for instructional method differences in gain scores were statistically significant for the first year of implementation ($F(1, 429)=79.43$, $p < .001$, $d = 0.66$), but not for the second ($F(1, 429)=1.40$, $p > .05$). That is, while overall gains for SWH group from the beginning of the first year to the end of the second year were higher than for the traditional instruction group, most of that gain occurred during the first year of implementation.

Table 17. Means, Standard Deviations and Sample Sizes for CCTT-X Scores by Year of Implementation, Instructional Approach, and Gender for Data Set 4

Grade		SWH			Control			Total
		Male	Female	Combined	Male	Female	Combined	
Pre-test	Mean	36.04	37.80	36.92	37.17	38.95	38.06	38.20
	SD	8.03	7.63	8.55	8.23	8.10	9.06	8.74
Post-test	Mean	41.09	41.15	41.12	40.47	39.61	40.04	41.84
	SD	7.78	7.23	8.00	8.77	7.98	8.39	8.73
Year I Gain Score	Mean	5.05	3.35	4.20	2.17	2.04	1.98	3.64
	SD	3.03	3.63	2.43	3.23	3.17	1.94	3.58
Pre-test	Mean	43.34	43.46	43.4	40.99	39.34	40.17	40.19
	SD	3.11	3.33	8.39	7.71	8.28	8.39	8.76
Post-test	Mean	45.89	44.79	45.34	42.45	41.97	42.21	42.11
	SD	8.78	8.09	8.25	8.71	8.78	8.97	8.75
Year II Gain Score	Mean	1.80	1.98	1.94	1.88	2.32	2.05	1.92
	SD	2.60	2.86	2.21	3.31	2.91	2.64	2.92
	N	110	117	227	103	103	206	433

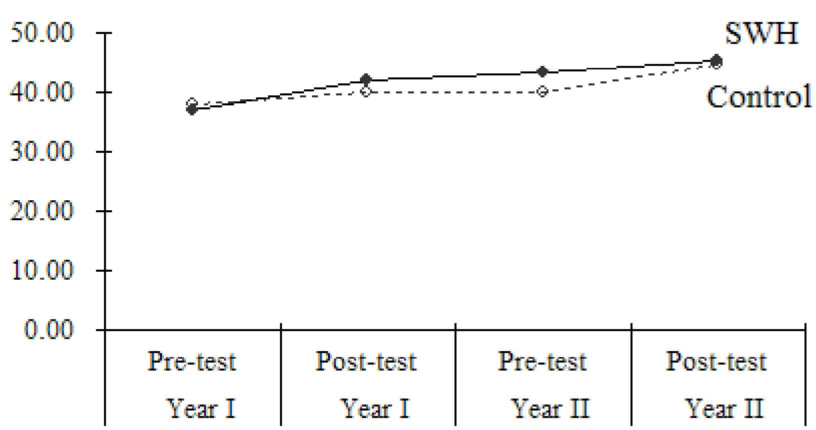

Figure 5. SWH and Traditional Instruction CCTT-X Mean Scores for Two Consecutive Years for Data Set 4

Table 18. Split-Plot ANOVA Summary Table for Data Set 4

Source	Sum of Squares	df	Mean Square	F	p
Between-subjects	5760.14	432			
Instructional Approach	444.10	1	444.10	35.84***	0.000
Gender	0.44	1	0.44	0.04	0.850
Instructional Approach x Gender	0.44	1	0.44	0.4	0.850
Within-group error	5315.16	429	12.39		
Within-subjects	4051.81	433			
Year of Implementation	587.47	1	587.47	84.59***	0.000
Year of Implementation x Instructional Approach	482.04	1	482.04	69.41***	0.000
Year of Implementation x Gender	2.82	1	2.82	0.41	0.525
Year of Implementation x Instructional Approach x Gender	0.10	1	0.10	0.02	0.904
Error (Years of Implementation within Subject)	2979.39	429	6.95		

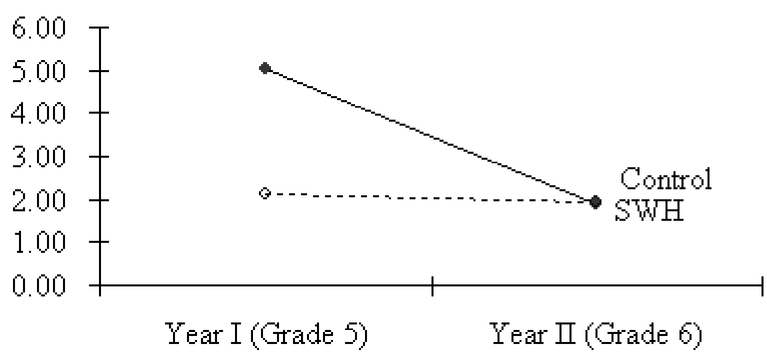

Figure 6. Instructional Approach by Year of Implementation Interaction for CCTT-X Gain Scores in Data Set 4

Means, standard deviations, and sample sizes for CCTT-X scores for Data Set 5 appear in Table . The plot of CCTT-X mean scores for the beginning and end of each year of implementation in Figure 7 reveals gains over that period for both the SWH and traditional groups. Consistent with results for the previous data set, statistically significant effects on yearly gain scores were found for Instructional Approach ($F(1, 507)=77.35$, $p< .001$), Year of Implementation ($F(1, 507)=1039.32$, $p< .001$)) and the Instructional Approach by Year of Implementation interaction ($F(1, 507)=44.35$, $p< .001$, see Table). The interaction, which is shown in Figure 8, reveals that CCTT gains were higher for the SWH group in the first year of implementation ($F(1, 507)=70.48$, $p < .001$, $d = 0.60$), but not for the second ($F(1, 507)=1.44$, $p > .05$). The overall gain in CCTT-X scores from the beginning of the first year to the end of the second year was still higher for the SWH group than for the traditional instruction group, but most of that gain occurred during the first year of implementation.

Table 19. Means, Standard Deviations and Sample Sizes for CCTT-X Scores by Year of Implementation, Instructional Approach, and Gender for Data Set 5

Grade		SWH			Control			Total
		Male	Female	Combined	Male	Female	Combined	
Pre-test	Mean	36.04	39.52	37.78	37.17	39.19	38.18	38.34
	SD	8.03	7.63	8.80	8.23	8.10	8.87	8.48
Post-test	Mean	41.09	44.55	42.82	40.47	40.11	40.29	41.45
	SD	7.78	7.23	8.35	8.77	7.98	8.67	8.44
Year I Gain Score	Mean	5.05	5.03	5.04	2.17	2.04	2.11	3.11
	SD	3.03	3.63	3.35	3.23	3.17	3.20	1.99
Pre-test	Mean	43.34	43.42	43.38	40.99	44.29	42.64	42.66
	SD	3.11	3.33	8.63	7.71	8.28	9.03	8.89
Post-test	Mean	45.89	44.65	45.27	42.45	46.73	44.59	44.65
	SD	8.78	8.09	8.85	8.71	8.78	9.05	8.85
Year II Gain Score	Mean	1.80	1.98	1.89	1.88	2.32	1.95	1.99
	SD	2.60	2.86	2.74	3.31	2.91	3.11	2.43
	N	126	133	259	123	129	252	511

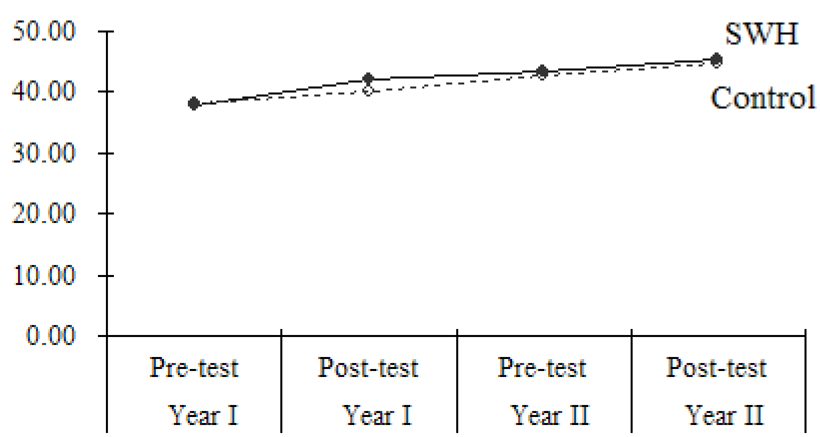

Figure 7. SWH and Traditional Instruction CCTT-X Mean Scores for Two Consecutive Years for Data Set 5

Table 20. Split-Plot ANOVA Summary Table for Data Set 5

Source	Sum of Squares	df	Mean Square	F	p
Between-subjects	2591.73	510			
Instructional Approach	342.56	1	342.56	77.35***	0.000
Gender	0.96	1	0.96	0.22	0.643
Instructional Approach x Gender	2.99	1	2.99	0.67	0.412
Within-group error	2245.23	507	4.43		
Within-subjects	10113.95	511			
Year of Implementation	6600.29	1	6600.29	1039.32***	0.000
Year of Implementation x Instructional Approach	281.67	1	281.67	44.35***	0.000
Year of Implementation x Gender	1.11	1	1.11	0.17	0.677
Year of Implementation x Instructional Approach x Gender	11.14	1	11.14	1.75	0.186
Error (Years of Implementation within Subject)	3219.76	507	6.35		

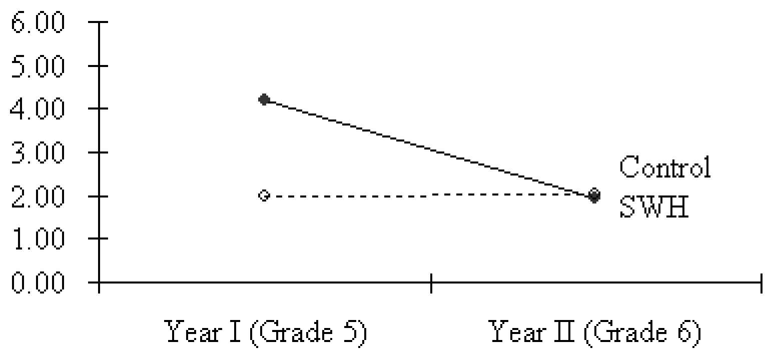

Figure 8. Instructional Approach by Year of Implementation Interaction for CCTT-X Gain Scores in Data Set 5

Summary

Results detailed in this chapter revealed that CCTT-X scores at each occasion of administration displayed high levels of internal consistency (r_{sh} ranged from .89 to .97) and yearly gain scores sufficiently reliable to evaluate group differences (i.e., gain score reliability coefficients varied from .47 to .81). Statistically significant improvements in critical thinking skills were found for each year of instruction for all data sets, but these improvements were higher on average in the SWH group (d-values ranged from 0.22 - 0.61, mdn = 0.53) than in the traditional group (d-values ranged from 0.13 - 0.33, mdn = 0.24). Analyses of yearly gain scores for all three data sets that represented a single year of implementation revealed statistically significant differences favoring SWH over traditional instruction with standardized mean differences (d-values) ranging from 0.58 to 0.65. Analyses for two of these same data sets uncovered statistically significant interactions between gender and grade level in which higher average gain scores were found for females at lower grade levels and for males at higher grade levels. Analyses from the two data sets that included two years of implementation revealed higher overall gains for SWH instruction than for traditional instruction but most of those gains were achieved during the first year of implementation.

CHAPTER FIVE
DISCUSSION OF RESULTS

The purpose of this study was to examine the effects of an immersion approach, the Science Writing Heuristic (SWH), to teach argument-based inquiry on student critical thinking skills. Gains in critical thinking skills, measured by the Cornell Critical Thinking-Form X (CCTT-X) were compared to students exposed to conventional approaches to science instruction. The study was conducted within elementary and secondary science classrooms where science was taught as part of each school's curriculum. Teachers were not asked to change the curriculum required by the school or district, but rather to shift from the traditional didactic orientations to science teaching to much more student-centered orientations. Specifically, teachers were asked to implement an argument-based inquiry approach that required them to focus on student learning by involving students in posing questions, generating claims using supporting evidence, and publicly negotiating meaning of the "big idea" of the topic, wherever possible.

All data for the current study were collected at either elementary and/or secondary school science classrooms in public schools. The participants took a critical thinking test at the beginning and end of the academic year. The five data sets examined allowed for exploration of effects between treatment and control groups both cross-sectionally and longitudinally over a two-year period.

Three research questions were addressed:

Research Question 1: To what extent do critical skills change from the beginning to the end of each academic year? (Data sets 1-5)

Research question 2: What are the main and interactive effects of instructional approach, gender, and grade level on CCTT-X gain scores across a single academic year? (Data sets 1-3)

Research Question 3: What are the main and interactive effects of year of implementation, instructional approach (SWH versus traditional), and gender on CCTT-X gain scores across two academic years? (Data sets 4-5)

Summary and analysis of results for the research questions.

Research Question 1: Changes in critical thinking skills from the beginning to the end of each academic year.

Statistically significant improvements in critical thinking skills were found in each year of instruction for all data sets, but these improvements were higher on average in the SWH group than in the traditional group. The general improvements in students achievement and critical thinking skills would be expected as they gain experience and meet challenges of advancing coursework. Ennis, Millman and Tomko (2005), for example, report "rough but gradual improvement (p.17)" in critical thinking over time with correlations between CCTT-X scores and age ranging from .3 and .39. Kuhn (1993, 1999) proposed three stages of developmental in critical thinking. Kuhn noted that an individual's thinking ability can grow with maturation but may be affected by developmental factors, coming from family, environment and education.

In conventional science classrooms, students frequently are not expected to work actively and collaborate, to think about concepts as much as memorize facts, or to develop and support a written statement or argument as they would with SWH instruction. While the acquisition of content typically requires choosing and applying a certain concept to a given situation, critical thinking goes beyond and requires evaluating, questioning, and synthesizing new information. Critical thinking, thus, is the basis from which students' scientific reasoning can emerge (Kuhn, 1993). Students have a natural curiosity to explore content beyond rote memorization of fact to a more complex higher-order thinking that requires advanced analysis and evaluation. This higher-order thinking can be enhanced when students' natural inquisitiveness is inspired and cultivated through an inquiry-based learning process (McNeill et al, 2006).

Prior research indicates that the inquiry strategy and writing process are effective because students must conceptually organize and structure their thoughts as well as their awareness of thinking processes (Langer and Applebee, 1987; Ackerman, 1993; Holliday, 1994; Rivard, 1994). Taylor and Sobota (1998) studied the effect of writing intervention in ten biology sections by testing students' critical thinking skills (as measured with Cornell Critical Thinking Test Z, CCTT-Z). The lack of any significant change in analysis, inference, or evaluation skills in the control group indicated that the traditional lab instruction used in the control courses did not help students develop critical thinking skills. As part of the writing process, students can shape their thoughts at the point of construction and continually analyze, review, and clarify meaning through the processes of drafting and revision by necessarily engaging and applying analysis and inference skills (Klein, 1999; Hand and Prain, 2002). The process of writing may be closely linked to critical thinking gains. Students who write also may experience a greater cognitive demand than non-writing students simply because the writing act required them to hypothesize, debate, and persuade (Rivard, 1994; Hand and Prain, 2002) rather than simply memorize (Rivard, 1994; Hand and Prain, 2002).

The Science Writing Heuristic (SWH) approach differs from the conventional approaches of teaching argument structure in several important ways by requiring students to collect data from their own investigations, make decisions about which of the data points will be used as evidence, and construct a logically connected explanation using these data points. Students are required to make decisions about what data are appropriate and provide reasoning on how the data points form evidence to support their claim. Each of these procedures and epistemological viewpoints are useful in constructing knowledge and facilitating science learning.

Various studies have supported the benefits to teachers and students who experienced the SWH approach. Hand and colleagues (2007), for example, conducted a case study of a teacher who attempted to implement the SWH approach in a science classroom over a two-year period. Using dialogical analysis and the Reform Teaching Observation Protocol (RTOP), the researchers found that the teacher successfully moved from teaching science as the traditional approach to an argument-based inquiry approach that allowed the focus and direction of class to be determined by students' voices. The teacher also shifted from heavily using factual recall questioning to using various questioning patterns that encouraged students to contribute in the classroom. Students were reported to have had ownership in their learning and engaged in argument structure by investigating their own questions, claims and evidence. Hand et al also reported that students who learned with the SWH approach were more successful than students who learned with standard methods in the lecture and laboratory courses. Wallace and Hand (2004) found that by engaging in a series of scaffolded writing tasks that centered on the development of ideas through argument, students' improved their performance on conceptual questions at the completion of a unit on cells. These improvements occurred primarily not through engagement in writing but rather through talking, discussing, and arguing to evaluate claims supported by evidence. Gunel et al (2009) also found that the SWH approach helped students develop scientific argument strategies. Students who engaged in the SWH approach were provided opportunities for small-group and whole-class discussion and negotiation. These were crucial components of the SWH approach. Their results revealed that negotiation with peers could promote reflective thinking about patterns of analysis. Taken as a whole, these studies reveal that the SWH approach can be an effective tool to promote and scaffold scientific argument within science classrooms, and these processes in turn are logically related to improvements in critical thinking (Cavagnetto & Hand, 2008; Hand et al, 2008; Keys et al, 1999; Martin & Hand, 2010). The present research provided empirical verification of such a relationship.

Research question 2: Effects of instructional approach, gender, and grade level on CCTT-X gain scores across a single academic year?

ANOVA results for CCTT-X gain scores revealed a statistically significant main effect for Instructional Method across all three data sets examined and a Gender by Grade level interaction in two of the three data sets. Students who received SWH instruction improved more in critical thinking over the school year than did students who received traditional instruction, and this difference was more than half of a standard deviation higher (d values from .58 to .65) for the SWH group in all instances. Possible reasons for the superior performance for SWH over traditional instruction were discussed in the previous section focusing mainly on the fundamental requirement of students needing to apply critical thinking to a greater extent under SWH than under traditional instruction. A key new finding for the current research question was the consistent Gender by Grade Level interaction observed within two data sets. This interaction revealed a reversal of gender differences in yearly CCTT-X gain scores between grades with females achieving greater average gains than males at the 6^{th} grade level and males achieving greater gains than females at the 7^{th} and 8^{th} grade levels.

Studies examining the effects of gender on critical thinking conducted at the college level have yielded conflicting findings. No differences are reported in most instances (e.g., Angeli & Valanides, 2009; Kalman, 2002), and differences favoring either female (e.g., Higgins et al, 2004) or male (e.g., Down, 2008; King, 1990; Rogoff, 2003) in other instances. For research specifically related to the CCTT-X, Ennis et al. (2005) cited 14 studies that compared differences between male and female elementary and secondary students, ranging from 5^{th} through 12^{th} grade with sample sizes from 28 to 1,126. Three of these studies showed gender differences with girls outperforming boys in relation to judging the credibility of sources and observations. The remaining studies showed no differences between boys and girls in critical thinking. In examining components of higher thinking, Costs, Terracciano and McCrae (2001) note that girls tend to act in a more consensus-making manner and are more agreeable compared to boys, who are less open to ideas and more assertive in presenting information. Yenilmez and Sungur (2005) reported that boys have higher scores than girls on proportional, probabilistic and combinational reasoning, whereas girls have higher scores on controlling variables and correlational reasoning.

When gender was examined as a main effect within the present three data sets, no trustworthy differences were found in critical thinking gain scores between the sexes.

Only after taking grade level into account did such differences emerge. The present data do not provide a mechanism for determining why this interaction occurred, but one can speculate that factors such as differences in educational experiences and sex stereotyping may play a role. King (1999), for example, notes that some educational systems encourage male students more so than female to become better critical thinkers, and this practice might become increasingly prevalent as students get older. Some researchers also have noted that cultural differences might play a role (see, e.g., Bataineh & Zghoul, 2006; Belenky, 1987) with gender differences in Eastern countries tending to be more pronounced than in Western countries.

Another possible reason for the Gender by Grade interaction may relate to course selection. The data collected here for 6th grade, for example, came from a general science classroom, whereas the data for 8th grade came mostly from physics and chemistry classrooms These higher level science course may attract stronger male students on average that would more heterogeneous general science classes.

Research Question 3: Effects of year of implementation, instructional method and gender on CCTT-X gain scores.

Split-plot ANOVA results for the two data sets examined revealed statistically significant effects on yearly gain scores for Instructional approach (treatment versus control), Year of Implementation (one year versus two years) and the Instructional Approach by Year of Implementation interaction. As was the case in the analyses already discussed, average gains were again higher under SWH than under traditional instruction, and performance improved on average over time. However, the Instructional Approach by Year of Implementation interaction revealed that CCTT gains were higher for SWH than traditional instruction in the first year of but not for the second year. That is, the overall gains in CCTT-X scores from the beginning of the first year to the end of the second year was still higher for the SWH group than for the traditional instruction group, but most of that gain occurred during the first year of implementation.

One possible explanation for the greater rise in performance for the first year with the SWH condition is a "Hawthorne" or novelty effect. Participants in SWH treatment group may have intentionally improved or modified their learning behavior to meet the goal of argument-based inquiry that led to higher gains in critical thinking skills, because they knew that they are being studied or simply were experiencing something beyond the status quo. These effects, in turn, may have led to fast improvement in the beginning or first year of the intervention. In a review of educational research, Clark and Sugrue (1991), for example, noted that uncontrolled novelty effects cause on average 30% of a standard deviation rise, which can decay to small levels after 8 weeks; specifically, 50% of a SD for up to 4 weeks; 30% of SD for 5–8 weeks; and 20% of SD for > 8 weeks.

A second possible explanation for this result may be a plateau effect similar to that sometimes observed in standardized test score means that diminish as grade level increases (Center for Education Policy, 2006; Chudowsky & Chudowsky, 2009). Similarly, in some assessment-based accountability systems, students' initial gains in test results during the first few years are impressive, but difficult to sustain over time (Chudowsky & Chudowsky, 2009; Korsky et al, 2013). Proficient trends also can follow a wide variety of trajectories, including some plateaus, some steady increases, and fluctuating "zigzag" patterns that still move in an overall upward direction. This might suggest some critical points, or capstone stages, during the development of certain ability, all of which were not captured in the present analyses.

One could speculate that, in the present case. SWH no longer has too much effect on critical thinking, or that the CCTT-X is less sensitive to improvements beyond the first year levels due to some student topping out on the score scale. A common finding in second language studies, for example, (see, e.g. Yuan et al, 2009) is that students, after a certain period of time, might reach a saturation point and stop paying attention. Students hit a plateau in their language learning, stabilize at a set point, and seemingly cannot progress further. The saturation point occurs in second language learners who have been in the target culture for six months to a year. Eventually, the student can move beyond the saturation point on his/her own and begin to absorb more language, but may that take some time. In another context, Quitadamo et al (2009) reported that when Organic Chemistry I and II were taught in successive terms by the same instructor, students showed a 6 percentile gain for the first terms but only a 4 percentile point gain for the second term. A similar trend might have occurred here with critical thinking.

Performance gains over time also depend on the nature of the assessment, the development procedure, and testing format (Chudowsky & Chudowsky, 2009). Some skills are more easily and quickly taught and assessed. For example, math computation skills tested with multiple-choice items is relatively easy to improve, compared to complex problem-solving situation that may require more cumulative, long-term instruction. Accordingly, tests that focus on lower-order skills might be more susceptible to short-term gains that would tests of higher-order thinking.

In pulling these various perspectives together in the present research, one could speculate that students' greater gains in Year I than in Year II may have occurred because initial exposure to SWH instruction opened a new window for students in thinking about and learning science. By the end of the first year, the effect may reflect a more saturated level of impact. While the same students continue to be exposed to SWH approach in the second year, their critical thinking skills might improve but not as dramatically as the first year. Similarly, compared to students with one year of SWH exposure, students with two years of exposure have a higher starting point in critical thinking skills, as measured by the pre-test, with more limited improvement and measurement sensitivity possible.

Implications

An important implication of the present findings is that students can develop critical thinking skills within a fairly short nine-month period of time and that the SWH approach may play an essential role in that process. These consequences are consistent with findings from other studies of critical thinking. For example, Yeh (1998), who compared traditional and explicit instruction on critical thinking in a different academic setting, found that students' critical thinking skills improved within one year. Wertsch (2008) also concluded that implementing different instructional strategies, such as project-based learning improved students' critical thinking and their academic performance.

Requiring students to construct and critique science knowledge through the process of inquiry also is aligned with the ideas put forward by Ford and Forman (2006) who suggested that students need opportunities to gain a "grasp of practices" (p. 3) in science. By this, they refer to the social and material aspects of science highlighting the dialogical aspects of science construction as well as the content of science. Importantly, Ford and Forman emphasized the need for students to be engaged with both the construction and the critique of science concepts.

Expanding on this idea in a recent review of the argumentation literature, Cavagnetto (2010) argued that there are three major argument intervention approaches used within science classrooms: teaching the structure of argument, emphasizing the interaction of science and society, and an immersion approach. In discussing these different approaches with respect to the opportunities students have to develop and critique perspectives, Ford and Forman (2006) highlighted several issues with the teaching the structure of argument and the interaction of science and society approaches. For Ford (2008), teaching of an argument approach fails to involve students in the critique of scientific practice and, thus, prevents them from really grasping the practices of science. This approach focuses on the explicit teaching of the structure of argument as a skill to be learned. Ford further argued that the interaction of science and society approaches tend to move beyond science in that students will engage in political, social, economic, and cultural issues related to these approaches. Thus, he is uncertain whether the students will be able to clearly understand the practices of argument being advocated because of the constraints of the broad range of issues advocated in such approaches. By contrast, the Immersion approach to learning science argumentation is based on the concept that students need to be actively engaged in the process of argumentation as a means to learn about science argument (Hand et al, 2008); that is, they are required to both construct and critique science knowledge as a critical function of engaging in scientific inquiry. This approach parallels the Immersion approach advocated for learning critical thinking due to its focus on the ideas of argumentation as opposed to the skills of argument.

The results of the present study reinforce the idea that students can benefit from student-centered argument-based inquiry and be exposed to it as early and often as possible, if not explicitly taught critical thinking skills. Teachers also need to create learning environments where students can negotiate their big ides with peers by using claims and evidence to construct understanding about the natural world through essential scientific practices (Hand, 2008). The SWH approach creates coherent opportunities for students to engage in multiple aspects of scientific inquiry while constructing knowledge and working toward better explanations through argumentation (Cavagnetto & Hand, 2012). Therefore, students with SWH instructions have opportunities to critique arguments based on the credibility of information that is used as evidence and stay open-minded to the source that allows for critique.

Critical thinking also may have an important implication for transfer of knowledge and application of problem solving skills to novel situations (Brown, 1990). Evidence suggests that complex cognitive skills can be systematically taught (Moss & Koziol, 1991; Nickerson, 1988) and remain an overarching goal of education (NRC, 2000). In this regard, the methods used in current study might be extended to other disciplines to further examine the impact of transferability to different contexts. While Halpern (1998) stated that critical thinking ability can be taught, he also highlighted a number of challenges in teaching critical thinking and evaluating the effectiveness of such instruction. Differences between pre-test and post-test scores can reflect stability, or losses in performance but often not the sources of changes in scores. Cognitive skills often improve gradually with practice with effects of instruction taking time to become apparent and are also affected by learners' natural maturation and experiences obtained from other sources. As a result, it seems advantageous to adopt a comprehensive multidimensional and multidisciplinary approach to understanding development and changes in critical thinking skills over time and transferability of critical thinking skills to novel situations and new domains.

Future research

Fruitful future extensions of the present research into critical thinking might go beyond gender, age, and gender level effects and include examination of other co-variables such as prior knowledge (Tusi, 2002; Rogoff, 2003), academic achievement, (King et al, 1990), and instructor implementation (Kuhn, 1993). Because critical thinking assessments have been shown to have a strong relationship to academic achievement, it is important to consider prior achievement in evaluating the effectiveness of instructional interventions. Previous research also indicates that teaching style can influence certain aspects of student learning, but we know little about such effects on critical thinking per se.

Development is another important aspect of critical thinking that requires further clarification. Many previous studies of critical thinking were snapshots at specific college levels, and the present study of elementary and secondary classrooms was extended longitudinally over a maximum of two academic years. Future research might be focused on a longer time frame to explore critical points of plateaus and development and transferability across subject matter areas (see e.g., Boscolo & Mason, 2001).

Construction of critical-thinking friendly learning environment is another topic for further investigation. The SWH approach is designed to engage students in justifying their claims relate to the real world by using evidence and critical reasoning skills, considering explanations and alternative explanations, and building upon and critiquing different perspectives with their classmates. Other techniques and sequencing of material may provide additional insights into how critical thinking might be facilitated. Ideally, students would learn foundational skills of critical thinking at very young ages, and be able to hone these skills as they progress through the post-secondary education by adopting the best strategies for cognitive development.

Limitations

To interpret the results for this study in a proper way, its limitations should be noted and addressed in future investigations. First, the present results came from participants who were not randomly sampled from a clearly defined target population and as such may have limited generalizability to groups that deviant from the present samples demographically. Second, the study followed students for a maximum of two academic years and therefore does not provide information about possible changes in critical thinking beyond that period. Third, critical thinking was assessed using only the CCTT-X, which provides only one operationalization of critical thinking skills. Fourth, the research design included only instructional method, gender, grade level, and year of implementation as independent variables thereby omitting other potentially important factors such as prior achievement, previous instructional experiences, type of science course taken, teacher style, and cultural influences. Finally, the study focused on the outcomes more so that the specific processes by which thinking skills might have improved.

REFERENCES

Abbott, C. & Wilks, S. (2005). Developing an effective classroom climate for higher order thinking. In S. Wilks (Ed.), *Designing a thinking curriculum* (pp. 1-11). Camberwell: ACER Press.

Abd-El-Khalick, F., BouJaoude, S, Duschl, R, Lederman, N. G., Mamlok-Naaman, R, Hofstein, A, Niaz, M, Treagust, D, & Tuan, H.-L. (2004). Inquiry in science education: International perspectives. *Science Education*, 88(3), 397-419.

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, Dai. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, *78*(4), 1102–1134.

Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical-thinking skills in further education students. *Instructional Science*, 29, 1-32.

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom□s taxonomy of educational objectives.* New York: Longman.

Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: performance on ill-defined issues. *Learning and Instruction*, 19, 322–334.

Bailin, S. (2002). Critical thinking and science education. *Science & Education*, *11*(4), 361–375.

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies*, *31*(3), 285–302.

Bataineh, R. F., & Zghoul, L. H. (2006) Jordanian TEFL graduate students' use of critical thinking skills (as Measured by the Cornell Critical Thinking Test, Level Z), *International Journal of Bilingual Education and Bilingualism*, *9*(1), 33-50.

Bandura, A. (1986). Social foundations of thought and action: a social cognitive theory. Englewood Cliffs, N.J.: Prentice-Hall.

Baron, J.B. (1987). Evaluating Thinking Skills in the Classroom. In Baron, J.B. & Sternberg, R.J. (Eds.), Teaching Thinking Skills: Theory and Practice (pp. 221-247). New York: W.H. Freeman & Company.

Becker, A. (2006). A review of writing model research based on cognitive processes. In Becker, A. 62 Homing, A. (Eds.), Revisions: History, theory and practice, 25-49. West Lafayette, IN: Parlor Press.

Belenky, M.F. (1986). Women's Ways of Knowing: The Development of Self, Voice and Mind. New York: Basic Books.

Bee, H. & Boyd, D. (2004). *The developing child*. Boston: Pearson.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*(5), 765-793.

Berland, L.K. & Reiser, B. (2009). Making sense of argumentation and explanation. *Science Education, 93*, 26-55.

Boscolo, P., & Mason, L. (2001). Writing to learn, writing to transfer. In P. TynjaÅNlaÅN, L. Mason, & K.Lonka (Eds.), *Writing as a learning tool. Integrating theory and practice* (pp. 83–104). Dordrecht: Kluwer Academic Publishers.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24,* 61-100.

Brown, A. L. (1990). Domain specific principles affect learning and transfer in children. *Cognitive Science*, 14, 107–133.

Bybee, R. Taylor, J., Gardner, A., Van Scotter, P., Carlson Powell, J., Westbrook, A., and Landes, N. (2006). *The BSCS 5-E instructional model: Origins and effectiveness*. A Report Prepared for the Office of Science Education, National Institutes of Health. Colorado Springs, CO.

Cavagnetto, A. R. (2010). Argument to foster scientific literacy: a review of argument interventions in K-12 science contexts. *Review of Educational Research*, 80(3), 336-371.

Cavagnetto, A. R., & Hand, B. (2011). The importance of embedding argument within science classrooms. In K. Myint (Ed.), *Perspectives on Scientific Argumentation: Theory, Practice and Research.* Dordrecht, The Netherlands: Springer.

Chiapetta, E.L. (2008). Historical Development of Teaching Science as Inquiry. In J. Luft, R. Bell, and J. Gess-Newsome (Eds.), Science as inquiry in the secondary setting (pp. 21-30). Arlington, VA; National Science Teachers Association.

Chudowsky, N., & Chudowsky, V. (2009). Is there a plateau effect in test scores? Washington, DC: Center on Education Policy.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge Falmer.

Crawford, B. A. (2000). Embracing the essence of inquiry: New roles for science teachers. Journal of Research in Science Teaching, 37(9), 916-937.

Dam, G. T., & Volman, M. (2004). Critical thinking as a citizenship competence: teaching strategies. *Learning and Instruction, 14*, 359–379.

Downs, C. (2008). Rethinking critical thinking. *Higher Education Review, 41*(1), 60-79.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. Science Education, 84(3), 287-312.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education 32*, 268–291.

Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*(1), 39-72.

Erduran, S. (2008). Methodological Foundations in the Study of Argumentation in Science Classrooms. In S. Erudran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education. Perspectives from Classroom-Based Research.* Dordrecht: Springer

Erduran, S., Simon, S., & Osborne, J. (2004). Tapping into Argumentation: Developments in the Application of Toulmin's Argument Pattern for Studying Science Discourse. *Science Education, 88*, 915-933.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44-48.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). New York: Freeman.

Ennis, R. H. (1989). Critical thinking and subject specificity: clarification and needed research. *Educational Researcher*, 18, 4-10.

Ennis, R. H. (1995). *Critical Thinking*. Prentice-Hall, New Jersey.

Ennis, R. H., Millman, J., & Tomka,T. N. (2005). *Manual: Cornell Critical Thinking Tests*. Pacific Grove, CA: Midwest.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of toulmin's argument pattern for studying science discourse. *Science Education*, 88(6), 915-933.

Educational Testing Service. (1998). *Manual: Critical Thinking Assessment*. Princeton, NJ: Riverside.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA: The California Academic Press.

Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relation to critical thinking skill. *Informal Logic, 20*(1), 61–84.

Facione, P (2007). Critical thinking: What it is and why it counts. California: Insight Assessment.

Finocchiaro, M. (2005), 'Fallacies and the Evaluation of Reasoning', American Philosophical Quarterly 18, 13–22.

Fisher, R. (2001). *Critical thinking: an introduction*. Cambridge: University Press.

Fisher, R. (2003). *Teaching thinking: philosophical enquiry in the classroom*. New York, London: Continuum.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911.

Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. *Review of Research in Education*, 30(1), 1–32.

Galbraith, D. (1999). Writing as a knowledge-constituting process. In T. Torrance & D. Galbraith (Eds.), Knowing what to write: conceptual processes in text production (pp.139-150). Amsterdam: Amsterdam University Press.

Galbraith, David; Ford, Sheila; Walker, Gillian; Ford, Jessica (2005) The contribution of different components of working memory to planning in writing. L1 – Educational Studies in Language and Literature 15, 113-145.

Galbraith, David; Hallam, Jenny; Olive, Thierry; Le Bigot, Nathalie (2009) The role of different components of working memory in writing. In Taatgen, Niels; van Rijn, Hedderik; Schomaker, Lambert; Nerbonne, John (eds.) Proceedings of the 31st Annual Conference of the Cognitive Science Society. Amsterdam: Cognitive Science Society, 3028-3033.

Galbraith, David; Torrance, Mark; Hallam, Jenny (2006) Effects of writing on conceptual coherence. Proceedings of the 28th Annual Conference of the Cognitive Science Society,1340-1345.

Gellin, A. (2003). The effect of undergraduate student involvement on critical thinking: A meta-analysis of the literature 1991–2000. *Journal of College Student Development*, *44*(6), 746–762.

Geier, R., Blumenfeld, P., Marx, R., Krajcik, J., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. Journal of Research in Science Teaching, 45(8), 922-939.

Gelman, S. A., & Markman, E. M. (1986).Categories and induction in young children. *Cognition*, 23, 183–209.

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, *7*(1), 22-30.

Grandy, R. & Duschl, R.A. (2007). Reconsidering the character and role of inquiry in school science: Analysis of a conference. *Science & Education*, 16, 141-166.

Hand, B. (2008). Introducing the science writing heuristic approach. In B. Hand (Ed.), *Science inquiry, argument and language: A case for the science writing heuristic*. Rotterdam, the Netherlands: Sense.

Hand, B. (2012). The Importance of Embedding Argument Within Science Classrooms. In *Perspectives on Scientific Argumentation: Theory, Practice and Research*. New York, NY: Springer.

Hand, B., & Prain, V. (1998). Writing as a Learning Tool in Science: Lessons Learnt and Future Agendas. In *Second International Handbook of Science Education*. New York, NY: Springer.

Hand, B., Wallace, C. W., & Yang, E.-M. (2004). Using a Science Writing Heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: quantitative and qualitative aspects. *International Journal of Science Education*, 26(2), 131-149.

Halonen, J. S. (1995). Demystifying critical thinking. *Teaching of Psychology, 22*(1), 75–81.

Halpern, D. F. (1996). *Thought and knowledge: an introduction to critical thinking*. Mahwah, N.J.: L. Erlbaum Associates.

Halpern, D. F. (1998).Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449–455.

Halpern, D. F. (2001) Assessing the effectiveness of critical thinking instruction. *The Journal of General Education*, *50*(4), 270–286.

Hendricks, C.C. (2001). Teaching Causal Reasoning through Cognitive Apprenticeship: What are Results from Situated Learning? *The Journal of Educational Research,* 94(5), 302-311.

Hernstein, R.J., Nickerson, R.S., de Sanchez, M., & Swets, J.A. (1986). Teaching Thinking Skills. *American Psychologist*, 41(11), 1279-1289.

Heyman, G. D. (2008). Children's critical thinking when learning from others. *Current Directions in Psychological Science, 17*(5), 344–347.

Heyman, G. D., & Legare, C. H. (2005).Children's evaluation of sources of information about traits. *Developmental Psychology, 41*(4), 636–647.

Higgins, S., Baumfield, V., Lin, M., Moseley, D., Butterworth, M., Downey, G., Gregson, G., Oberski, I., Rockett, M., & Thacker, D. (2004). *Thinking skills approaches to effective teaching and learning: what is the evidence for impact on learners? (research Evidence in Education Library).* London: EPPI Centre, Social Science Research Unit, Institute of Education.

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

Halliday, M. A. K., & Martin, J. R. (1994). *Writing science: Literacy and discursive power.* Pittsburgh, PA: University of Pittsburgh Press.

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science,* 17(9), 757–758.

Jimenez-Aleixandre, M. P., & Erduran, S. (2008). Argumentation in science education: An overview. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 3 – 28). Dordrecht, The Netherlands: Springer.

Johnson, D. W., & Johnson, R. T. (1993). Creative and critical thinking. Through academic controversy. *American Behavior Scientist*, 37, 40–53.

Kalman, C. S. (2002). Developing critical thinking in undergraduate courses: a philosophical approach. *Science and Education*, 11, 83-94.

Kennedy, M., Fisher, M. B., & Ennis, R. H. (1991). Critical thinking: literature review and needed research. In L. Idol, & B. Fly Jones (Eds.), Educational values and cognitive instruction: Implications for reform (pp. 11-40). Hillsdale, NJ: Erlbaum.

Keys, C. W., Hand, B., Prain V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching,* 36(10), 1065-1084.

King, P. M., Wood, P. K., & Mines, R. A. (1990). Critical thinking among college and graduate students. *The Review of Higher Education,* 13, 167 – 186.

Knight, G. (2005). *Critical, creative, reflective and logical thinking in the NEMP assessments (A National Education Monitoring Project Probe Study Report).* New Zealand: Ministry of Education.

Knight, G. (2006). *Student responses to critical thinking tasks in the NEMP assessments (A National Education Monitoring Project Probe Study Report).* New Zealand: Ministry of Education.

Knudson, R. (1992). Analysis of argumentative writing at two grade levels. Journal of Educational Research, 85, 169-179.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*(6), 1261–1277.

Ku, K. Y. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity, 4*(2009), 70–76.

Kuhn, D., & Dean, D. (2004). A bridge between cognitive psychology and educational practice. *Theory into Practice, 43*(4), 268–273.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, 77(3), 319-337.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28, 16-25.

Kuhn, D. (2010), Teaching and learning science as argument. *Science Education*, 94: 810–824.

Kuhn, D., & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance. *Cognitive Development, 13*, 227–247.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction*, 15(3), 287–315.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child development*, 74(5), 1245-1260.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lai, E. R. (2011). *Critical thinking: a literature review*. Retrieved from http://pearsonassessments.com/hai/images/tmrs/CriticalThinkingReviewFIN AL.pdf

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice, 32*(3), 131–137.

Lipman, M. (1988). Critical thinking—What can it be? *Educational Leadership, 46*(1), 38–43.

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development, 73*(4), 1073–1084.

Martin, A. M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom: A longitudinal case study. *Research in Science Education,* 39(1), 17-38.

Martinez, M. E. (2006). What is metacognition? *Phi Delta Kappan, 87*(9), 696–699.

McGregor, D. (2007). *Developing thinking; developing learning: a guide to thinking skills in education*. Maidenhead, Berkshire, England: Open University Press.

McGuinness, C. (1999). *From thinking skills to thinking classrooms.* Nottingham: DfEE Priced Publications.

McInerny, D. M. & McInerny, V. (2002). *Educational Psychology: constructing learning.* Frenchs Forest, NSW: Prentice Hall.

McPeck, J.E. (1981). Critical Thinking and Education. New York: St. Martin's Press.

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19*(4), 10–12.

McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, 93(2), 233-268.

McNeill, K. L. (2011). Elementary students' views of explanation, argumentation and evidence and abilities to construct arguments over the school year. *Journal of Research in Science Teaching*, 48(7), 793-823.

McNeill, K. L., Lizotte, D. J, Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, 15(2), 153-191.

McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education*, 94(2), 203-229.

McPeck, H. (1990). Critical thinking and subject specificity: a reply to Ennis. *Educational Researcher*, 19, 10-12.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry based science instruction - what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47, 474-496.

Moss, P. A., & Koziol, S. M. (1991).Investigating the validity of a locally developed critical thinking test.*Educational Measurement: Issues and Practice*, *10*(3), 17–22.

National Research Council. (2000). Inquiry and the national science education standards: A guide for teaching and learning. Washington, D.C., National Academy Press. CHAPTER 2.

National Research Council. (2007). Taking science to school: Leaning and teaching science in grades K-8. Washington, DC: National Academies Press. CHAPTER 9.

National Research Council. (2008). Ready, set, science: Putting research to work in K-8 science classrooms. Washington, D.C.: National Academies Press. CHAPTER 5.

National Research Council (NRC) (2013). *National Science Education Standards*. Washington, DC: National Academy Press.

Newton, P., Driver. R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, 21(5), 553-576.

Nelson, C. E. (1994). Critical thinking and collaborative learning. *New Directions for Teaching and Learning*, *1994*(59), 45–58.

Nickerson, R.S. (1987). Why Teach Thinking? In Baron, J.B. & Sternberg, R.J. (Eds.), *Teaching Thinking Skills: Theory and Practice* (pp. 27-37). New York: W.H. Freeman & Company.

Nickerson, R. S. (1988). On improving thinking through instruction. *Review of Research in Education*, *15*(1988–1989), 3–57.

Norris, S. P. (1989). Can we test validly for critical thinking? *Educational Researcher*, *18*(9), 21-–26.

Norris, S.P., Phillips, L.M., & Osborne, J.F. (2008). Scientific Inquiry: The Place of Interpretation and Argumentation. In J. Luft, R. Bell, and J. Gess-Newsome (Eds.), *Science as inquiry in the secondary setting* (pp. 87-98). Arlington, VA; National Science Teachers Association.

Norton-Meier, L., Hand, B., Hockenberry, L., & Wise, K. (2008). *Questions, claims, and evidence: The important place of argument in children's science writing.* Portsmouth, NH: Heinemann.

O'Hare, L. O., & McGuinness, C. (2009). Measuring critical thinking, intelligence, and academic performance in psychology undergraduates. *The Irish Journal of Psychology*, *30*(3–4), 123–131.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in science classrooms. *Journal of Research in Science Teaching*, 41, 994–1020.

Perkins, D.N. (1987). Thinking Frames: An Integrative Perspective on Teaching Cognitive Skills. In Baron, J.B. & Sternberg, R.J. (Eds.), *Teaching Thinking Skills: Theory and Practice* (pp. 41-61). New York: W.H. Freeman & Company.

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research,* 42(3), 237–249.

Prain, V., & Hand, B. (1996). Writing for learning in secondary science: rethiking practices . *Teaching and Teacher Education*, 12 (6), 609-626.

Prawat, R. S. (1991). The value of ideas: The immersion approach to the development of thinking. *Educational Researcher*, 20(2), 3-10

Quitadamo, I. J., & Kurtz, M. J. (2007). Learning to improve: Using writing to increase critical thinking performance in general education biology. *Life Sciences Education,* 6,140–154.

Reece, G. (2002). Critical thinking and transferability: A review of the literature Retrieved June, 2006 from http://www.library.american.edu/Help/reseach/lit_review/critical_thinking.pdf

Reece, G. (2007). Critical thinking and cognitive transfer: Implications for the development of online information literacy tutorials. *Research Strategies*, 20, 482-493.

Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., & Holowchak, M. (1993). Reasoning in conversation. *Cognition and Instruction*, 11(3-4), 347-364.

Rivard, L. P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching, 31,* 969-983.

Rogoff, B. (2003). *The cultural nature of human development*. Oxford, England: Oxford University Press.

Royalty, J. (1995). The Generalizability of Critical Thinking: Paranormal Beliefs versus Statistical Reasoning. *The Journal of Genetic Psychology*, 156(4), 477-487.

Sampson, V., & Clark, D. B. (2008a). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92(3), 447–472.

Schraw, G., Crippen, K. J., & Hartley, K. (2006).Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36 (1-2), 111–139.

Siegel, H. (1992), Educating Reason: Rationality, Critical Thinking, and Education, Routledge, New York.

Silva, E. (2008).*Measuring Skills for the 21st Century* [Report]. Washington, DC: Education Sector. Retrieved from http://www.educationsector.org/usr_doc/MeasuringSkills.pdf.

Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement, and improvement* National Institute of Education. Retrieved from http://eric.ed.gov/PDFS/ED272882.pdf.

Stewart, J., Cartier, J. L., & Passmore, C. M. (2005). Developing understanding through model-based inquiry. In M. S. Donovan and J. D. Bransford (Eds.), *How students learn* (pp. 515-565). Washington, D.C.; National Research Council.

Taggart, G., Ridley, K., Rudd, P., & Benefield, P. (2005). *Thinking skills in the early years: a literature review*. Slough: National Foundation for Educational Research.

Taylor, K.L., and S.J. Sobota. 1998. Writing in biology: An integration of disciplines. The *American Biology Teacher,* 60 (5): 350–53.

Thayer-Bacon, B. J. (2000). *Transforming critical thinking: Thinking constructively*. New York, NY: Teachers College Press.

Tindal, G., & Nolet, V. (1995). Curriculum-based measurement in middle and high schools: Critical thinking skills in content areas. *Focus on Exceptional Children*, 27(7), 1–22.

Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education, 40*, 185–200.

Tsui L. (2002). Fostering critical thinking through effective pedagogy: evidence from four institutional case studies. *Research in Higher Education*. *73*(6):740–763.

Van Gelder, T. (2005).Teaching critical thinking: Some lessons from cognitive science. *College Teaching*, 53(1), 41–48.

Wertsch, J. V. (2008). From social interaction to higher psychological processes: a clarification and application of Vygotsky‘s theory. Human Development, 51(1), 66-79.

Whimbey, A., &Lochhead, J. (1986). *Problem solving and comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wilen, W.W. & Phillips, J.A. (1995). Teaching Critical Thinking: A Metacognitive Approach. *Social Education, 59*(3), 135-138.

Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator*, 8–19.

Wilson, C. D., Taylor, J., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on student knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, 47(3), 276-301.

Windschitl, M., Thompson, J. & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, *92*(5), 941-967.

Yeh, S. S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher*, 30(9), 12-17.

Yeh, S.S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle school students. *Research in the Teaching of English*, 33(1), 49–83.

Yore, L. D., & Treagust, D. F. (2006). Current Realities and future possibilities: Language and science literacy-empowering research and informing instruction. *International Journal of Science Education*, 28, 291-314.

Zembal-Saul, C. (2009). Learning to teach elementary school science as argument. *Science Education*, 93(4), 687-719.