Spring 2009

The impact of embedding multiple modes of representation on student construction of chemistry knowledge

Mark Andrew McDermott
University of Iowa

THE IMPACT OF EMBEDDING MULTIPLE MODES OF REPRESENTATION ON STUDENT CONSTRUCTION OF CHEMISTRY KNOWLEDGE

by

Mark Andrew McDermott

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Science Education
in the Graduate College of
The University of Iowa

May 2009

Thesis Supervisor: Professor Brian Hand

ABSTRACT

This study was designed to examine the impact of embedding multiple modes of representing science information on student conceptual understanding in science. Multiple representations refer to utilizing charts, graphs, diagrams, and other types of representations to communicate scientific information. This study investigated the impact of encouraging students to embed or integrate the multiple modes with text in end of unit writing-to-learn activities. A quasi-experimental design with four separate sites consisting of intact chemistry classes taught by different teachers at each site was utilized. At each site, approximately half of the classes were designated treatment classes and students in these classes participated in activities designed to encourage strategies to embed multiple modes within text in student writing. The control classes did not participate in these activities. All classes participated in identical end of unit writing tasks in which they were required to use at least one mode other than text, followed by identical end of unit assessments. This progression was then repeated for a second consecutive unit of study. Analysis of quantitative data indicated that in several cases, treatment classes significantly outperformed control classes both on measures of embeddedness in writing and on end of unit assessment measures. In addition, analysis at the level of individual students indicated significant positive correlations in many cases between measures of student embeddedness in writing and student performance on end of unit assessments. Three factors emerged as critical in increasing the likelihood of benefit for students from these types of activities. First, the level of teacher implementation and emphasis on the embeddedness lessons was linked to the possibility of conceptual benefit. Secondly, students participating in two consecutive lessons appeared to receive greater benefit during the second unit, inferring a cumulative benefit. Finally, differential impact of the degree of embeddedness on student performance was noted based on student's level of science ability prior to the initiation of study procedures.

Abstract Approved: ______________________________
Thesis Supervisor

____
Title and Department

____
Date

THE IMPACT OF EMBEDDING MULTIPLE MODES OF REPRESENTATION ON STUDENT CONSTRUCTION OF CHEMISTRY KNOWLEDGE

by

Mark Andrew McDermott

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Science Education
in the Graduate College of
The University of Iowa

May 2009

Thesis Supervisor: Professor Brian Hand

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

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

Mark Andrew McDermott

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Science Education at the May 2009 graduation.

Thesis Committee: ______________________________
Brian Hand, Thesis Supervisor

____
Alicia Alonzo

____
Jeffrey Denburg

____
James Maxey

____
Soon-hye Park

To Lindsay

ACKNOWLEDGMENTS

I would like to sincerely thank the members of my committee, Dr. Alicia Alonzo, Dr. Jeffrey Denburg, Dr. James Maxey, and Dr. Soon-hye Park for their help with this study and throughout my graduate studies. I would like to especially thank Dr. Brian Hand for his continual support and guidance throughout my graduate career. I have always felt challenged, yet encouraged, and the wide range of learning experiences you have provided for me and guided me through have been invaluable in my own professional development.

I would like to thank the teachers who participated in this study. I am well aware of the demands placed on practicing teachers and the added demands participation in this study have placed on you. I am grateful for all of your extra work in helping me accomplish my goal of finishing this study. I also thank the students who were members of the participating classes for their efforts in completing the activities for this study.

I would also like to thank all of the teachers and students I have been fortunate to share schools and classrooms with throughout my secondary teaching career. I have always gained more from my interactions with you than from any other experiences I have had and I appreciate all the great people I have been able to work with.

Finally, I would like to thank my family. Thank you Ron, Jan, Brad and Andrew for all the help in making my graduate studies possible. To my sisters Becky and Jane and my brother-in-law Kenny, you have been great models of teachers for me but even better role models as people. To my parents, you have always been able to expect the best but help me through the worst. I am continually reminded of how lucky I am to have such great parents. To Mallory and Jace, my greatest hope is that all the time that went into this project will mean much more time with you two in the future. And to Lindsay, nothing I ever do is possible without your constant love and support. I hope I am always worthy of that and I hope you always remember I love you most of all.

ABSTRACT

This study was designed to examine the impact of embedding multiple modes of representing science information on student conceptual understanding in science. Multiple representations refer to utilizing charts, graphs, diagrams, and other types of representations to communicate scientific information. This study investigated the impact of encouraging students to embed or integrate the multiple modes with text in end of unit writing-to-learn activities. A quasi-experimental design with four separate sites consisting of intact chemistry classes taught by different teachers at each site was utilized. At each site, approximately half of the classes were designated treatment classes and students in these classes participated in activities designed to encourage strategies to embed multiple modes within text in student writing. The control classes did not participate in these activities. All classes participated in identical end of unit writing tasks in which they were required to use at least one mode other than text, followed by identical end of unit assessments. This progression was then repeated for a second consecutive unit of study. Analysis of quantitative data indicated that in several cases, treatment classes significantly outperformed control classes both on measures of embeddedness in writing and on end of unit assessment measures. In addition, analysis at the level of individual students indicated significant positive correlations in many cases between measures of student embeddedness in writing and student performance on end of unit assessments. Three factors emerged as critical in increasing the likelihood of benefit for students from these types of activities. First, the level of teacher implementation and emphasis on the embeddedness lessons was linked to the possibility of conceptual benefit. Secondly, students participating in two consecutive lessons appeared to receive greater benefit during the second unit, inferring a cumulative benefit. Finally, differential impact of the degree of embeddedness on student performance was noted based on students' level of science ability prior to the initiation of study procedures.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER ONE

GENERAL OVERVIEW AND PURPOSE OF STUDY

Introduction

Much science education research is centered on the search for effective pedagogical techniques, tools, and interventions based on a solid theoretical background, with the ultimate goal of informing classroom practices that will lead to greater student learning. Often, this search involves a cyclical process in which research supported theoretical ideas lead to new questions, that when clarified, initiate a new cycle of research. The research proposed here is the product of a similar cyclical path for the author initiated through explorations of research supported practices in science classrooms, including his own, and questions that have arisen as a result of exploring these practices. The ultimate intent is to help clarify the characteristics of writing-to-learn activities that utilize multiple modes of representing science information in science classrooms in order to maximize the ability of these activities to improve student conceptual growth.

Science Literacy

Much debate in science education surrounds the question of what it means for students to be "scientifically literate" and what teachers must do to encourage science literacy (Yore & Treagust, 2006; Cavagnetto, 2006). While no universally accepted definition of science literacy exists, several theoretical and curricular goals related to this idea have been suggested in the literature, including improving student awareness of the nature of science, improving student critical thinking, improving student understanding of core science concepts, improving student use of science process skills, as well as others (Yore & Treagust, 2006; National Research Council, 1996). Historically, proposed definitions of scientific literacy have tended to emphasize one or more of these specific goals. Early models focused on the links between science and society (Hurd, 1958). Later, an emphasis on developing effective citizens through appropriate habits of the mind (Rutherford & Ahlgren, 1990) or through a proper understanding of science concepts, the nature of science, and science relationships to technology (Miller, 1996) was promoted. Bybee (1995) agreed in part with these assertions by proposing different aspects of science literacy, including functional (science vocabulary and terms), conceptual (big ideas and connections) and multidimensional (links to technology and society). In addition, a fourth aspect, procedural science literacy was offered that involved student understanding of the processes and methods of science (Bybee, 1995).

Norris and Phillips (2003) offered a more contemporary view of science literacy. Their definition includes two components, a “derived sense” and a “fundamental sense”. In this view, the derived sense involves the aspects mentioned previously, including student understanding of science terms, science concepts, science processes, science links to technology, and the history and nature of science. However, Norris and Phillips add the fundamental sense that involves developing students’ ability to read, interpret, and critique science arguments (Cavagnetto, 2006). Importantly, this view asserts this fundamental sense is perhaps most critical in classrooms, and that students must be given the opportunity to evaluate the strengths and weaknesses of scientific argument. Arguing a similar position, Ford (2008) proposes that science education must focus more on students understanding how to evaluate the accountability of claims, and encourage having students make “scientific sense of content” rather than making their “own sense of content.”

Yore and Treagust (2006) drawing heavily on Norris and Phillip’s position, summarize a definition of science literacy with two statements (p.293):

1. The meaningful understanding of knowledge about the big ideas or unifying concepts / themes of science like the nature of science, scientific

inquiry, and major conceptual themes in the biological, earth-space, and physical sciences.

2. A literacy component that stresses the cognitive abilities, critical thinking, habits of mind, and information communication technologies (ITC) to understand the big ideas in science, to inform and persuade others about these ideas, and to participate more fully in the public debate about STSE issues.

Teaching toward these goals would necessarily involve a focus on the "known" ideas, processes, and connections in science, on the ways to link these to everyday life, and on ways to critically evaluate these, as well as focus on student development of their own scientific understanding. Yore and Treagust (2006) also suggest that the degree to which each of these areas are emphasized, as well as the standards students are held to when assessing these areas can be manipulated based on the particular students involved. One of the most important questions arising from this discussion is what teaching techniques will have the greatest chance at success in accomplishing this goal of science literacy.

Writing-to-Learn

One particular strategy that has been suggested as a means of helping students achieve this expanded sense of scientific literacy is the use of writing. Writing has been promoted both as a communicative tool, where students can display their science understandings, and as a generative tool, where students can develop conceptual understanding through the use of writing tasks (Klein, 1999; Prain & Hand, 1996; Prain, 2006). In terms of the aforementioned definition of science literacy, writing would seem an appropriate vehicle to achieve many of the stated objectives. Further, the relationship between writing and reading has been described as constructive, both in terms of the author's and reader's ability to build knowledge (Yore & Treagust, 2006; Keys, 1999;

Ruddell & Unrau, 1994). This implies that the use of writing can encourage other beneficial activities in a classroom, many of which would enable the critical review of science knowledge, both from peers and outside sources, that Ford (2008) and others have called for.

A significant amount of discussion in research literature has centered on the appropriate way to incorporate writing into science classrooms to achieve these hoped for benefits. While many questions remain unanswered, there is emerging support for the use of writing from both theoretical perspectives and from practical research studies indicating realized benefits.

Cognitive Models Associated with Writing-to-Learn

Theoretical work in the area of writing in science classrooms has centered on the cognitive action present when students write. Emig (1977) promoted an early model when she posited that writing led to more student awareness of relationships and connections among ideas due to its constant feedback. The student writer dealt with both knowledge of content and composition and interacted with the developing text in a way that promoted learning (Wallace, Hand, & Prain, 2004). Following Emig's initial ideas, cognitive models developed describing writing as a problem solving activity. Flower and Hayes (1980) proposed a goal-oriented model in which writer's balance goals of the task with goals dealing with content. Bereiter and Scardamalia (1987) followed with a problem solving model still recognized as a dominant model in education today (Klein, 1999). They proposed "knowledge-telling" writing involves relating to another audience recall of information already known. This contrasts with "knowledge-transforming" writing in which a dynamic process takes place when information the student creates in text cognitively reorganizes information previously known. This process involves interplay between three phases of writing (planning, generating, and revising) and two factors (task environment and writer's long-term memory). The cognitive content space of the writer interacts with cognitive rhetorical space as goals shape text production and information held in the writer's content space is transformed (Galbraith & Torrance, 1999).

Galbraith (1999) proposed a model that agreed with some aspects of Bereiter and Scardamalia's, but also accounted for the unpredictable nature of writing and for the possibility of knowledge generation, in addition to transformation. In Galbraith's model, tacit knowledge becomes explicit due to interaction between the writer's neural representations of information and constraints on the writing, including linguistic, rhetorical, and task factors. Galbraith used the terms *writer's disposition* to describe cognitive content knowledge and *linguistic network* to refer to rhetorical knowledge. The activation of a "dispositional dialectic" through consideration of already produced text, and feedback loops through cognitive networks potentially leads to writing encouraging the constitution of new knowledge for the writer (Galbraith, 1999). Galbraith's major assertion is that under the appropriate writing conditions, the activation of the dispositional dialectic will lead to knowledge generation independent of new input in regard to rhetorical planning.

Types of Writing Encouraging Learning

The cognitive models associated with writing to learn predict that learning as a result from writing is a possible outcome, not a guaranteed result. The practical issue that develops in light of these theoretical models is how to design writing tasks for authentic classroom situations that will increase the likelihood of students realizing conceptual gains. Much research is currently dealing with this question.

Science education research has shown that one way to encourage knowledge transformation or generation is through diverse or non-traditional writing tasks calling for student use of everyday language in communicating about science concepts (Prain, 2006). These types of writing tasks use creative formats and allow students to communicate their scientific understanding in their own words. The promotion of this pedagogical method stems from a theoretical claim that using these types of writing instead of more traditional genres employed in science classrooms allows students to connect science concepts they engage in with their past experiences (Rowell, 1997). This connection then allows students to develop personalized meanings related to the more technical terms and concepts as described in the vocabulary of science (Prain, 2006). Finally, these activities have been offered as ways to encourage students to participate in more practical endeavors by arguing that they encourage knowledge consolidation through methods that are actual ways adults, scientists and non-scientists alike, write about science (Prain & Hand, 1996).

Non-traditional writing-to-learn techniques also enjoy empirical support from a wide variety of research studies. Positive effects for students have been demonstrated in terms of motivation (Hildebrand, 2004), communication, argumentation, justification, clarification, knowledge display, and improvement in learning outcomes (Prain, 2006; Hildebrand 2004; Gunel, Hand & Gunez, 2006). However, in spite of the growing body of research support for the benefit of these techniques in classrooms, Prain (2006) admits there is still considerable debate about the goals, rationale, and importantly, the specifics of the pedagogical practices associated with their use.

Multiple Modes of Representing Information in Writing-to-Learn

The use of multiple modes of representing information (multimodal representations) is a developing area of study in education (Prain, 2006; Gunel et. al, 2006; Eilam & Poyas, 2008). Research in this area related to writing-to-learn has generally centered on student use of different modes of representing information such as diagrams, pictures, math equations, tables, or graphs with text, as opposed to using only text (a unimodal representation). Supporters of this strategy argue that multi-modal representations are a typical occurrence in the field of science (Prain, 2006) and in education (Eilam & Poyas, 2008). With emerging technologies, they are increasingly used by students in their "everyday" life experiences (Gunel et. al, 2006). In addition, research has indicated that since multi-modal representations are built on unimodal representations, (Bernsen, 1993) student movement between unimodal and multi-modal representations as they develop written products requires a translation process (Pineda & Garza, 2000). This idea is linked to a general theoretical position underpinning the use of non-traditional writing tasks in science that argue translation processes between different types of "language" can benefit student learning. In fact, some researchers have gone as far as saying that learning involving the movement between different modes is predicated upon and encourages enhancement of multiple characteristics including the efficient retrieval of information, the integration of information from different sources, and the application of information to new contexts (Eilam & Poyas, 2008). It is the production of multi-modal tasks with these kinds of benefits within the framework of writing-to-learn strategies that is the research focus driving this proposed study.

Purpose of the Study

Building on the aforementioned research suggesting a benefit of using multiple modes of representing information in writing-to-learn activities, a previous study was initiated in the secondary chemistry classroom of the author. The intent of the study was to determine if requiring the use of multi-modal representations in writing tasks impacted student conceptual understanding. Emerging from that study was a realization that most students, when asked to provide multiple modes of representing information simply added modes other than text at the end of their written product, with little regard for how the multiple modes were related. Student writing typically lacked integration of the multimodal representations with each other or with the text. This initial finding led to a consideration of what has been termed for purposes of this study the "degree of embeddedness" in a multimodal writing task.

Degree of embeddedness refers to the level of integration among the different modes representing science information in student writing. If the translation benefit of students moving between modes as suggested by Bernsen (1993) and Pineda and Garza (2000) is to be realized, it is likely that students must strive to connect the different modes and the text. By drawing student attention to this characteristic of their writing, greater translation and cognitive action as a result of the writing task may be realized, leading ideally to greater conceptual development through the writing task.

Two issues appear critical in promoting a greater degree of embeddedness. First, students must be made aware of techniques typically employed in sources they are familiar with to accomplish embeddedness. Textbooks, websites, and magazine articles, among other sources, provide examples of these techniques. Once this awareness is promoted, students must then be encouraged to employ similar techniques in their own writing. To this end, a pilot study was initiated in the chemistry classroom of the author in which students were both encouraged to identify techniques used to embed multiple modes of representing science information in familiar sources and then to use these techniques. Analysis of the findings from this pilot study indicated that there were positive correlations between the degree of embeddedness in student writing and their performance on end of unit evaluations. These correlations were realized for two different topics when assessment took place immediately following the writing task, as well as for a third topic in which assessment was delayed for two weeks following the writing.

While the results from this pilot study were positive, several limitations were present. First, the correlations detected are indicative of a relationship between degree of embeddedness in writing and student performance on end of unit tests, but they do not indicate causality. Second, evidence from one classroom with one teacher is not likely to encourage widespread adoption of this pedagogical technique. More powerful evidence would involve showing differential student achievement between groups of students who participate in "embeddedness encouraging activities" and groups who do not, as well as evidence from a variety of classroom settings.

The overall intent of the study is to add to the developing research base relative to writing-to-learn activities utilizing multiple modes of representation in science by advancing understanding of the impact embedding multiple modes with text can have on student learning. Specifically, the study is designed to explore whether or not more direct evidence about the effectiveness of encouraging students to utilize embedding strategies for multi-modal writing tasks can be attained in a variety of settings. It is a step forward in the determining whether this instructional approach, rather than some underlying variable related to both embeddedness and student performance, is responsible for conceptual benefit.

Research Questions

The research proposed here will take place in four different chemistry classroom settings. In all settings, some students will participate in a lesson designed to identify strategies that can be used to integrate multiple modes of representing science information with text in a cohesive way. This lesson will culminate in the production of a class-generated checklist that can be used to assess a written product in reference to its degree of embeddedness. All students in a particular setting will then participate in identical diverse writing-to-learn tasks in which students who received embeddedness instruction will self assess their writing for embeddedness and the other students will not. All students will take identical end of unit assessments. If possible, the writing and assessment cycle will be repeated for a second consecutive unit of study in each classroom. The research questions guiding this study are the following:

1. Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to a greater degree of embeddedness in student writing?
2. Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to greater conceptual understanding as measured by end of unit assessments?
3. Can correlations between degree of embeddedness in writing and student performance be detected in a variety of classroom settings?

Dissertation Overview

Chapter two will describe the theoretical framework upon which the study presented here is set. To do this, a discussion on what it means to be scientifically literate will be followed by consideration of what an appropriate conception of the nature of science is. Analysis of the research related to these issues will be combined with research describing how students learn science to provide an overview for an effective science learning environment and rationale for the use of writing tasks as a part of this environment. Discussion will then proceed to clarify the cognitive models upon which writing-to-learn tasks are based, as well as why non-traditional writing tasks may be a more effective way to accomplish the goals of science teaching and incorporate the positive aspects of the cognitive models. Finally, specific issues related to the use of multi-modal writing tasks and how they relate to the previous issues will generate the impetus for this study.

Chapter three will provide a brief rationale for the use of a quasi-experimental, quantitative approach in this study based on pragmatic and practical concerns. Description of the student populations, as well as the teachers and schools involved in this study will be followed by a complete description of the procedures and activities the students participated in.

Chapter four will center on the data analysis of this study. Student data pertaining to embeddedness and to performance on the end of unit assessments will be presented, along with analysis of relationships between these factors. Data will be analyzed both at the level of treatment groups and at the level of individual students for each of the four settings. This will culminate in a summation of the overall findings.

Chapter five will present a discussion emerging from the data analysis. This discussion will focus on how the data addresses the research questions, as well as overall ideas about the relationship between multi-modal writing-to-learn tasks and student conceptual performance.

Chapter six will expand on limitations of the study, as well as implications from the study. In this discussion, these two issues will be combined to suggest appropriate as well as necessary further research in this area.

CHAPTER TWO

THEORETICAL FRAMEWORK OF STUDY

Any pedagogical intervention that aims to increase conceptual understanding in science must be grounded in knowledge of the theoretical positions that indicate its potential effectiveness. This chapter explores the theoretical framework that underpins the use of multi-modal writing tasks as a classroom tool to improve science understanding. All activity associated with conceptual understanding in science must be developed with consideration of several factors impacting learning in the science classroom. These factors include recognizing what science literacy involves, applying an appropriate view of the nature of science, and understanding how students learn science. Discussion of these factors will provide the initial focus of this chapter.

The theoretical discussion will continue with a focus on how writing tasks, particularly those that call for the use of multiple modes of representing science information, are a potential source for encouraging science learning that integrates current understanding about the important factors impacting learning in a science classroom previously mentioned. To develop the rationale for the use of these tasks, the cognitive models supporting the use of writing-to-learn tasks will be discussed, along with the practical aspects of utilizing non-traditional writing tasks in science that have been shown in empirical studies to promote the beneficial cognitive action hypothesized by the models. Finally, the use of multi-modal writing tasks will be proposed as a particular avenue for realizing student conceptual gains built on interaction of all the theoretical factors presented here.

Factors Influencing Learning in the Science Classroom

Science Literacy

One fundamental question in any science classroom and for any science teacher is what the overarching goal of science education should be. Teachers in their classrooms, teacher education practitioners developing methods classes and professional development, and science education researchers determining what questions are appropriate to explore must all contend with the dilemma of first determining what the overriding objective of science education is. If this question is not addressed, it is unlikely that consistent, meaningful learning will take place for students. Often termed *science literacy*, the ideas about what this overarching goal is have evolved historically. Accompanying the evolution of the concept of science literacy have been changes in the suggested pedagogical strategies proposed for the development of student learning leading to science literacy. The fact that some researchers have gone as far as to argue that "scientific literacy" is really nothing more than a "slogan" that is useful only as a rallying point to encourage improved science education practices (Bybee, 1997) necessitates a careful review of what scientific literacy has meant historically and how it influences effective science pedagogy.

Historical Definitions of "Science Literacy"

The term "scientific literacy" was first used in discussing science learning in the late 1950's (Hurd, 1958; McCurdy, 1958; DeBoer, 2000). The idea, however, that there should be a universally accepted goal of what science education should look like in the United States began to develop near the beginning of the 20th century (DeBoer, 2000). As DeBoer (2000) points out when describing particular aspects of the evolution of the term science literacy, the major dilemma in defining the term has typically been the question of how to balance the importance of understanding the concepts of science and how scientific thought operates with the practical aspects of how science can be used in everyday life. As the following discussion will illustrate, the balance has historically shifted back and forth between these competing issues.

The initial push in America for including science in the curriculum came from practicing scientists in the 1800's (DeBoer, 2000). The prevailing attitude of the time in terms of education was that the goal of educating children could be most appropriately realized by a focus on humanities education. Thus, as scientists argued for the inclusion of science in school curriculum, they were faced with the dilemma of promoting both the applicability of science understanding to useful, practical problems, as well as the potential for science learning to promote the more nebulous idea of higher level thinking and intellectual development (DeBoer, 2000). The strategy employed was to highlight three benefits. First, science learning could help train inductive thinkers who would be able to use observation of nature as a way to make claims about how it worked. Secondly, science learning could promote an "attitude of independence" among citizens about decisions related to the advancement of technology taking place in the country and world as the turn of the century approached. Finally, science learning that accomplished the first two goals was promoted as a way to develop citizens who could function more effectively in a democratic society.

The rationale utilized to promote inclusion of science in the curriculum was echoed as conceptions of what science literacy meant began to develop. Beginning with Eliot (1898), the usefulness of science education as a way to promote "effective power in action" in the larger scheme of the educational enterprise was a consistent theme with many educational theorists. The progressive era that was ushered in the early 20th century by the likes of John Dewey increasingly combined this action idea with calls to promote science education that both highlighted relevance to the lives of students studying the science and recognized the reality of the social implications of scientific work (DeBoer, 2000). As is still the case today, criticism emerged to this focus on making science education relevant and socially centered that claimed the subject matter and content of science were being sacrificed and neglected. This led to the assertion of a broader idea of science literacy summarized in a document from 1932 entitled *A Program for Teaching Science.* This document maintained the goal of science education should include aspects related to improving the individual ability to utilize science, improving intellectual ability in order to function in a democratic society, and improve the understanding of science as a cultural force and as a means to appropriately describe nature (National Society for the Study of Education, 1932).

Two major issues related to World War II and its aftermath impacted the perception of science in general and specifically what was expected of schools in terms of teaching science. On one hand, the end of the war and the methods used to achieve that end led some to believe that science was taking humankind in a negative direction, destined to make the world more violent (DeBoer, 2000). This perception led some educators to insist one role of science education should be to help familiarize students with the work scientists do to provide a balanced view of how scientific progress impacts the world. In addition, the growing realization that science learning was potentially linked to national defense was a motivating factor for some to call for making science an area of heavy emphasis in schools. One way to encourage this emphasis, particularly recognized within the political realm, was to develop grass roots public support for science education by linking it with national security:

> The security and prosperity of the United States depend today, as never before, upon the rapid extension of scientific knowledge. So important, in fact, has this extension become to our country that it may reasonably be said to be a major factor in national survival. (President's Scientific Research Board (1947), p.3, as quoted in DeBoer (2000)).

As the 20th century reached its mid-point, there was a growing realization that science education must confront two equally important goals as the world became more technologically advanced and as the rate of technological advancements increased. First, instruction in science must provide a large output of technically competent and well trained students who could participate in the work of science. Second, instruction in science should prepare all students to make intelligent decisions about the appropriate uses and applications of science (NSSE, 1960). In order to accomplish these goals, students needed not only to understand the concepts of science, but also appreciate how scientists worked and functioned in developing the technology that came from these concepts. It was in this spirit in the late 1950's that the terms "science literacy" and "scientifically literate citizenry" began to be used (DeBoer, 2000).

Hurd (1958) became one of the first writers to use the term "science literacy" publicly when he discussed his views on balancing the demands of understanding science concepts and being able to utilize the practical aspects of science. Hurd asserted that four "forces" were intertwined in the endeavor of science, those being scientific, social, economic, and cultural. While unwilling to do more than define science literacy in vague terms, Hurd did point out that the goal of curriculum designers and practitioners of science education should be the development of tasks in which students realize the intellectual or conceptual aspects of science, while simultaneously experiencing the procedural aspects of science as a discovery. In doing this, Hurd felt students would develop an awareness of the "spirit of scientific discovery" (Hurd, 1958). While agreeing with Hurd's call for an awareness of the cultural aspects of science, McCurdy (1958) warned that science education of the time was in need of a greater emphasis on science concepts. McCurdy worried that too great a focus on technology was impeding the sound understanding of science content necessary to provide a general scientific literacy that accurately conveyed the way science understanding impacted the interpretation of daily events. This concern foreshadowed events of the next decade.

During the 1960's the pendulum of what was important in promoting a scientifically literate citizenry swung toward an emphasis on content knowledge. Academic work became the focus, and coursework was often designed by scientists instead of educators. This coursework tended to emphasize more abstract descriptions of the natural world based on science concepts, as opposed to practical applications of science or instances of science phenomenon in everyday life (DeBoer, 2000). It was theorized that this increase in rigor would attract more academically gifted students, thus developing an emerging scientific work force of the highest caliber. Those who were not academically able to handle the increased rigor, it was thought, would at least develop a "sympathetic view" towards science which was another goal of the science community (DeBoer, 2000).

Inevitably, the focus on academic rigor in defining what was critical for attaining science literacy met criticism that large segments of the population were neglected, particularly due to the lack of concern for interest and development of students. To counteract this, science in a social context was again promoted as the goal of science education (NSTA, 1971). Some researchers, including Hurd (1970) went as far as to claim that in order to promote general science literacy, science had to be taught within a social context. This drive culminated in the promotion of a science-technology-society (STS) curriculum that explicitly called for the central position of the social context of science by suggesting social issues should be the unifying theme of science education and all curricula should be designed and organized around particular societal issues. In doing so, proponents argued not only would the science education of students become more personal as instruction dealt with topics relevant to everyday life, but it would also promote adequate awareness of and understanding about science issues for citizens to be informed decision makers (Gallagher, 1971; Hofstein & Yager, 1982).

Inherent in this call for social-topic organized science instruction was a drive toward social action. It was expected that the increased awareness of the relationship between science and the technology that it utilizes along with the societal issues in which it is manifested would lead to an increased willingness to take action to deal with concerns in ones local area that were of a scientific nature (DeBoer, 2000). As had been the case in the past, criticism developed in response to these initiatives, particular from those who felt that "basic" science understanding was being neglected in deference to emphasizing societal issues (Kromhout & Good, 1983). As the STS proponents and their critics battled over the relative importance of basing science learning on relevant issues or basing it on a solid understanding of scientific facts, three publications emerged that dramatically influenced practicing educators' views of what the characteristics of scientific literacy should be.

The first noteworthy report came in 1983 from the National Commission on Excellence in Education and was entitled *A Nation at Risk: The Imperative for Educational Reform*. The main premise of this report was that a domino effect had emerged in which lower academic standards for American students had led to lower achievement, which in turn was leading to a lessening economic position for the United States (NCEE, 1983). The second major report came from the American Association for the Advancement of Science in 1989 and was entitled *Science for All Americans*. The main assertion of this document was that it was critical for the United States to reform science education due to the fact the country had been slow to respond to the increased need to train children to function in a more technologically and scientifically complex society. The third document, published in 1996 was the *National Science Education Standards* (NSES) from the National Academy of Sciences. This comprehensive report attempted to provide a structure for science education reform by outlining national standards in several areas including content, teacher actions, and science teacher education.

The upshot of these documents was three main goals: an increased national awareness of the issues relating to science education, a broadening definition of what it meant to be scientifically literate, and the development of specific standards intended to clarify a framework for attaining scientific literacy. These goals are crystallized in the NSES and the definition of science literacy in that publication included most historically promoted characteristics of science literacy. A scientifically literate person according to the NSES would be able to ask and answer questions that develop from their own curiosity, describe and discuss natural phenomenon, use this knowledge of natural phenomenon to predict natural occurrences, read about science in typical popular press articles and intelligently discuss the information presented, evaluate decisions on the local as well as national level based on scientific principles, critique findings based on an understanding of the methods used in comparison to appropriate scientific methodology, and participate in and evaluate arguments based on or purportedly based on scientific evidence (NSES, 1996).

The emergence of the aforementioned documents along with their accompanying broad definition of science literacy was intertwined with another push for reform in science education during the 1990's. Driving this reform was a stance that a main reason for defining science literacy in the first place is to aid in the development of pedagogical tools that will bring about the intended characteristics in students. Promotion of the goals of science literacy with the intent of determining practical routes to achieve them began to dominate the literature. Miller (1996) succinctly argued three main areas must be considered in the classroom: the understanding of basic science concepts and vocabulary, the nature of science, and the relationship between science, technology, and society. Miller's first area was split into two components, functional literacy intent on understanding vocabulary and conceptual literacy concerned with understanding the main concepts in science by Bybee (1995). Miller's final two goals were combined and termed multidimensional literacy. Bybee (1995) defined this as understanding the link between science and technology along with a historical recognition of the nature of science. Finally, Bybee (1995) called for procedural literacy that focused on the methods and techniques that are used to develop scientific understanding. DeBoer (2000) provided a comprehensive summary of both the historical development of science literacy and the characteristics currently considered critical in establishing scientifically literate students.

His summation included nine major themes to be presented by teachers for students and in turn demonstrated by students to indicate scientific literacy: science as a cultural force in the modern world, science as preparation for the world of work, direct applications of science to the everyday world, science as a way to develop informed citizens, science as a particular way to examine the natural world, an understanding of reports and discussion about science in popular media, appreciating the aesthetic appeal of science, preparation of citizens who are sympathetic to science, and identification of the nature and importance of technology and its relationship to science.

The current ideas about science literacy and the recent call for reform of science teaching practices promotes a conception of science literacy as a standards-based way for all people to develop appropriate conceptual abilities and habits of mind that allow them to understand scientific "big ideas", apply these to local and national issues involving science, technology, and society, and discuss these ideas with others in a way that could persuade them to take action (Hand, Prain, & Yore, 2001). In the next section, the link between these ideas about science literacy and language is explored as one way to help develop sound teaching practices.

Language and Science Literacy

The idea of science literacy as presented by the NSES and as summarized by many including DeBoer (2000) has certainly had criticism leveled against it. Shamos (1995) has argued that science literacy still focuses too much on traditional science conceptual knowledge and presents an unrealistic picture of what students can actually do and understand. Shamos calls for a more attainable goal of "scientific awareness" in which students are helped to develop an appreciation for how science works without necessarily requiring students to develop a thorough understanding of science concepts. Others, typified by Mayer (1997) have called for a greater social action commitment to be affirmed as a critical aspect of what it means to develop science literacy. However, one of the most crucial questions that must be dealt with in terms of making science literacy more than a theoretical idea and turning it into a driving force for practical application in the classroom is how should the literacy component of science literacy be defined and utilized.

Most historical as well as current definitions of science literacy at least implicitly call for scientifically literate students having the ability to communicate, argue, debate, and discuss science concepts. However, few definitions of science literacy speak directly to the appropriate use of language in defining and realizing science literacy (Yore, Bisanz, & Hand, 2003). Lemke (1990) was an early proponent of the need to situate science literacy in the arena of language as a way to talk and argue about science. Others, including Kelly & Green (1998) have recognized the importance of framing attainment of science literacy in a social context calling for appropriate communication about science. However, as Reveles, Cordova, and Kelly (2004) point out, students are not automatically able to participate in scientific communication just because they understand or have been presented science concepts. If science literacy is to include scientific communication, specific instruction in the literacy component is necessary.

Since the late 1970's, efforts have been made to promote a combined view of science learning with language (Yore et. al, 2003). Central to this call is a realization that learning is a multi-sensory experience involving some combination of oral discourse, reading comprehension, and writing within a social setting (Holliday, Yore, & Alvermann, 1994). Lemke (1998) asserts that any time students "do science", talk about, read about, or write about science, they are utilizing many different types of communication skills including but not limited to verbal, mathematical, graphical or visual, and motor expression. In addition, scientists themselves rely on multiple forms of communication and expression in conducting and communicating scientific endeavors. Therefore, as Yore et al. (2003) point out, language is involved in both the "doing" of science and the communicating about science. Language can help students develop and construct meaning about science concepts, as well as provide the medium for students to communicate what methods and techniques they have used to develop scientific understanding. Further, the evaluation, argumentation, and discussion about science concepts including relevant social, political, and cultural issues is predicated on the use of language. Language, then, is connected to every facet of science literacy (Yore et al., 2003).

Norris and Phillips (2003) developed a multi-faceted definition of science literacy based on this connection between language and science literacy. Their definition developed through consideration of both a classical analysis of language and an analysis of the philosophy of science (Yore et al., 2003). Norris and Phillips clarify two "senses" of science literacy, the fundamental sense and the derived sense. The fundamental sense deals with the ability to use language discourse tools such as writing, reading, and speaking to discuss and explore science. In terms of current standards aligned with attaining science literacy, this sense is concerned with the abilities and emotional attitudes necessary to appropriately convey science information. The derived sense, on the other hand, deals with the established body of knowledge in science and is related to knowing and understanding science concepts. This sense relates to content standards and standards invoking the application of science understanding (Yore et al., 2003). Yore and Treagust (2006) suggest a vision of science literacy based on this theoretical framework from Norris and Phillips resulting in the following two components:

1. The meaningful understanding of knowledge about the big ideas or unifying concepts / themes of science like the nature of science, scientific inquiry, and major conceptual themes in the biological, earth-space, and physical sciences.
2. A literacy component that stresses the cognitive abilities, critical thinking, habits of mind, and information communication technologies (ITC) to understand the big ideas in science, to inform and persuade others about these ideas, and to participate more fully in the public debate about STSE issues. (p. 293)

In this approach, the historically developed ideas of science literacy are linked with language in a way that utilizes the language to both develop science understanding and communicate that to others. One important pedagogical question then becomes what classroom actions will most likely allow for development of this type of science literacy for students. Development of appropriate classroom activities to attain this science literacy depends on a combination of theoretical and practical factors. The next section will consider one theoretical issue central to any concept of science literacy and important when developing activities in science classrooms.

Nature of Science

One of the goals of science education often stated in literature is to convey an appropriate view of the nature of science to students (Kimball, 1967; Alters, 1997). In fact, some science educators will go as far as stating this should be *the* major goal of science education (Matthews, 1994). In addition, contemporary views on science literacy such as those discussed previously are predicated on an appropriate view of the nature of science. Unfortunately, the literature on nature of science continually reports this goal is not met (Kimball, 1967; Alters, 1997) with regard to teachers (Eve & Dunn, 1990; Johnson & Peeples, 1987) or students (Lederman & O'Malley, 1990; Ryan & Aikenhead, 1992). A myriad of factors are given for the failure to achieve this goal. Some blame teacher education programs and claim they have not done an adequate job of preparing pre-service teachers in regard to the nature of science (Kimball, 1967, McComas, 1996). Others place blame on "traditional" educational systems that fail to convey an appropriate nature of science. Faulty views on how learning in general takes place are often cited as another root cause of misconceptions in terms of the nature of science (McComas, 1996).

Several fundamental issues are connected to the general dilemma of helping students form an appropriate view of the nature of science. In the context of the research that will be reported here, the main focus is to ascertain what is an appropriate view of what science is and how it is done so that any pedagogical interventions promoted to improve conceptual understanding in science are built upon and are encouraging an adequate understanding of the nature of science in students.

The Characteristics of the Nature of Science

The characteristics of the nature of science have been a part of the conversation in science education for at least the last fifty years (Kimball, 1967; Alters, 1997). During that time period however, the specific views on what these characteristics are have changed. Kimball (1967) provided an early vision of the nature of science in which curiosity was the driving force behind scientific study. In this vision, the process of science was emphasized as a dynamic activity that aims for simplification and comprehensiveness. Mathematical relationships were seen as the most effective way to present scientific understanding. Kimball viewed the physical universe as "susceptible" to human understanding if appropriate values, such as dependence on sense experience and evaluation in terms of reproducibility were consistently utilized, rather than a strict adherence to a technique driven "scientific method". Finally, science was characterized by an "openness of mind" manifested in a willingness to change ones opinion when faced with evidence and this led to science being characteristically tentative and uncertain.

Between the mid 1980's and the early 1990's, research literature began to report challenges to the previously mentioned views of the nature of science (Stenhouse, 1985; Duschl, 1985, 1988; Hodson, 1986). This led to a set of assumptions about the nature of science that placed more emphasis on the personal and uniquely human aspects of the discipline (Cleminson, 1990). Among these were assertions that scientific understanding was never to be equated with truth because it was always developed through human interpretation. This human interpretation is always the result of a theoretical lens built on prior knowledge and creative inference from data obtained through scientific techniques. Because of the personal nature of this knowledge development, the abandonment of "cherished knowledge that has been falsified" happens with great reluctance (Cleminson 1990). Finally, these assertions emphasized the interaction of the scientists as a part of the world explored, rather than as an entity outside of the studied concepts (Cleminson, 1990). Later, Ryan and Aikenhead (1992) built on these assumptions by contrasting a "worldly view of science" with a "naïve" or logical positivist view of science. According to their findings, important characteristics of an appropriate nature of science also included focus on the social purposes of science and the social construction of scientific knowledge through consensus.

Just as the National Science Education Standards (NSES) had promoted a particular view of science literacy appropriate for classrooms, this document also asserted that science teachers were responsible for contributing to student understanding of the nature of science. This publication put forth a list of the characteristics of what the nature of science entailed based on three related areas that teachers should identify and discuss in relation to all scientific understanding. First, science should be presented as a human endeavor. Second, the nature of scientific knowledge and the processes that have been used to develop it must be considered. These two areas were very much in line with previous conceptions of the nature of science. The third area consisted of a characteristic not as commonly considered, the historical perspective of science. The guidelines in the NSES claimed students should also be made aware of the historical factors, including political, cultural, and religious viewpoints that have impacted how science was done and how science understanding was developed historically.

More recently, McComas (2004) compiled a list of characteristics of the nature of science that is often cited as a guide for science educators. The list summarized many of the aspects previously discussed and included the following:

- Science demands and relies on empirical evidence.

- Knowledge production in science includes many common features and shared habits of mind. However, there is no single step-by-step scientific method by which all science is done.
- Scientific knowledge is tentative but durable. Science cannot prove anything because the problem of induction makes “proof” impossible, but scientific conclusions are still valuable and long lasting because of the way that knowledge eventually comes to be accepted in science.
- Laws and theories are related but distinct kinds of scientific knowledge.
- Science is a highly creative endeavor.
- Science has a subjective element.
- There are historical, cultural and social influences on science.
- Science and technology impact each other, but they are not the same.
- Science and its methods cannot answer all questions.

In addition to the dilemma of defining the “true” characteristics of an appropriate nature of science, there is an increasingly contentious debate surrounding the appropriate philosophical stance to take in relation to science. Emerging cultural relativist positions are based on a philosophical bent that asserts science has no legitimate claim to durable standards of truth and objectivity (Norris, 1997), while multi-culturalist stances assert a philosophy of multiple sciences that are culturally based and referenced (Stanley & Brickhouse, 2001). A more typical philosophical position used in developing contemporary views on the nature of science is what Yore et al. (2003) term “naïve realism” with an “evaluativist interpretation”. This perspective is based on an ontological position that there is a “reality” to nature that science attempts to explain, but on an epistemological position that multiple explanations for this reality will likely emerge from different people. Science as a discipline then becomes the process of developing consensus about how nature works by submitting the multiple interpretations to public debate and review. In this sense, the speculative and temporary individual views are used to build more durable ideas that serve as "scientific knowledge". The idea of inquiry learning is built on the concept that through experience with scientific concepts, students can develop their own personal science understandings that can then be publicly negotiated and compared to other scientific findings in order to develop understanding (Yore et al., 2003). The real key for educators is to recognize what view of the nature of science is driving the instructional methods they utilize in their classrooms and consider whether these views are helping develop an effective understanding of how science is done with their students. In this way the theoretical ideas discussed in the previous two sections become a practical concern.

Combining Nature of Science and Science Literacy to Guide Instruction

In combining the view of the nature of science previously related with many of the issues discussed in the section on science literacy, Hurd (1998) proposed the following description of characteristics of a scientifically literate person:

- A scientifically literate person distinguishes experts from the uninformed, theory from dogma, data from myth and folklore, science from pseudo-science, evidence from propaganda, facts from fiction, sense from nonsense, and knowledge from opinion.
- A scientifically literate person recognizes the cumulative, tentative, and skeptical nature of science; the limitations of scientific inquiry and causal explanations; the need for sufficient evidence and established knowledge to support or reject claims; the environmental, social, political, and economic impact of science and technology.
- A scientifically literate person knows how to analyze and process data; that some science-related problems in a social and personal context have more than one

accepted answer and that social and personal problems are multidisciplinary having political, judicial, ethical, and moral dimensions.

Hurd's summation provides a research based background for establishing a general goal for the science education of students, regardless of specific subject matter. The enduring dilemma emerging from this theoretical discussion is identifying what tasks will most likely lead to the establishment of these characteristics in students. Before identifying one type of task that may be an avenue for achieving development of scientifically literate students, the way students learn science in general must be considered.

How Students Learn Science

The promotion of any sound pedagogical tool must at minimum take into account current research suggesting how students learn science. A teaching intervention based on promoting an appropriate sense of science literacy and an appropriate view of the nature of science, but offered for students in a way that is not consistent with appropriate theories of how students learn will likely lead to little benefit. There is no shortage of research on learning theories and how they relate to science. In this section, several proposed explanations for student learning in science will be discussed, culminating in a rationale for linking views stemming from cognitive science with the use of writing in the science classroom.

General Categories of Cognition and Learning

From a theoretical perspective, Greeno, Collins, and Resnick (1996) categorize learning theories into three main viewpoints: the behaviorist / empiricist view, the cognitive / rationalist view, and the situative / pragmatist-socio-historic view. The behaviorist / empiricist view posits that knowledge is developed by organizing associations and developing specific skills or component skills that demonstrate these associations. More specific traditions are found within this view including associationism, behaviorism, and connectionism. Associationism is the oldest of these theories and from this perspective learning is viewed as the establishment of new associations. Behaviorism focused primarily on observed actions and defined learning as the strengthening or weakening of connections between stimuli and response. Connectionism is a more recent adaptation of the behaviorist / empiricist tradition that establishes learning as the strengthening or weakening of patterns in neural pathways or networks (Greeno, Colllins, & Resnick, 1996).

The cognitive / rationalist perspective is built on the importance of cognitive abilities including but not limited to problem solving skill, reasoning skill, and language comprehension. This position has been strengthened recently through the development of new research techniques, particularly imaging techniques related to neuroscience (Bransford, Brown, & Cocking, 2000; Byrnes & Fox, 1998). Gestalt psychology, with its emphasis on insight, has historically been categorized within this position. More relevant to educational issues today are the other two major traditions grouped under this broad category. Constructivism, which focuses on how conceptual understanding of individuals is developed and symbolic information processing which emphasizes language acquisition and understanding, have both been foundational for many educational interventions (Mayer, 1996, Henriques, 1997, Cavagnetto, 2005). These three traditions are linked by a definition of learning involving growth and changes in conceptual understanding. This growth and change is typically the result of reorganization of the way concepts are stored cognitively. Important to this cognitive growth in these theories is the development of metacognition and metacognitive strategies (Greeno et al., 1996).

The third perspective on learning and cognition is the situative / pragmatist-sociohistoric perspective. Three research positions typically aligned with this perspective are ethnography, ecological psychology, and situation theory. These positions agree that knowledge should be conceived of as being distributed among people and the environment the people are a part of. For example, ethnography is primarily concerned with patterns that develop within cultures and how these patterns impact what is "known" in the community and by the community members. In general, learning theories within this general perspective describe learning as something that is done both at the collective level and at the level of the individual. Learning progresses as those members of the community begin to understand the constraints, as well as the "affordances" that they function under as a part of a larger social group. The emphasis here is on the regular, consistent patterns within the group that impact knowledge and knowledge making (Greeno et al., 1996).

Constructivism and Science Learning

Much classroom methodology has been and continues to be based on behavioral learning models emerging from the behaviorist / empiricist perspective (Novak, 1977; Yager, 1991; Bransford, Brown, & Cocking, 2000). However, as early as the 1960's, researchers began pointing out dilemmas and inconsistencies with basing theories of student learning (and the pedagogy arising from these theories) on behaviorist notions (Ausubel, 1967). Ausubel (1967) argued that the "psychological structures" associated with behaviorist, or what was termed rote learning at the time, were different than those associated with the deeper learning desired in the classroom. Novak (1977) supported this notion of incongruence between behaviorist theories and psychological structures with an assertion that evidence gained from classrooms did not support the use and benefit from behavioral learning theories. Gowin (1981) summarized a major criticism of behavioral learning theories by noting that the observable behaviors taken to be the foundation of learning in a behaviorist model are actually outcomes of non-observable behaviors such as thinking and listening. Gowin's contention was that the actual "learning" is the process involving the non-observable behaviors, rather than the products present in the observable behaviors. As technological advances began to make some of these non-observable behaviors open for analysis (Bransford et al., 2000), cognitive learning theories began to gain favor.

Information processing theories provided a historical bridge between learning theories based on behaviorist tradition and those centered on cognitive principles. While some consider information processing theory a type of constructivism (Henriques, 1997; Greeno et al., 1996), this learning theory is more typically described as a forerunner to constructivist theory. Mayer (1996) argues for this position by referring to information processing as the “intellectual precursor” to constructivism. In general, information processing posits that much like a computer, humans take in information (input), process the information internally by applying different mental operations, and then produce some sort of product (output) (Mayer, 1996). Implied in this description is recognition that alteration of either the input or the mental processes involved can change the output created. The goal of a classroom built on this theory of learning is to involve all students in similar activities, logically designed to provide input that will lead all students to conclusions agreed to be authentic knowledge (Henriques, 1997). In a science classroom, this would typically involve students participating in tasks with a series of sub-procedures culminating in an understanding of the target concept that agrees with current scientific understanding. All students are led to the same endpoint, what is deemed by the teacher as “correct” scientific understanding (Zahorik, 1995).

Information process theorists recognize the existence of different prior conceptions in individual students, and describe one aspect of the learning process as a “continuous goodness of fit” (Henriques, 1997, p. 20) between the new information the student encounters and the existing conceptions they hold. However, it is supposed that through carefully designed instruction, all students can arrive at a pre-determined endpoint. One criticism of this position is the seemingly paradoxical view that while students may hold individual conceptions before instruction, they will hold identical conceptions after instruction. One radical interpretation of information processing involves teachers simply reciting "correct" science information to be transferred to students, who would then form similar conceptions of the target idea. As arguments developed historically in favor of more individual conceptions resulting from instruction, constructivist learning theories began to gain prominence.

Alternative Conceptions of Constructivist Theories

Constructivist learning theories have traditionally been identified as originating with Vico in the 1700's. A multitude of positions within this broad category tied to a variety of individuals, including Dewey, Vygotsky, Piaget, Von Glasersfeld, and others (von Glasersfeld, 1989; Matthews, 1994; Phillips, 1995) have been discussed in the literature. All these positions, however, are connected at minimum by a common belief that students "construct" knowledge when their experiences (including the social context of the experiences) interact with their prior knowledge to create a unique conception of information (Henriques, 1997). Appleton (1993) describes four possible outcomes resulting from this process for students:

- Assimilation: New information encountered by the student is compatible with their prior conceptions. The new information is integrated into the students existing cognitive structures and no significant change takes place.
- Accommodation: New information encountered is not compatible with the student's existing understanding. Restructuring of the student ideas takes place leading to a change in understanding that is manifested in changes in the way the information is cognitively structured.
- Waiting for "Right Answer": Students realize that their previous conception is inadequate when faced with new information. Rather than accommodating new information and changing cognitive structures, students wait to hear the "right answer" from an authority figure. Students may repeat "right answer" back when

questioned in similar context but will often fail to apply the concept correctly in new situations.

- Ignore Conflicting Evidence: Students realize their conceptions are not accurate or inadequate but do not make effort to change their ideas or wait to hear new ideas. Prior conceptions are maintained even in light of conflicting evidence.

While all manifestations of constructivist theory can potentially lead to these possible outcomes, different perspectives are based on different epistemological positions, and lead to different pedagogical interventions. In the following sections, two of the most extreme views of constructivist theory, radical constructivism and social constructivism will be contrasted and critiqued. A third position, interactive constructivism combining aspects of each of the first two positions will then be discussed. A rationale for the use of this third perspective as the learning theory informing the educational intervention focused on in this study will be offered.

Radical Constructivism

From a radical constructivist viewpoint, the focus in learning is squarely on the individual. Learning is the result of Piagetian disequilibrium, in which the individual is confronted with information that does not align with his or her pre-existing conception of an idea (Duffy & Cunningham, 1996). While knowledge may be impacted by the social context that an individual is a part of, the knowledge resides in the mind of the individual. From an epistemological stance, radical constructivism allows for multiple interpretations of the world, and as long as explanations offered are in line with accepted social and cultural norms, all explanations are equally valid (Henriques, 1997). The relativistic nature of this position is due to its adherence to a commitment to the individuality of human experience, and therefore, a unique interpretation of each experience from individuals (Matthews, 1994). Therefore, each individual's own understanding must be considered valuable (Ertmer & Newby, 1993; Yore & Shymansky, 1997). Ontologically, this position posits that knowledge is a human construct and rejects the notion of an objective reality (Phillips, 1995).

The main objection to the radical constructivist position in terms of science learning is the mismatch between the postmodern view espoused by this learning theory and an appropriate view of the nature of science discussed earlier. The postmodern, relativistic slant of radical constructivism is based on the core belief that all knowledge is a human construct (Prawatt, 1999). This being the case, judgment on the validity of ideas falls to the individual, and the natural outcome of this individual assessment is that all ideas judged adequate by individuals are equally acceptable (von Glaserfeld, 1989; Yore & Shymansky, 1997). This standard of validity does not align with a philosophy that science is a search for an understanding of the natural world, by testing in the natural world, and by analyzing which descriptions of the natural world are supported to a greater or lesser degree by these tests. As Osborne points out (1996), radical constructivism provides no avenue for the determination of which of a number of alternative explanations is better in terms of accurately describing nature. From a pragmatic viewpoint, this characteristic of radical constructivism would make assessment of student understanding problematic in that most ideas students posit would necessarily be judged appropriate.

Social Constructivism

Social constructivism proposes a similar ontological view of the world, as does radical constructivism in that it asserts no objective reality (Henriques, 1997). Epistemologically it proposes a different pathway to our knowledge construction about the world. Social constructivism posits that knowledge is built through disequilibrium resulting from interactions between individuals and followed by a process of social negotiation in reaching a consensus understanding (McCarthey & Raphael, 1992, Yore & Shymansky, 1997; Henriques, 1997; Cavagnetto, 2005). This is the opposite of the individual cognition responsible for knowledge growth proposed in radical constructivist theory. This position is typically traced to Vygotsky's (1978) ideas promoting the initiation of all higher human functions arising from social interaction. The foundation of this position is the premise that the mind is located in an "individual-in-social interaction" context (Cobb, 1994) and that validity is established when the community or group reaches consensus. From this perspective, individual knowledge is modified, mitigated, and evaluated by community action, with value being placed on the socially negotiated meaning (Yore & Shymansky, 1997; Cavagnetto, 2005).

As with radical constructivism, this position can also be critiqued in terms of its relativistic slant. While the origin of the relativistic generation of knowledge switches to the group, the same neglect for standards of validity based on a match with nature exist with this position. Community judgment and group consensus becomes more important than nature (Henriques, 1997; Cavagnetto, 2005). Osborne (1996) asserts that the upshot of this philosophy when specifically applied to science learning is that social cultural interactions are favored over evidence from nature. In addition, in emphasizing the consensus building aspect of science knowledge production, the social constructivist position neglects the importance of individual scientists engaging in independent thought to move science understanding forward (Cavegnetto, 2005). This would also tend to discourage the creative aspect of the nature of science.

Interactive Constructivism

The inadequacies of the previous two positions in accomplishing the dual task of adequately describing how students learn and aligning that viewpoint with an accurate view of the nature of science have led to calls for the adoption of an intermediate position (Cobb, 1994; Phillips, 1995; Prawat, 1999). Interactive constructivism has been offered as a viewpoint that not only recognizes aspects of radical and social constructivism, but also involves tenants of all three of the major positions on learning and cognition referred to earlier. The major premise of this position is that learning has both public and private aspects (Henriques, 1997). Public learning involves the social interactions, negotiations, and experiences with the broader science community that impact student cognition. Personal aspects of learning involve the reflection, interpretation, and meaning making that an individual student will undergo (Henriques, 1997). Both public and private learning follow a pathway consistent with the general constructivist position in which student prior knowledge is made explicit and then confronted with experiences that either agree with or challenge these understandings, followed by a process of modifying (or retaining) existing knowledge structures. The strength of the interactive constructivist position when applied to science learning is that it allows for the impact on ideas from social discussion and interaction, as well as individual contemplation, but it also holds these understandings accountable to nature (Cavagnetto, 2005). With this viewpoint, evidence from nature is analyzed to determine its match with either individually created ideas or group built findings. Henriques (1997, p. 22) summarizes the characteristics of interactive constructivism as follows:

- Alignment among outcomes, instruction, resources, and assessment
- Outcomes of conceptual change, conceptual growth, and metacognitive strategies all impact learning
- Does not rule out direct instruction embedded in natural context of need
- Supports big ideas / unifying concepts (AAAS, 1993; NRC, 1996), science literacy and habits of mind needed to attain science literacy
- Requires students to gain ability to construct the construction, think critically, to communicate their constructions and persuade others of their value or utility
- Encompasses guided inquiry, learning cycles, conceptual change, and generative approaches
- Teaching involves accessing, engaging, experiencing / exploring, justifying / rationalizing, consolidating / integrating old and new, and applying knowledge

As is evident from the list above, the interactive constructivist position utilizes aspects of all other major positions on learning. In addition, it relies on and perpetuates a view of the nature of science that aligns with currently accepted ideas. Finally, it allows for the promotion of both the fundamental and derived senses of science literacy.

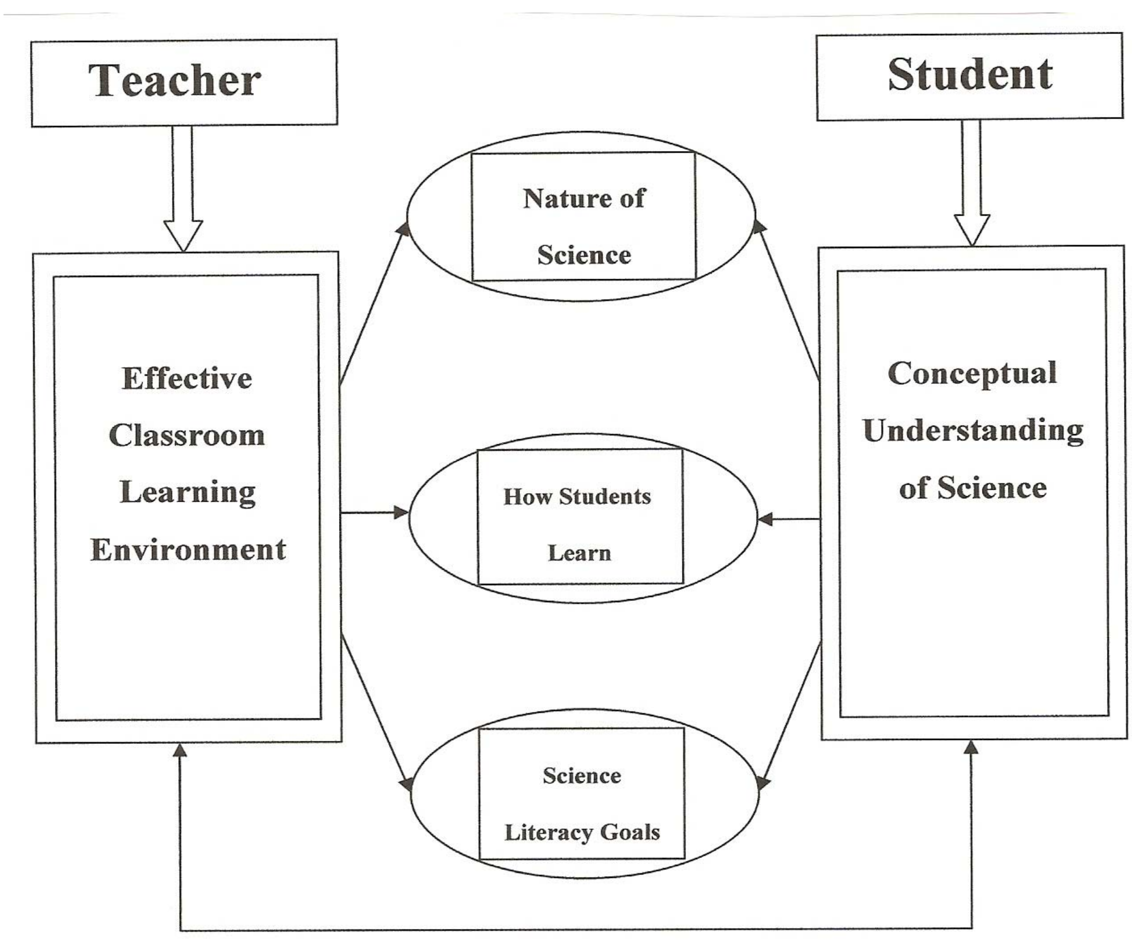
Summary of Factors Influencing Learning in a Science Classroom

The three broad areas previously discussed (science literacy, nature of science, and how students learn) dynamically interact in any science classroom to impact the learning that takes place. Figure 1 represents a schematic representation of this author's overview of how this takes place.

The intent of figure 1 is to graphically depict the learning environment not only in terms of the theoretical ideas already covered, but also in terms of the major participants in the classroom, the teacher and the students. Teachers are ultimately responsible for designing and implementing practices that will result in an effective learning environment for students. Ideally, this would involve the implementation of pedagogical tools that give students the greatest possibility of developing a sound conceptual understanding of science topics.

As pictured, the on-going learning environment the teacher develops should be based on an appropriate goal in terms of science literacy, should communicate and foster an appropriate view of the nature of science, and should utilize procedures in the classroom to develop concepts that are based on knowledge of how students learn science. The other main participants in the classroom environment are obviously the students and the conceptual understandings that they develop will reflect their own goals of what they believe they should be obtaining from participation in the class (in essence, their view of what their particular "science literacy goals" are) and a view of the nature of science that characterize the learning environment as well as their own overall experiences. Further, student understanding will be impacted by the actual, individual way that each student learns science interacting with the way information has been dealt with and presented in the classroom. In addition, each individual student will have multiple social interactions with all other students and the combined social dynamic will impact the learning of all students. The ideas on student learning the teacher uses to design the learning environment will have differing degrees of match between the actual way each student learns and the level of agreement will to some extent determine the level of conceptual development attained by an individual student.

Figure 1. Factors Interacting in a Science Classroom



The arrow at the bottom of figure 1 linking the teacher developed learning environment with the student developed conceptual understanding is meant to further illustrate the interactive nature of the classroom dynamic. The emerging understanding of each individual student is constantly interacting with the learning environment, including the understandings of all the other students. Ideally, as student development of conceptual understanding takes place, the emerging ideas will be assessed by the instructor and initiate further planning or modification of the learning environment by the instructor.

The description of this learning environment highlights the complexities associated with endeavoring to develop a practical classroom situation based on the theoretical ideas discussed in the previous sections of this chapter. To aid teachers in dealing with this complexity and to ultimately promote greater student learning, investigation of pedagogical tools that are likely to contribute to this type of effective and dynamic learning environment in science and how they can be practically implemented is needed. The remainder of this chapter will provide information from the literature suggesting that the utilization of multi-modal writing tasks may be one possible route to accomplishing the goal of promoting science learning.

Relating Theoretical Positions to Practical Classroom Activities with Writing-to-Learn

The first part of this literature review has argued that sound science instruction leading to attainment of science literacy incorporates many theoretical facets that must be considered. The major pragmatic question stemming from this theoretical argument becomes what types of actual instructional methods would best be suited to reaching these learning goals. Research exploring the use of writing-to-learn has begun to establish that writing tasks, used appropriately, may provide one option. (Yore et al., 2003). As meta-analyses in research literature during the 1980's began to indicate that the promotion of hands-on-activities lacking a "minds-on" connection was less effective than anticipated in terms of student gains (Shymansky, Kyle, & Alport, 1983; Willett, Yamanshita, & Anderson, 1983; Wise & Okey, 1983), the push to utilize writing activities in the role of this missing minds-on component began in earnest. The assumptions that writing could provide this mental counterpart to hands-on activities were based on emerging cognitive models detailing what happened as students wrote. The first portion of this section will consider the discussion surrounding the cognitive foundations supporting the use of writing-to-learn activities.

Cognitive Science and Implications for Writing

The vast area of cognitive science is where the idea of constructivism as a learning theory originally developed. When determining appropriate pedagogy and interventions for science classrooms, awareness of an emerging avenue of research in cognitive science is helpful. Klein (2006) provides a relevant theoretical description of cognition in general in his article contrasting first and second generation cognitive science views. First generation cognitive psychology generally portrayed knowledge, thinking, and language as part of a denotative process while developing second generation views provide a more expressive outlook. The first generation view posits language is a "window into thinking" in which language follows thinking. The specific language of a discipline (in this case science) reports the knowledge of that discipline. This view on cognition is aligned with a philosophical view of science that is rational, and relies on inductive and deductive reasoning to generate knowledge that can then be communicated in a clear way using the forms and vocabulary unique to the discipline. These views on cognition and the nature of science would align with the use of more traditional writing tasks in science (Yore & Treagust, 2006).

Conversely, the second generation view of cognition is more aligned with the use of non-traditional writing tasks and is based on a modern view of science that is more hypothetico-deductive in nature. This conception of cognition and its relationship to language is less straightforward and more "fuzzy". Proponents of this view of cognition propose that language in general, and writing in particular, when applied to a particular discipline such as science, can not only communicate ideas, but also shape them. This development of ideas is accomplished by the strengthening of connections between related aspects of concepts dealt with in the act of using the language to describe the science (Klein, 2006; Yore & Treagust, 2006). Yore and Treagust (2006) suggest that adoption of this second generation perspective has two potentially important results. First, it would allow for a more "pragmatic" view of logic coupled with a "flexible mixture of natural and disciplinary language" (p. 299) that can enhance connections between experiences, oral discourse and text. Secondly, if this view of cognition, and the accompanying position of the use of more expressive and creative writing are more accurate in representing the way cognition occurs in students (as Klein (2006) suggests), then engagement and student motivation for learning can be increased (Yore & Treagust, 2006). Important to the study presented here, Klein (2006) suggests that the use of non-traditional writing tasks can provide a bridge between more denotative outlooks typical of the discipline of science and typical student thinking, again with the potential for improving student learning.

The discussion about general ideas of cognition offered here set the stage for discussion about specific cognitive aspects of writing. The next section will provide an overview of the historical ideas about specific cognitive action when students write.

Historical Overview of Cognitive Models of Writing

As focus on using writing as an instructional tool increased over the past half century, models attempting to describe what is happening cognitively in these situations emerged. In this section, a brief historical overview of the proposed cognitive models for the foundation of writing-to-learn activities will be presented. Then, a more in-depth analysis of a current cognitive model that will be used as the basis for the research presented here will be addressed.

Problem Solving Models

An initial cognitive model associated with writing developed when Emig (1977) posited that writing led to more student awareness of relationships and connections among ideas due to its constant feedback. The student writer needed to deal with both knowledge of the content written about and of composition in general. By dealing with both, the writer was able to interact with the developing text in a way that promoted learning (Wallace, et al. 2004). Emig's main assertion was that writing provided an instructional technique more powerful than discussion because it allows for a cognitively rich process in which

> …the symbolic transformation of experience through the specific symbol system of verbal language is shaped into an icon (the graphic product) by the enactive hand. (p. 124)

Emig believed that writing could play this transformative role due to its unique relationship to learning strategies. These relationships are summarized table 1.

Cognitive models of the 1980's developed around the theme of writing as a problem solving activity. Flower and Hayes (1980) proposed a goal-oriented model in which the writer balanced goals of the task with goals dealing with the content. In describing their model, Flower and Hayes carefully define the "discovery" aspect of writing in a way that includes important cognitive action:

> A writer in the act of discovery is hard at work searching memory, forming concepts, and forging a new structure of ideas, while at the same time trying to juggle all the constraints imposed by his or her purpose, audience, and language itself. Discovery, the event, and its product, new insights, are only the end result of a complicated intellectual process. (p. 21)

Flower and Hayes asserted that three main activities, planning, translating, and revising or reviewing, interact cognitively to create meaningful writing. This interaction is framed in reference to a “rhetorical problem” that the author must solve. In essence,

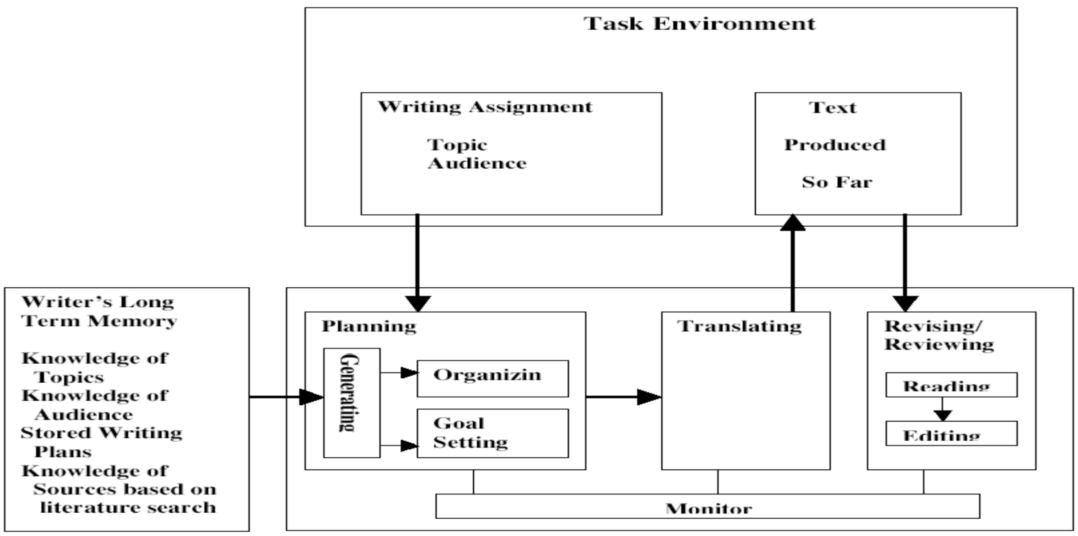
Table 1. Unique Cluster of Correspondences between Certain Learning Strategies and Certain Attributes of Writing

Selected Characteristics of Successful Learning Strategies	Selected Attributes of Writing, Process and Product
1) Profits from multi-representational and integrative reinforcement	1) Represents process uniquely multi-representational and integrative
2) Seeks self-provided feedback:	2) Represents powerful instance of self-provided feedback:
(a) immediate	(a) provides product uniquely available for immediate feedback (review and re-evaluation)
(b) long-term	(b) provides record of evolution of thought since writing is epigenetic
3) Is connective:	3) Provides connections
(a) makes generative conceptual groupings, synthetic, and analytic	(a) establishes explicit and systematic conceptual groupings through lexical, syntactic, and rhetorical devices
(b) proceeds from propositions hypotheses, and other elegant summarizers	(b) represents most available means (verbal language) for economic recording of abstract foundations
4) Is active, engaged, personal – notably self-rhythmed	4) Is active, engaged, personal – notably, self-rhythmed

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

this requires the writer to either search their existing stored memory for a representation of the type of writing they are to produce, or conversely, create a unique representation for the particular task at hand. Flower and Hayes found that more expert writers were able to create writing that took into account their own goals for the writing as well as their understanding of the potential reader in creating a product with more conceptual depth. Novice writers, on the other hand, produced writing more concerned with simply meeting rhetorical requirements (number of pages, characteristics of the required genre, grammatical concerns) than with expressing appropriate conceptual goals. The aspects of this model can be summarized in the diagram in Figure 2.

Figure 2. Flower and Hayes Cognitive Model of Writing Process

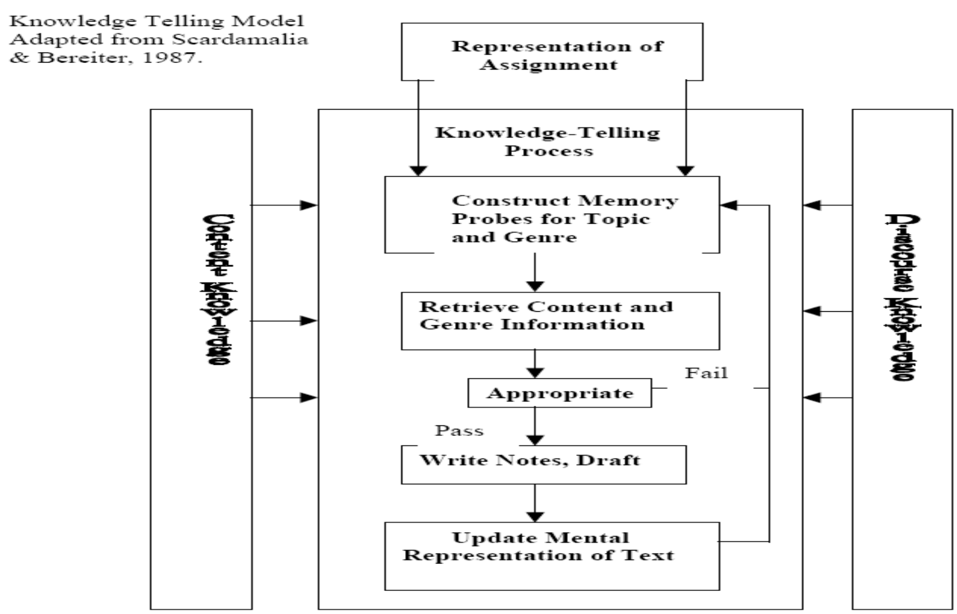


Source: Flower, L., & Hayes, J. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication, 31*, 21-32.

Bereiter and Scardamalia (1987) followed with another problem solving model that is still recognized as a dominant model in education today (Klein, 1999). Agreeing with Flower and Hayes that writing was basically a problem solving task, Bereiter and Scardamalia clarified the cognitive processes involved by describing a dynamic interaction between the cognitive content space and rhetorical space of the writer. Relying on data from think-aloud protocols and focusing on the differences between expert and novice writers, this model posited differential levels of cycling between these mental spaces. The differential activity led to two very different outcomes.

The Bereiter and Scardamalia model contrasts two types of writing. "Knowledge-telling" writing involves relating to another audience recall of information already known. "Knowledge-transforming" writing involves a dynamic process where the information that the student creates in text cognitively reorganizes information previously known. This involves interplay between three phases of writing (planning, generating, and revising) and two factors impacting writing (task environment and writer's long-term memory). In general, the cognitive content space of the writer interacts with their rhetorical space as goals shape text production and information held in the writer's content space is transformed (Galbraith & Torrance, 1999). Figures 3 and 4 show representations of the two contrasting models.

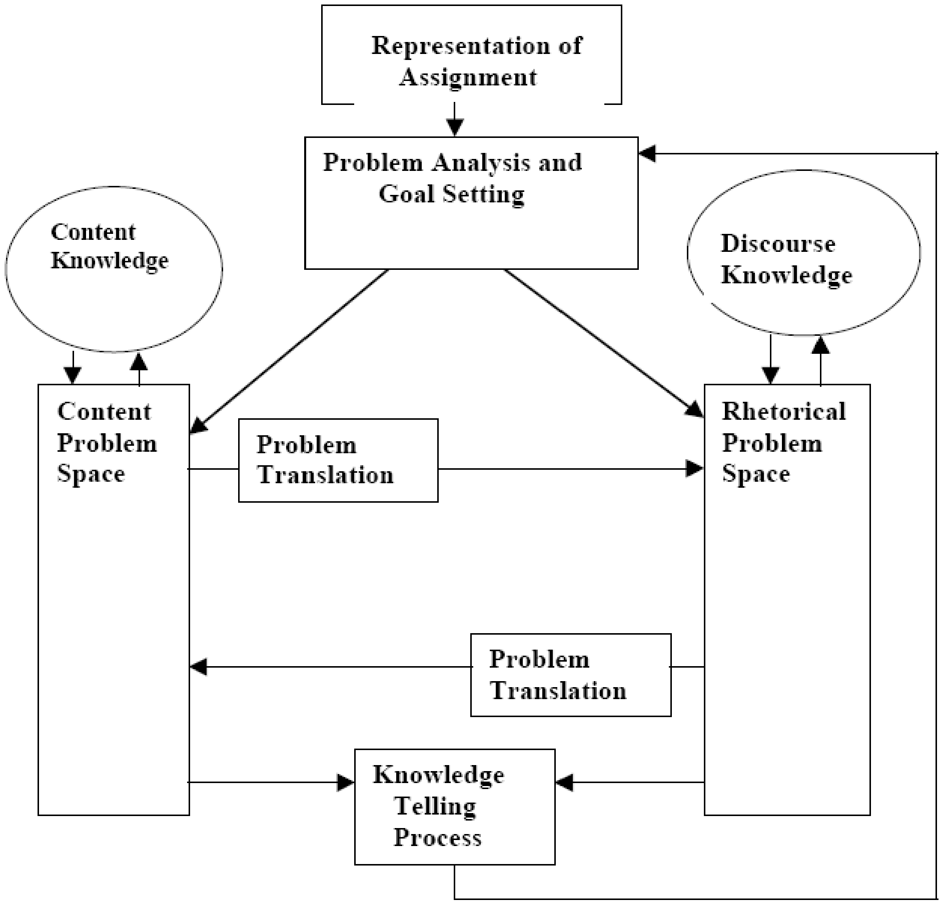
Figure 3. Bereiter and Scardamalia Knowledge Telling Model of Writing



Source: http://www.uic.edu/classes/psych/psych303/Psych303

Figure 4. Bereiter and Scardamalia Knowledge Transforming Model

Knowledge-Transforming Model
Adapted from Scardamalia &
Bereiter, 1987



Source: http://www.uic.edu/classes/psych/psych303/Psych303

Research in the mid 1980's indicated that most writing tasks that were being used in science classrooms were consistent with a knowledge-telling perspective. These writing tasks were characteristically short, informational passages for the teacher (Langer & Applebee, 1987). Yore et al. (2003) reported that these writing tasks were utilized by the teacher as an evaluative tool because they represented students' attempts to put into print mental representations they held of science information. The writing process associated with these tasks was described as mechanical and linear, emphasizing recall, and manifested in formal short answer or essay products that were produced with little peer discussion or interaction. However, Holliday et al. (1994) when discussing the knowledge-transforming outlook, described an emerging realization amongst science educators in the late 1980's and early 1990's that writing certainly could be and often was much more than this:

> Bereiter and Scardamalia (1987) described the interactive and constructive processes involved in the knowledge-transforming model of writing that parallels the generative model of science learning in that it involves long-term memory, working memory, and sensory-motor activity. The knowledge-transforming model appears to be far more interactive and recursive than linear. (1994, pp. 885-886)

Accompanying this realization was a corresponding shift in pedagogical action associated with writing tasks. Instructors began to utilize a variety of writing tasks and audiences, for a variety of purposes, and associated with a variety of instruction. This varied instruction included the fundamental recognition of the use of language as a symbolic way to represent science understanding, explicit focus on the writing process and how it relates to scientific inquiry, mention of the specifics of different types of science genres and their characteristics, and emphasis on metacognitive awareness and control in writing (Ferrari, Bouffard, & Rainville, 1998; Sawyer, Graham, & Harris, 1992).

Other models have been proposed to complement, supplement, or even replace those discussed so far. While all models have their own unique attributes, Klein (2006) attempted to highlight some important similarities among the many cognitive models associated with writing. First, all models in some way refer to the separate but interacting factors of content knowledge and rhetorical knowledge. While the terminology used to refer to these is not always the same, all models recognize that in writing, what students know about the concepts involved and what they know about the act of writing in general interact to influence the learning that takes place. Secondly, all models indicate some level of cycling through or feedback from different cognitive areas if writing promotes learning. The proposed ways this cycling and feedback is initiated and sustained differ, but the presence of this type of cycling is indicative of more cognitive action and ultimately, more knowledge construction. Third, the models refute the "strong text argument" that writing in and of itself leads to better achievement and recognize the need to develop writing tasks in specific ways to encourage learning (Klein, 2006). The general conclusion resulting from comparison of the models is that writing can do more than reflect understanding, it can transform it. The final model considered here takes this notion one step further.

Galbraith's Model of Knowledge Constitution

While agreeing with some aspects of the Bereiter and Scardamalia model, Galbraith (1999) proposed a model that accounted for the unpredictable nature of writing and for the possibility of knowledge generation, rather than just transformation. In Galbraith's model, tacit or implicit knowledge becomes explicit due to interaction between the writer's neural representations of information and constraints on the writing, including linguistic, rhetorical, and task factors. Galbraith uses the term *writer's disposition* to indicate cognitive content knowledge and the term *linguistic network* to indicate cognitive rhetorical knowledge. The activation of what he calls a "dispositional dialectic" and feedback loops through cognitive networks as a result of the act of writing lead to text production that may, if the writing task is set up appropriately, actually encourage the constitution of *new* knowledge for the writer (Galbraith, 1999). This knowledge constitution is the factor in Galbraith's model setting it apart from previous models. His model does not dispute that in some writing situations, knowledge is either transferred or transformed, but it suggests that in other cases, original thinking and new knowledge development could emerge from the writing itself.

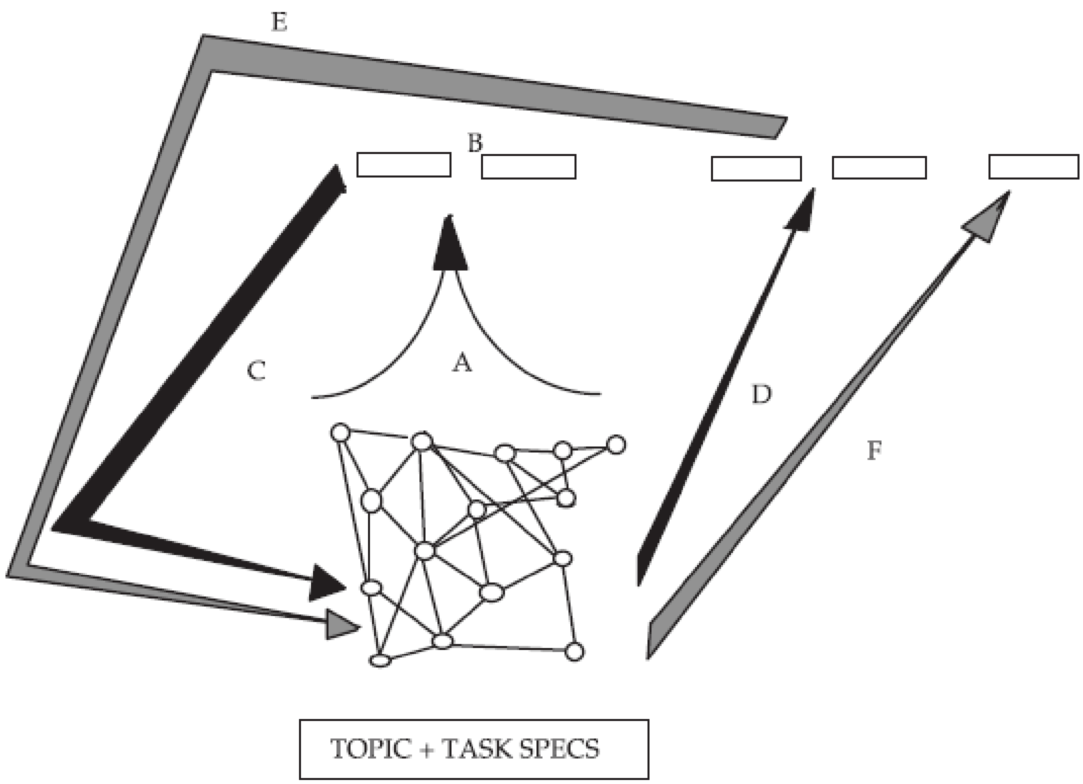
Galbraith's cognitive model was based on evidence gathered from three main sources. First, he used evidence from characteristics of expert and novice writers to help distinguish processes involved in writing. Secondly, he used evidence from "think-aloud" protocols in which students who were writing described what is happening "in their heads" as they write. Finally, he based much of his work on experimental situations in which different types of writing tasks and task conditions were used as treatments. Data collected from all of these situations indicated that when specific task conditions were applied, writing allowed students to generate new knowledge (Galbraith and Torrance, 1999).

Galbraith's main conclusion from his empirical work was that two processes are possible in writing. First, there is an aspect of writing that involves rhetorical planning, but this does not lead to the formation of new ideas, simply the reorganization of existing ideas. Secondly, a process Galbraith called dispositional spelling out, can potentially lead to the writer generating thought as they are producing text. Galbraith asserted that the main difference between the two was whether knowledge was implicit or explicit. He argued that the cognitive representation of knowledge involved when writers use the interaction of their background knowledge with rhetorical goals to guide their writing involves explicit knowledge. This would be "prior" knowledge that the writer has about the concept they are dealing with in the writing task. The interaction of the rhetorical goals for the task with this explicit knowledge is in reality reorganizing knowledge that was already there. This reorganization may lead to transformed knowledge, but not to knowledge generation. Knowledge related to the actual production of text, however, is implicit and is stored in what Galbraith called a distributed network of conceptual relationships. This information, represented in activation patterns in a network, is not something that can be retrieved (indicating it was already present knowledge), it is something that is generated at the moment of text production (Galbraith, 1999).

The act of the writer reading the newly emerging text can then initiate another cycle of cognitive activity. Figure 5 is a representation of Galbraith's model. In the figure, the topic and task specifications interact with the web, representing the writer's dispositional network (A). This interaction leads to text production, indicated by the small white rectangles at the top (B). Notice it is the emergence of the initial text that causes a feedback to the writer's dispositional network (C). This process continues as further text emerges (D,E,F).

It is Galbraith's contention that writing tasks structured in a particular way can potentially encourage a greater amount of dispositional dialectic that will ultimately lead to greater knowledge construction and generation for students through participation in the writing activity. To accomplish this, students must be involved in writing tasks where they cannot simply fulfill rhetorical and task goals by repeating back information they already understand. In order to create the optimal conditions to provide the opportunity for knowledge construction and generation, the students must be given opportunities to engage in clarification of the conceptual ideas through tasks that will require them to cognitively deal with making connections and developing relationships in order to fulfill the task requirements. These types of tasks will likely initiate processes that Galbraith details. The pedagogical dilemma then becomes determining what types of tasks are composed of the characteristics that allow for students to participate in these cognitive actions to improve understanding. The next section will discuss what types of writing may be most beneficial in this regard.

Figure 5. Overview of Galbraith's Knowledge Constitution Model



Source: Galbraith, D. (1999). Writing as a knowledge-constituting process. In D. Galbraith, & M. Torrance (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 139-159). Amsterdam: Amsterdam University Press.

Types of Writing and Student Benefit

Little argument can be offered for the contention that current science reform, as proposed by documents such as the National Science Education Standards (National Research Council, 1996) promotes an interactive-constructivist viewpoint in science classrooms. As writing in science has come to be viewed as valuable for its ability to "enable the discovery of knowledge" (Connolly, 1989), a natural theoretical collaboration between writing and interactive constructivism has been discussed (Yore et al., 2003).

One barrier, however, to this potential collaboration has been ongoing debate as to the most effective types of writing for accomplishing science literacy. This debate is also critical in answering the pragmatic question of what types of writing will lead to the student benefit that has been discussed previously in this chapter.

Agreement and consensus exists on the more global questions of what writing-to-learn tasks in science should strive to do and what classrooms employing these techniques should strive to look like. The general goal of the tasks is to improve student science learning by encouraging student reflection, consolidation, and elaboration of concepts. In addition, student development of deep conceptual understanding is promoted through higher order thinking skills when these tasks encourage hypothesizing, interpretation, synthesis, and persuasion (Yore et al., 2003). Current research promotes the establishment of several major characteristics of science writing as a part of a well-integrated science learning classroom. Writing should include narratives, descriptions, explanation, instruction, and argumentation (Gallagher, Knapp, & Noble, 1993; Yore 2000; Yore et al., 2003). Debate, however, surrounds the question of whether traditional writing tasks (such as lab write ups or technical lab reports) or alternative writing tasks (encompassing a huge variety of tasks including but not limited to brochures, letters, personal narratives, poems, or non-traditional lab reflections) are more likely to include these characteristics and lead to greater learning in science classrooms.

Traditional Writing in Science

Supporters of the use of traditional writing genres generally hold that science as a discipline and worldview has developed a specific style of writing designed to accurately convey scientific ideas and connections among these ideas (Prain, 2006). Researchers with this view promote the need for students to understand and to replicate the unique structural and vocabulary components of these types of writings in order to become scientifically literate (Halliday & Martin, 1993; Unsworth, 2001; Martin & Veel, 1998).

Teaching techniques offered to accomplish this include instruction to develop awareness of linguistic features of these typical science writing texts, along with practice producing texts similar to those used "by scientists" (Martin, 1999). Additionally, supporters of this view contend that everyday language is an inappropriate means for communicating about science because it is not technical or precise enough to accurately describe scientific phenomenon (Gee, 2004). Therefore, students would not benefit in terms of learning science from activities that do not require the use of scientific language, rather science learning would result from the production of replications of the types of writing done by scientists (Prain, 2006). In addition, Martin (1993) suggests that using non-traditional writing is "patronizing" to learners, implying that they are not able to understand or use the "official" writing style of the science community. Others have argued that diverse writing is in fact geared towards more "linguistically gifted" students (Martin & Veel, 1998) and disadvantages a significant segment of the student population.

Utilizing Alternative Writing Tasks in Science Classrooms

Proponents of the use of non-traditional writing in science classrooms offer an alternative viewpoint. In general, the implementation of non-traditional writing tasks is seen as a way to broaden the perspective of the writing students do in science classrooms and as means to generate scientific knowledge and knowledge about science. Students would encounter the attitudes and perspectives associated with the processes science uses to develop content knowledge (Prain, 2006). To accomplish this, non-traditional writing tasks seek to manipulate the types of writing, along with the purposes of writing and the audiences written to. Rowell (1997) succinctly summarizes the intent of these tasks by describing them as a way to connect emerging science knowledge and the technical vocabulary of science to students' everyday language and their past experiences. Employing these types of tasks can be accomplished through a number of different types of writing, but as Prain (2006) points out in summarizing a vast array of research in this area, when students write to “paraphrase, re-word, elaborate, unpack and re-represent meanings, express uncertainties, analyze comparisons, and reconstruct understandings” (p.185) there is opportunity to strengthen connections in their understandings and to go beyond the recording of past understandings. Ideally, this leads to science learning that can last beyond the classroom.

Comparison of Traditional and Alternative Writing Tasks

Empirical support exists for both the use of traditional and alternative writing tasks. However, as Bangert-Drowns, Hurley, and Wilkinson (2005) point out in their meta-analysis of writing tasks in education, the level of benefit is not only a function of the statistical analysis utilized, but also the specific characteristic of the sought after benefit from the writing task. Writing tasks designed more in line with a knowledge-telling perspective may demonstrate benefit when benefit is conceptualized as a recall of information presented by the instructor, but may do little to promote the transformation of existing concepts or the generation of unique conceptual understandings. Therefore, when comparing the effectiveness of the use of traditional writing tasks with the use of alternative writing tasks, clarity of the nature of the targeted benefit is imperative. In terms of knowledge transformation and generation, research tends to indicate greater benefit from utilizing alternative writing tasks (Yore, 2003).

Analysis of research reporting success in utilizing more traditional writing indicates that the writing tasks in this category tend to promote duplication of knowledge rather than knowledge generation (Yore, 2003). The writing tasks are typically supplemented with explicit instruction on the usage of grammar, the appropriate conception of audience, and word usage from the perspective of how scientists “normally” communicate (Yore, 2003; Kelly & Takao, 2002; Rice, 1998; Koprowski, 1997). The main thrust of these tasks, as Yore (2003) asserts, is “replication of the norm.”

The utilization of alternative writing tasks, conversely, is based on the premise that student understanding of science knowledge can be transformed or even created to some extent through participation in the writing task (Hand, et al., 2001). Typically, these tasks are designed to allow students to experience a “write-react-revise” scenario in which their personal understanding is presented to an authentic audience in an attempt to clarify and develop that understanding through explanation to a non-expert reader (Yore, 2003). The potential types of writing tasks are immensely varied and those studied in the literature have ranged from poetry (Watts, 2001) to anthropomorphic writing (Hildebrand, 2002) to modified lab write-ups (Hand & Keys, 1999). Typically, the use of these tasks is paired with a rich classroom experience including social negotiation of conceptual understanding, constructive argument, and learning opportunities stemming from experience and activity and is embedded in the holistic science learning experience (Hand et al., 2001). Student benefit has been observed not only in conceptual understanding through standardized and teacher generated assessment, but also in terms of student writing performance, student understanding of the processes and nature of science, student attitudes about both science and writing, and improved student argumentation and higher order thinking skills (Yore, 2003; Keys, Hand, Prain, & Collins, 1999; Tucknott & Yore, 1999; Wray & Lewis, 1997).

Evidence from previous research also suggests that incorporating alternative writing-to-learn activities in science classrooms can have positive effects at all grade levels. Jaubert and Rebiere (2005) reported improved construction of science knowledge when using personal writing about science controversies with primary school students. Boscolo and Mason (2001) showed that fourth graders involved in activities using writing-to-learn strategies were able to achieve deeper conceptual understanding and metacognitive awareness of their learning compared to students not writing. Keys, Hand, Prain, and Collins (1999) and Hand, Wallace, and Yang (2004) reported benefits for middle school students using an alternative writing-to-learn approach called the Science

Writing Heuristic (SWH). These benefits included improvements in data analysis, in building connections among scientific procedures, in metacognition, in understanding the nature of science, and in performance on extended response conceptual questions. Hand, Hohenshell, and Prain (2004) also found that two writing experiences were more beneficial than just one for tenth graders when concepts were tested two months after completion of the treatment.

Factors to Consider in Designing Alternative Writing Tasks

Yore and Treagust (2006) explicitly point out that the diversified writing tasks encouraged in the holistic approach to science teaching through utilization of alternative writing tasks are not equivalent to what most people call “creative writing”. Consideration of each specific writing task in terms of its ability to correspond to underlying learning theories, its ability to promote authentic science learning and discourse, and its ability to incorporate known effective science teaching practices is necessary. Further, as the meta-analysis by Bangert-Drowns, Hurley and Wilkinson (2004) mentioned earlier points out, not all writing tasks are beneficial. Klein (1999) indicates as much in noting that research refutes the “strong-text” argument that any type of writing will lead to student gains. These assertions led to research intended to clarify the characteristics necessary to include in alternative writing-to-learn tasks to encourage student benefit.

Prain and Hand (1996) offered a model for developing alternative writing-to-learn activities based on five main elements: topic, type, audience, method of text production, and purpose. The intent was to provide an interactive model in which different combinations of the five elements could be used to develop a multitude of writing tasks. Topic referred to the particular science concept under study. Teachers would typically determine which topics dealt with lend themselves to writing tasks. Type referred to the specific writing product called for and could include many possibilities ranging from narratives, to brochures, or even more technologically recent items such as powerpoint productions. Audience dealt with selection of the authentic group written to. Method of text production indicated the mode of creating text (handwritten, typewritten). Finally, the purpose selected related to the specific curricular goal of the activity (exploration of a concept, clarification, application). The model was intended to give instructors a framework to guide consideration of relevant factors in setting up the tasks as opposed to simply asking students to write about a topic.

The Use of Multiple Modes of Representation in Writing-to-Learn Tasks in Science

Researchers agree that many unanswered questions in relation to writing-to-learn remain (Rivard, 1994, Holliday, Yore, & Alverman, 1994; Klein, 1999; Yore, Bisanz & Hand, 2003). Ongoing research is attempting to fully understand aspects of each of the particular elements associated with these types of tasks discussed in the previous section, as well as how they relate to each other (Gunel, Hand & McDermott, in press). Much research has focused on audiences written to as well as type and purpose of writing. Less frequent and less comprehensive has been research related to delineating particular aspects of the writing the students create themselves. One emerging area of research in this vein is the utilization of multiple modes of representing conceptual information (Gunel, Hand, & Gunez, 2006; Hand, Gunel, & Ulu, 2008; Pineda & Garza, 2000). In the remaining sections of this chapter, research highlighting how the use of multiple modes of representing science information as a part of science writing-to-learn activities may be an effective way to accomplish the theoretical and curricular goals set up throughout this chapter.

The use of the term "multi-modal" in the research literature is somewhat confusing because it is often indiscriminately applied to different ideas. In general, multi-modal refers to utilizing different means of representing the same or similar concepts. The types of different modes are quite varied, but often include diagrams, pictures, charts, mathematical equations, photographs, or tables in addition to text (Gunel et. al, 2006). Some studies expand the notion of multi-modal to representations that include video, animation, audio sounds, or other technological enhancements (Pineda & Garza, 2003). For purposes of this study, multi-modal representations are utilized in relation to writing-to-learn tasks and include any mode the student author uses other than text in their writing that relates to the target concept. The terms “multi-modal” and “multiple representations” will be used synonymously to refer to the varied modes.

Three important connections can be highlighted to suggest the potential benefit for science student conceptual development through the use of multi-modal representations in their writing-to-learn activities. First, the use of multiple modes can be related to specific ideas on cognition and learning overlapping with the previously discussed ideas in this literature review. Specifically, this overlap can be conceptualized in relation to a second linkage between science learning, writing tasks, and multiple modes through the idea of “translation”. Finally, specific links between science as a field of study and the characteristics of effective science classrooms discussed earlier in this literature review suggest a positive connection between multi-modal usage, writing tasks, and science learning.

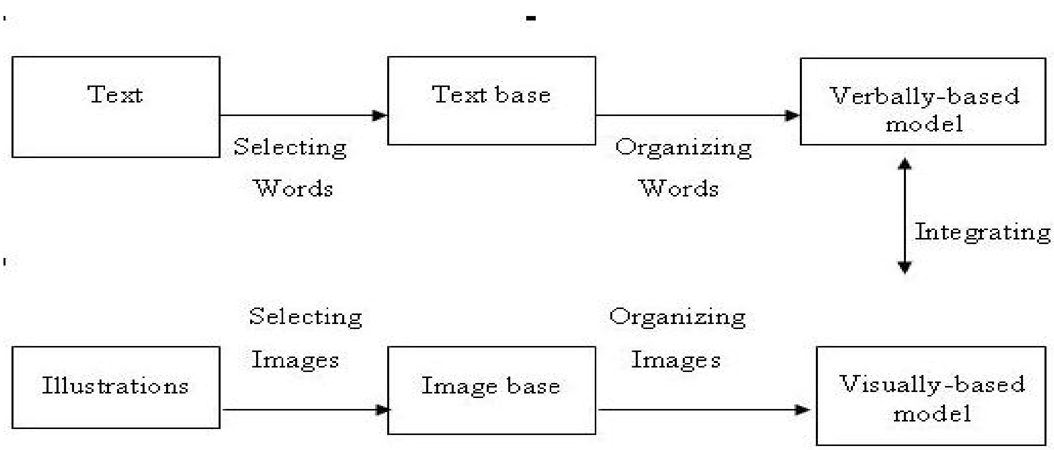
Cognition, Learning, and Multi-Modal Representations

Similar to the cognitive models proposed to describe what researchers believe takes place during writing-to-learn tasks, different models have been suggested to account for cognition and learning from the use of multiple modes of representation. Mayer (1997) focused on how students learn from viewing multi-modal presentations in what he termed “multimedia learning”. He called his model of cognition “a generative theory of multimedia learning” and built it on the twin ideas of dual coding theory (Paivio, 1986; Clark & Paivio, 1991) and generative theory (Wittrock 1974, 1989). The dual coding aspect of Mayer's theory asserts that verbal and visual cognitive processes take place in separate systems. The generative aspect of Mayer's theory is based on the contention that "meaningful learning" takes place in a series of three steps. First, the learner must select what is considered relevant from presented material, then the learner must organize selected information into some sort of meaningful representation, and finally the representation must be integrated with other representations referring to the same concept. Mayer's research to test this theory was based on the premise that the design of the learning environment a student was in would impact the degree to which the potential cognitive actions in the verbal and visual areas would take place and that:

> meaningful learning occurs when a learner builds coherent mental representations of a cause-and-effect system in verbal short term memory and in visual short term memory, and builds systematic connections between the verbal and visual representations. (Mayer, 1997, p. 6)

Figure 6 illustrates on overview of Mayer's model.

Figure 6. Mayer's Generative Model of Multimedia Learning

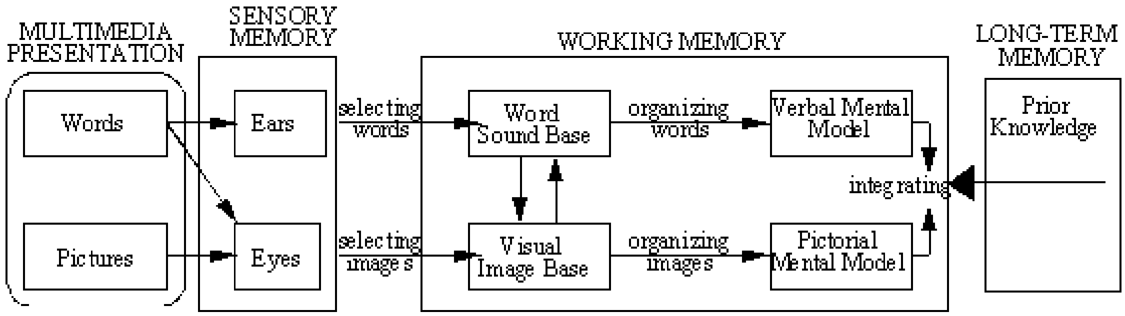


Source: Mayer, R. (1996). Learners as information processors: Legacies and limitations of educational psychology's second metaphor. *Educational Psychologist, 31*(3/4), 151-161.

Sweller, Merrienboer, and Pass (1998) offer another cognitive model that has described for some why utilizing multi-modal representations impacts learning. The model, termed Cognitive Load Theory is primarily concerned with cognitive architecture and the difference between working (conscious) memory and long-term memory. Sweller et al. (1998) assert that the goal of instruction is to create and automate schema, the organized long term storage units in human memory. This process of schema creation and automation is initiated in working memory through a process of selecting relevant material, similar to Mayer's selection process (Mayer, 1997). However, working memory has limited storage capacity, and this constraint may ultimately lead to less schema construction or automation and therefore less learning. The key issues impacting learning considered with the Cognitive Load Theory are intrinsic, extraneous, and germane cognitive load.

Intrinsic cognitive load is a fundamental characteristic of the targeted concept from instruction and for the most part is unaffected by instructional practices. Extraneous cognitive load is defined as "unnecessary" cognitive requirements existing as a result of the particular type of instruction utilized with a particular concept but that occupy some capacity of the working memory. Germane cognitive load is cognitive load that is able to be manipulated by instructional practices and aids in attainment of conceptual learning. According to Sweller et al.(1998), to reach the stated goal of schema development, extraneous cognitive load must be minimized and germane cognitive load maximized. Mayer's theory can be integrated with Cognitive Load Theory to create an integrative model that takes into account the capacity limit of working memory. Figure 7 illustrates a model based on this integration.

Figure 7. Integrated Cognitive Load and Generative Model of Cognition



Source: imej.wfu.edu/articles/2000/2/05/index.asp

In addition to the previously mentioned models, researchers have also categorized several levels of multi-modal "resources" (Hand et al., 2008). Mayer (1997) in setting up the discussion of his generative model noted three categories of representations: delivery media, presentation modes, and sensory modalities. Delivery media referred to the general "system" of informational presentation (for example a textbook or a computer), presentation mode referenced the actual mode utilized within the delivery media (text, picture, graph, etc.), and sensory modalities meant the cognitive processes utilized to interpret the representation (Mayer, 1997). Schnotz and Lowe (2003) offered a similar categorization with slightly different terminology in describing technical devices, semiotic or representational formats, and sensory modes. Bernsen (1993) and Pineda and Carza (2000) simplified this scheme into either representational modes (semiotic) or psychological modes (sensory). Key to all these categorizations is both the recognition of the specific aspects of each "resource" along with the interconnectedness of all resources when attempting to comprehend or utilize the modes.

A more holistic picture emerges considering all the aforementioned ideas related to cognition and multi-modal representations. From this holistic perspective, the cognitive action present when utilizing multi-modal representations becomes a dynamic activity composed of characteristics of cognitive architecture, characteristics of the instructional setting present in the learning environment, and characteristics of the particular modes of representation themselves. Any learning resulting from this cognition is dependent upon the interaction of all these factors. In addition, some researchers have broadly generalized learning itself as an inherently multi-modal activity. Alverman (2004) states explicitly that "all meaning-making is multi-modal" and that "learners must develop multiple meanings as a factor of learning" (p. 227). Airey and Linder (2006) posit that "learning is about allowing students to use different modes appropriate to the topic to make meaning for themselves." Upon realizing these connections, the necessity of determining the core or critical aspects of multi-modal instruction that leads to learning arises. In dealing with this question, researchers tend to focus on the idea of translation.

Translation and Multi-Modal Representations

In summarizing his position and promoting the use of multiple modes of representing information, Mayer (2003) asserted that knowledge construction is the result of movement between different modes of representation. Because different modes are dealt with by different aspects of working memory, the integration of the differential cognition resulting from exposure to multiple modes leads to deeper understanding than would be attained through exposure to a single mode of representation. Pineda and Garza (2000) assert a similar sentiment when they state that the construction of "rich understandings" is accomplished when students develop an understanding of how different modes all relate to a similar concept. This realization initiates a process of reasoning and inference making that they characterize as "incremental" and dictated by "inference constraints". Bernsen (1993) also points out that since multi-modal representations are built on unimodal representations and because there is no "universal decoding system" for all modes, the key to developing these types of deep and rich understanding is students' ability to attain additional information through the progressive consideration of modes. Gunel et al. (2003) state this in another way by saying that as students move between modes, they are forced to "cognitively process information language of one mode in a way that it can be utilized with another mode" (Gunel et al., 2003). Pineda and Garza utilize the term "translation" in describing this ability and assert that this causes greater cognitive activity for the author. This characteristic of translation also provides a direct connection between multi-modal use and writing-to-learn activities.

Translation has been posited as a desirable outcome of writing-to-learn activities due to its potential to increase cognitive activity and thus, conceptual understanding. Prain and Hand (2005) ground much of their reported benefit from alternative or non-traditional writing activities in the necessity of students moving from one "language" to another. The languages most often associated with these tasks are the "everyday" language of the student, the "classroom" language utilized by the teacher, and the "canonical" language of the discipline of science dealt with. Yore and Treagust (2006) find a parallel between science learners moving through and between these various languages as a part of writing-to-learn tasks and English as a second language learners moving between and through their "home" language and English. This translation process is seen as requiring not only greater cognitive action, but increases the likelihood of interaction between the rhetorical and content spaces likely leading to a greater degree of dispositional dialectic that Galbraith (1999) hypothesized in his knowledge constituting model of writing described earlier.

Manipulating the audience written for has generally been viewed as the key aspect of writing-to-learn activities leading to this translation. Gunel, Hand, & McDermott (in press) assert the benefit they detected when Year 10 biology students wrote to Year 3 students was due to the set up of the task encouraging the translation of the science information dealt with from the language it was delivered in the classroom to the language the younger students would understand for a writing task and back to the language of the classroom for assessment purposes. Additionally, McDermott and Hand (in review) note that students involved in alternative writing tasks perceive this translation process and indicate that it initiates a cycle of recognizing gaps in personal understanding and filling in those gaps to clarify the understanding.

Initial studies investigating multi-modal representations in science writing indicate that requiring students to utilize modes other than text in their writing initiates a similar cognitive translation process for the author (Gunel et al., 2003; Hand et al., 2008). The consideration of the different modes available to the author for description and discussion of the target concept in their writing task involves a personal translation process between different modes. However, there are indications that this translation is not a guaranteed outcome of writing tasks that require multi-modal representations. Rather, the translation is more likely to occur if integration of the multiple modes is emphasized as opposed to mere inclusion of the multiple modes (McDermott & Hand, 2008). If this integration can be accomplished, the translation can be cognitively beneficial both in terms of conceptual development and in terms of modeling "true" scientific learning, as Lemke (1998) summarizes:

> "to do science, to talk science, to read and write science, it is necessary to juggle and combine in canonical ways verbal discourse, mathematical expression, graphical-visual representations, and motor operations in the "natural" world." (Lemke, 1998, p. 90)

This connection to science will be elaborated in the next section.

Connections between Multi-Modal Representations and Science

In addition to the previous noted cognitive and learning benefit resulting from the use of multiple modes of representation, multi-modal usage has been promoted due to its specific connections to science. Hand et al., (in review) note that to truly realize the dual goal of science literacy offered earlier in this review with its derived and fundamental senses, representing concepts multi-modally is necessary. Gunel et al. (2006) broaden the objective of multi-modal usage, and the translation alluded to previously, by asserting a goal of not only translation of conceptual ideas by students in the context of the classroom, but also ideally translation through and with typical representational modes used in science.

Further, if one of the goals of science literacy is to emulate characteristics of practicing scientists in the classroom, then multi-modal usage can provide an avenue to pursue this goal. Kozma (2003) asserts the following in terms of how scientists act:

> Scientists coordinate features within and across multiple representations to reason about their research and negotiate shared understanding based on underlying entities and processes. (Kozma, 2003, p.296)

This summation obviously entails habits of the mind that would be essential to encourage in science learners. Other researchers have asserted that some aspects of multi-modal use in relation to science may be unique to the discipline, as Lemke (1998) summarizes:

> Scientists combine, interconnect, and interpret verbal text with mathematical expressions, quantitative graphs, informational tables, abstract diagrams, maps, drawings, photographs and other unique and specialized visual genres seen nowhere else. (Lemke, 1998, p. 89).

In concurring with Lemke's idea, diSessa (2004) goes even further in proposing that the utilization of "new forms of representation" as a part of normal science discourse is one way that scientific understanding in general has been advanced historically. The intent of using multi-modal representations as a part of writing-to-learn activities in science would be to encourage the personal advancement of scientific understanding in the minds of the students producing the multi-modal products.

The aforementioned qualities of multi-modal representations when applied to the context of writing tasks have been linked in the literature to a description of cognition during writing consistent with Galbraith's model of knowledge constitution. Pineda and Garza (2003) describe a cyclical process of cognition resulting from the translation between multiple modes of representation proceeding until all possible connections to concepts are exhausted. This process is quite reminiscent of the dispositional dialectic that Galbraith describes. Seufert (2003) offers a similar view when asserting that deep conceptual understanding in science is only truly realized when students are able to produce connections both within and between different representations. The building of this deep conceptual understanding is contrasted to simple recall by diSessa (2004) who points out that the ability to build a "meta-representational competency" is evidence that knowledge has been generated as opposed to simply recalled.

Two main issues emerge from the overall consideration of the literature discussed in this chapter in terms of the potential benefit of the utilization of multi-modal representations in science writing tasks. One issue aligned with diSessa's "meta-representational competency" idea is that students must develop some understanding of the use of multiple modes of representing information as a technique to communicate about science. It is unlikely that benefit will result if students do not have some "competency" associated with appropriate ways to utilize different representations. It is the contention of the researcher that one main aspect of this multi-modal competency is an understanding of how to integrate alternative modes outside of text with the text itself. The second main issue deals with whether or not the action of utilizing multiple modes of representing science information and embedding the modes in text can lead to greater science conceptual understanding. It would appear that to make multi-modal writing tasks an effective component of a productive learning environment, both of these issues must be considered. The research presented here will explore both these issues.

Previous Research and Goals of the Present Study

Previous research dealing with using multiple modes of representing science information has indicated mixed results (Ainsworth, 2006). However, much of this research has investigated using multi-modal representations created by others as an aid to learning rather than utilizing student created multi-modal products as learning tools. In exploring student created multi-modal products, Gunel et al.(2003) showed improved performance for physics students who created power point presentations with multi-modal representations versus students who produced text only summaries. In addition, this study showed increased benefit for students employing multi-modal strategies during the second round of testing, indicating some benefit from the consistent use over time of the strategies. Hand et al. (2008) also found benefit for students utilizing multi-modal representations in writing, particularly when a specific sequence of modal use was employed. Clearly, student created multi-modal writing tasks can be beneficial, but specific characteristics of these tasks need to be identified to make this benefit more likely.

Yore and Treagust (2006) report that little research has explored cognition in students as they think through the transformation from one mode to another or as the students move between different representations in their writing. One possible reason for this lack of research is the difficulty in setting up writing tasks that explicitly call for students to engage in this transformation. A pilot study leading to the study reported here indicated that many students, when required to utilize multiple modes of representing information in a writing task will simply add on additional modes after text has been completed, rather than integrating the multiple modes and text together (McDermott & Hand, 2008). This appears to indicate a low level of multi-modal competency, leading to little benefit in terms of science learning. If the hypothesized benefits in terms of student knowledge generation are to be realized, not only must a degree of match between multi-modal writing tasks and ideas on effective science learning environments be demonstrated, but questions about how to improve multi-modal competency and ultimately how to improve positive cognition leading to greater conceptual understanding must be addressed.

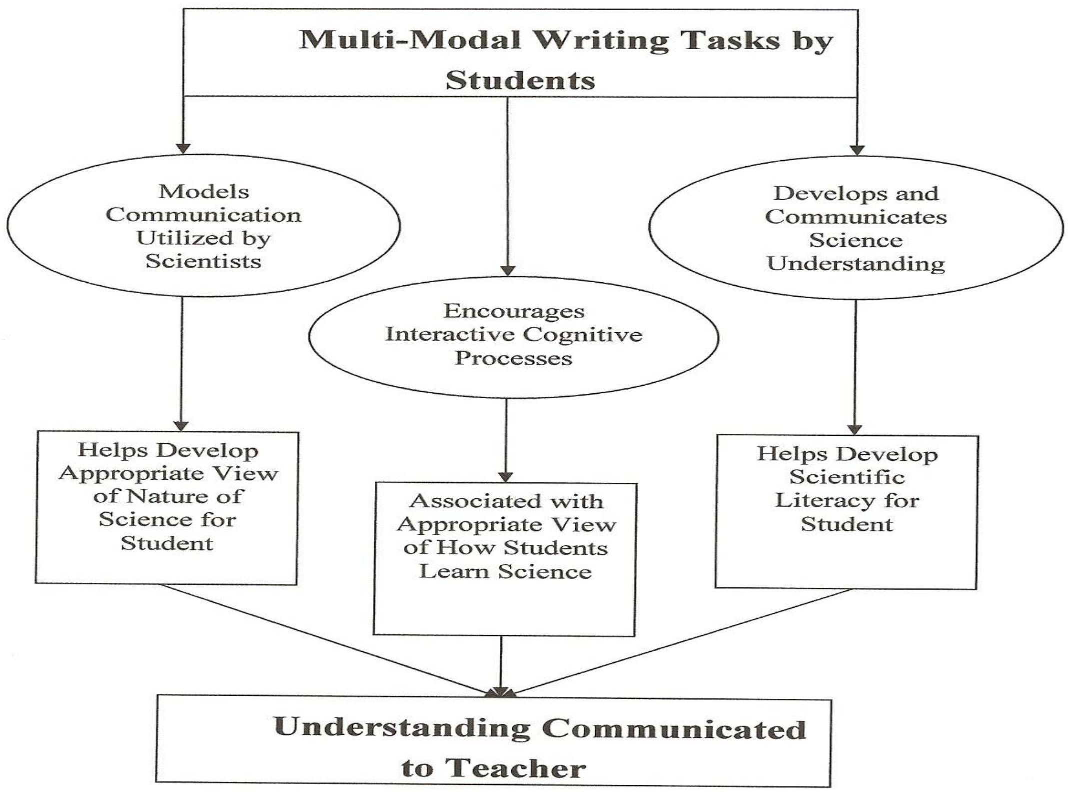
Recalling figure 1 developed earlier in this chapter to represent the interactive nature of the science learning environment is instructive in considering whether multi-modal writing tasks match desired attributes of an effective science learning environment. The previous sections of this chapter have provided references to research literature that highlight the connections between cognition and learning and multi-modal tasks, as well as connections to the way science is done by practicing scientists. Further, the multi-modal writing tasks would provide a vehicle for students to construct their own knowledge and for the students to make that knowledge known to the instructor in a way that would give the instructor valuable feedback. Figure 8 provides a representation of how multi-modal tasks integrate the features of an effective science learning environment. Key to this integration would be the effective promotion of the cognitive processes described in this literature review. However, unless students effectively embed and integrate the multiple modes in text, it is likely the desired cognitive results will not follow. Therefore, promotion of a solid multi-modal competency, including the ability to embed and integrate modes with text would seem an essential component of making these tasks a critical part of an effective science learning environment.

The research questions guiding the study presented here are listed below and stem from consideration of the dual issues of multi-modal competency and science conceptual understanding:

- Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to a greater degree of embeddedness in student writing?

- Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to greater conceptual understanding as measured by end of unit assessments?
- Can correlations between degree of embeddedness in writing and student performance be detected in a variety of classroom settings?

Figure 8 – Multi-Modal Tasks and Science Learning Environment



The first question deals with the idea of multi-modal competency by asking if specific instruction designed to promote awareness of the issue of embedding multiple modes in text leads to a greater understanding of how to effectively use multiple modes, as manifested in actual student writing. The second question is focused on the conceptual understanding of science and asks whether this can be improved through participation in activities that have the intent of improving multi-modal competencies. Finally, the third question seeks to explore whether positive benefits from higher degree of embeddedness and student science understanding can be realized in different settings, dealing with different topics, and taught by different instructors. The next chapter will describe the procedures used to investigate these questions.

CHAPTER THREE

METHOD

This chapter has three main purposes. First, a justification for the general methodology utilized in this study, grounded in research literature, will be provided. Second, an overview of the methodological procedures, including explanations of the participants, research protocol, and data collection will be given. Finally, a description of the data analysis techniques and instruments will be provided. In combination, discussion in these three areas will ideally provide a justification for the appropriateness of the particular methodological framework employed as a valid way to answer the research questions guiding this research.

While the main thrust of the research discussed here is to further the understanding of the particular intervention of utilizing multiple modes of representing science information in writing-to-learn activities, additional concerns impacted the choices associated with the methodology employed. The major issue was sensitivity to the students involved and their experiences as participants of the study. Obviously, the participating students were first and foremost actual students in actual schools attempting to learn chemistry. They were not simply research participants. The same was true of the teachers involved, all practicing teachers who agreed to allow their students to be utilized for purposes of this study. Therefore, it was a particular concern of the researcher to design a study that on the one hand allowed for meaningful research analysis, but on the other hand did not in any way disrupt the opportunity for the students involved to gain conceptual understanding and for the teachers involved to provide their students that opportunity. These practical concerns along with the philosophical issues relevant to research design combined to impact development of the study.

Justification of Research Methodology

Three areas of general interest when inspecting any research perspective are epistemology, ontology, and the particular methodology that stems from the first two. Epistemology refers to beliefs about truth, including what "counts" as truth, what can lead to the construction of knowledge or truth, and how the researcher and the participants interact in this search for truth (Grbich, 2007; Denzin & Lincoln, 2003). Ontology refers to beliefs about the nature of being and what constitutes "reality" (Denzin & Lincoln, 2003). Methodology refers to the actual research practices that are employed as a result of particular views on epistemology and ontology. Denzin and Lincoln (2003) state the combination of these three factors results in a paradigm for a researcher to function under. Guba (1990) defines paradigm as a "basic set of beliefs that guide action" and Denzin and Licoln (2003) posit that this "interpretive framework" regardless of how implicit or explicit it is for a particular researcher, will ultimately decide how they do research and what they believe about the research process itself. This underscores the importance of the researcher's awareness of what paradigm is guiding their work.

This research study has been designed with an ontological belief that is postpositivist, in line with what Denzin & Lincoln (2003) call "critical realism". With this perspective, an objective reality is assumed to exist, however, our ability as humans to obtain an understanding of this reality is questioned. It is not assumed, as with positivism, that through utilization of proper scientific methods, the true reality of nature can be ascertained, but rather that through careful use of experimental techniques, a meaningful interpretation of the true reality can be gained. Ideally, through analysis of multiple studies or multiple locations employing the study procedure, the degree of match between the interpretation of reality and the true nature of reality can be increased. The epistemological perspective guiding this research is aligned with an interactive-constructivist view, similar to that described in chapter two, in which the nature of knowledge in the classroom (what students believe and understand about science) and the nature of the understanding of what happened in the classroom gained through research are both personally and socially constructed. The goal of the methodology and analysis described in this chapter will be to facilitate an interpretation of the data by the researcher that describes the nature of the reality of the studied interventions as accurately as possible, with the understanding that the interpretation of that reality will also have a socially constructed aspect. This search for a balance between the practical constraints of human constructed understanding attempting to interpret and evaluate a situation that does indeed have an ultimate reality is what necessitates careful consideration and implementation of research procedures. The aforementioned issues characterize the research paradigm that guides this research and has been manifested in a practical sense in the utilization of a quasi-experimental, quantitative design.

Berliner (2002) in debating the merits of the U.S. government's push for solid research backed practices in education with the establishment of the controversial No Child Left Behind legislation notes the many difficulties associated with educational research. In particular, Berliner notes the difficulties associated with the multitude of variables present in any educational context (such as socioeconomic status, student prior experiences, teacher action, school setting and countless others), the multitude of interactions, and difficulties associated with changing social dynamics. In summarizing the challenges facing educational research, Berliner notes:

> We should never lose sight of the fact that children and teachers in classrooms are conscious, sentient, and purposive human beings, so no scientific explanation of human behavior could ever be complete. In fact, no unpoetic description of the human condition can ever be complete. When stated this way, we have an argument for heterogeneity in educational scholarship (Berliner, 2002, p. 20)

Berliner asserts that while he is certainly a proponent of randomized experimental design, to believe that this type of research is the "only scientific research…that yields trustworthy evidence" is a misconception that hampers educational research.

Berliner's recommendation is to promote discourse, evaluation of evidence, and debate through the utilization of multiple and varied experimental techniques that view educational interventions from different perspectives (Berliner, 2002). This philosophy is employed in this study. The bulk of the data analysis considered will be quantitative, but the analysis included tests from several different perspectives and utilized different techniques, in an attempt to thoroughly explore the research situation. Further, while there was no formal qualitative component, the interpretation of the data presented in the discussion of the results was aided by member checking procedures in which the participating teachers were given samples describing their classrooms and asked to verify the accuracy of the descriptions. In addition, feedback from teachers was utilized to assess the fidelity of the research in terms of how well the study procedures were implemented at different sites.

A quasi-experimental design was necessitated for this study due to the use of convenience, rather than random, sampling. Kerlinger (1980) calls the quasi-experimental design a "compromise position" useful for research in education because the random assignment of students to treatment groups in any sort of meaningful, school like situation is impractical at best, and impossible in many cases. The classes that were utilized in this study were pre-existing classes that had been established at the beginning of the school year, prior to the implementation of the study. Teacher involvement was predicated on the ability to use intact classroom groupings. In addition, from a pragmatic standpoint, if the techniques investigated here are to be more fully utilized in the future, they will most certainly be utilized in pre-existing, non-randomized settings, therefore this type of set-up is a more practical context.

Johnson and Christensen (2007) note that the major issue with a quasi-experimental design is whether or not a causal inference can be made due to the possibility of confounding variables not being controlled for due to the non-random design. Just as in "strong" experimental designs, three requirements must be met to establish a causal inference: the cause must precede the effect, the cause must covary with the effect, and alternative hypotheses must be shown to be implausible (Johnson & Christensen, 2007). Typically with the quasi-experimental design, the difficulty becomes showing that alternative hypotheses are invalid. One way to account for this is to utilize analysis of covariance to account for variables of concern in the data analysis. Another way is to accumulate as much evidence as possible to support the contentions shown from the research. This accumulated evidence may also include qualitative information to supplement quantitative data.

The data analysis utilized in the research methodology also stems from a pragmatic perspective. Obviously, the first concern of this research is to answer the research questions driving the study. However, in addition to improving our understanding of an aspect of science education, this study is also designed to hopefully provide practical information to practicing teachers. Therefore, the data analysis techniques were intended to address both the research and practical concerns. This perspective led to the main questions for the data analysis. First, information was sought to determine if a claim could be substantiated about whether the utilization of the pedagogical technique in a classroom of having students participate in writing tasks in which they embed multiple modes of representing science information with text leads to differential performance on writing tasks and on post unit assessment. This information was gathered using analysis of variance or covariance (if baseline testing indicated significant differences on key characteristics). In addition, effect size calculations were utilized to explore all differences in group performance. The second main question for data analysis was whether a relationship exists at the level of individual students between level of embeddedness in writing and performance on post test assessment. This question was investigated with correlation and regression analysis.

The overall intent of the research design and analysis is to give as complete and clear a picture of how using these writing activities with multiple modes of representing effects student performance. Ideally, what will be gained will be not only a complete picture of statistical and perceived differences in differential treatment, but also information and relationships that will be dually beneficial to researchers who wish to study these techniques further and teachers who wish to explore use of the techniques in their classrooms.

Research Design

Participants

Four unique groups of students were utilized in this study. Each group consisted of an existing set of general chemistry classes taught by one teacher. Prior to describing the progression of activities utilized in this research design, characteristics of each group will be described. For purposes of this study, each participating group will be referred to as a "site".

Site One

Site one included thirty-seven students in chemistry courses taught by the same instructor at a parochial high school in the middle-east United States who agreed to participate in the study procedures. Five other students did not consent to having their data used for the study. The students were all grade 11 females. The total enrollment at the school attended by these students is 740 and it is located approximately 10 miles outside of a large urban city. Demographically, the school includes approximately 5 % minority students and 95% white students. The school is classified as having middle to upper middle class socio-economic status. The instructor at this site has twelve years of secondary school teaching experience, all but two months of which have been at the current school. The instructor has a bachelor's degree in chemistry and a master's degree in secondary education. This instructor has taught a variety of courses including physics, marine biology, general biology, and geophysics, along with chemistry and advanced placement chemistry.

Site Two

Site two included sixty total students in general chemistry courses at a small, rural school in the southern United States who participated in the study and three who did not. The participant students included 21 females and 39 males. The school these students attend is a parochial school of approximately 396 total students. The school has less than 1% minority students and is middle class in terms of socio-economic status. The teacher for these students has taught chemistry for five years after a laboratory career in chemistry. This instructor earned a bachelor's degree in chemistry and was working on a master's degree in secondary education at the time of the study. All five years of teaching for the instructor have been at the current school and the instructor has taught chemistry courses exclusively.

Site Three

Site three included students enrolled in three separate general chemistry courses taught by the researcher. Eighty-three total students in these three classes including 41 males and 42 females participated in the study. Seven students did not provide consent for their data to be utilized. The students were in grades 10-12. The students attended a large high school of approximately 1800 total students in the midwest United States. The school enrollment includes 1% American Indian students, 9% Asian students, 5% Hispanic students, 10% black students, and 75% white students. Approximately 16% of the students are eligible for free and reduced lunches. Approximately 80% of the students from this school attend college after high school. The instructor has taught high school science for fourteen years, twelve of which have included teaching chemistry. The school year in which the study took place was the ninth year in which the instructor had been at the current school. The instructor earned a bachelor's degree in biology and secondary education and a master's degree in science education. The instructor was finishing coursework to earn a doctoral degree in science education at the time of the study.

Site Four

Site four included students enrolled in five separate general chemistry courses taught by a colleague of the researcher at the same school as the students in site three. One hundred nineteen total students were enrolled in these classes with 55 females and 64 males participating in the study. Ten students did not provide consent. Students were in grades 10 – 12. The teacher for this course was teaching for the second year in the district. This instructor had a total of 12 years of secondary school teaching experience, all of which have been teaching chemistry. The instructor has a bachelor's degree in secondary education and chemistry and a master's degree in organic chemistry. The instructor also has 3 years of post-secondary teaching experience and two years of chemistry lab work experience.

Research Procedure

The teachers at the four sites were given identical instructions for the procedure to follow. Initially, all classes were given a standardized baseline science competencies test. The test consisted of twenty-one multiple choice questions covering a variety of science concepts typical of an overall high school curriculum in the United States. This baseline competency test has been frequently used in research protocols dealing with writing to learn. Along with the baseline competency exam, students were also asked to provide a writing sample that covered a science topic in which they were asked to provide at least one mode of representation other than text. This sample was designed to provide the researcher with a baseline "embeddedness" score for each student to detect any significant differences in tendency to embed that existed in the treatment groups prior to participation in the research activities. For the baseline embeddedness activity, students were asked to write a letter to the principal of their school requesting that a proposed building project on school grounds be halted due to the negative impact on the environment. Students were required to use at least one mode other than text in their letter. Students were given one class period to complete both baseline assessments.

Following the baseline assessments, teachers were asked to begin instruction in the first unit for which a writing task would be completed. At each site, half of the classes were assigned to the treatment condition and half to the control condition. Teachers were asked to randomly assign classes to treatment conditions by generating random numbers for each class and assigning the classes with the lowest numbers to the treatment condition. Prior to the assignment of the writing task for the first unit, teachers were asked to provide a lesson focusing on how common sources of science information integrate multiple modes of representing science information to the treatment classes. The lesson culminated in the production of a student-teacher co-generated checklist for assessing the "embeddedness" of any written piece dealing with a science topic. The lesson outline was provided for the teachers by the researcher and is shown in Appendix A of this report. A sample student–teacher co-created checklist for embeddedness is provided as Appendix B. This sample was given to teachers as a model, but they were instructed to generate their own checklists with their classes. The teachers were instructed that the control classes were not to receive this embeddedness lesson. Although no common alternative task was given for the control classes, teachers were asked to make sure that the students in these classes did not receive additional instruction dealing with the concepts covered in the unit of study during this time in order to assure equal time on task for all classes.

All classes at each site (control and treatment) received identical writing tasks for the first unit of study. Teachers were asked to assign the writing tasks so that final drafts would be collected prior to the end of the unit assessment. All writing tasks were to involve writing to an audience outside of the instructor, receiving feedback from this audience after completing an initial draft, revising, and creating a final draft for evaluation by the instructor. Additionally, students in the treatment groups were asked to complete a "self-assessment" of their writing using the checklist for embeddedness created as a part of the lesson on embeddedness. Students submitted their final drafts, the feedback sheet they received from their authentic audience, and their self-assessment of embeddedness (if in the treatment groups) to the instructor. A sample writing task assignment sheet is shown in Appendix C.

Following the writing tasks, both treatment and control students at each site participated in an identical end of unit assessment consisting of multiple choice and extended response conceptual questions. Teachers were instructed to mark the assessments in whatever manner was most appropriate for their classroom and then send all assessments to the researcher for evaluation purposes for this study. The researcher scored all assessments utilizing rubrics created for each question. The researcher was blind to treatment condition of students when assessing these evaluations. The sequence of writing task and end of unit assessment was then repeated for the next chronological unit of study with the same students participating in the same treatment conditions. In this way, data could be analyzed for two consecutive units. The research procedures are summarized in table 2.

The original intention of the researcher was to have all participating sites utilize identical writing tasks for identical units of study and assess students with identical end of unit exams in order to facilitate overall analysis of the relationship between degree of embeddedness and student conceptual understanding. However, practical limitations due to scheduling conflicts and unanticipated outside factors did not allow for this intention to be realized. Due to the lack of conformity across groups, data analysis was calculated by site, rather than pooled. Interpretation of the data will include discussion of both characteristics of each site and of overall trends among all sites.

Table 2. Research Design for Each Participating Teacher

Control Classes	**Treatment Classes**
Baseline Test & Baseline Embeddedness	Baseline Test & Baseline Embeddedness
	Embeddedness Activity & Construction of Checklist
<u>WRITING ASSIGNMENT ONE</u>: Requirements: Must include at least ONE mode of representing other than text	<u>WRITING ASSIGNMENT ONE</u>: Requirements: Must include at least ONE mode of representing other than text & Students Self Assess Writing with Embeddedness Checklist
ASSESSMENT	ASSESSMENT
<u>WRITING ASSIGNMENT TWO</u>: Requirements: Must include at least ONE mode of representing other than text	<u>WRITING ASSIGNMENT TWO</u>: Requirements: Must include at least ONE mode of representing other than text & Students Self Assess Writing with Embeddedness Checklist
ASSESSMENT	ASSESSMENT

Instruments, Data Collection and Analysis

Baseline Science Understanding and Embeddedness Scores

Prior to initiation of treatment for the study, all students were assessed for level of scientific understanding using a baseline competency of science information assessment. The assessment was a multiple choice test consisting of 21 total questions covering a variety of science concepts. Questions for the test were collected from the National Assessment of Educational Progress (NAEP) and the Trends in International Mathematics and Science Study (TIMSS). This baseline test has been consistently utilized as a measure of student prior knowledge in a variety of writing-to-learn studies over the course of more than a decade. A Cronbach's alpha coefficient for the study has been calculated at 0.7. The intent of using the study in this context was to determine equivalency of groups prior to initiation of the study and to be used as a measure in regression analysis. Analysis of variance (ANOVA) tests were applied to the baseline science scores to determine if significant differences between participating classes were present. No significant differences at the level of alpha = .05 were detected between treatment and control groups for any sites.

In addition to analysis of baseline science understanding, students were also assessed for a baseline level of embeddedness. Student responses to the baseline embeddedness question (described earlier in this methods section) were analyzed using ANOVA to detect differences between classes in baseline embeddedness scores. Significant differences in terms of embeddedness were used to determine the necessity of utilizing a covariate for further analysis of writing scores. A significant difference at $p < .05$ was detected at site four between the treatment and control group. No other significant differences between treatment conditions were detected.

Embeddedness Rubric

A researcher generated embeddedness rubric was developed as a part of a pilot study leading to the research reported here to provide measures of degree of embeddedness utilized by students in their writing. Face validity of this measure was attained through consultation with researchers who had developed multi-modal studies in the past as well as research literature related to the use of multiple modes of representing information. In addition, classroom experiences with the utilization of the rubric in the pilot study informed the construction of the assessment tool. From the aforementioned sources, it was determined that three sub-categories of assessment could be combined to indicate how well students had integrated multiple modes with text in their writing.

The first writing assessment sub-category was the text the students produced. Of particular interest was whether the text covered the required topics from the assignment, was accurate, was complete, and was grammatically correct. This first area was assessed using a point scale ranging from 0-3 for each text factor (covered required topics, accuracy, completeness, grammar). A specialized rubric for each particular writing task was created to guide text assessment. These four factor scores were combined and referred to as the “text” score. Obviously, the text of the written task itself is critical in both communicating accurate science information and reflecting student conceptual understanding. Further, application of strategies to appropriately integrate alternative modes of representing information in text that is inaccurate or incomplete would not reflect a solid multi-modal competency or promote sound scientific understanding.

The second main area of assessment was the overall number of modes outside of text that were utilized and the number of science topics that were addressed through utilization of these modes. This assessment involved a count of the total number of appropriate representations outside of text, a count of the total number of different types of modal representations used (for example 3 pictures and 2 charts would be two types of modal representations, but five total representations), and a count of the number of science topics referred to. These three scores were added to form the “modal” score.

Finally, an “average embeddedness score” was determined for each piece of multi-modal writing. This score was determined by assessing each unique use of a mode outside of text in the student writing with a checklist of several key factors. The key factors were whether or not the modal representation was accurate, complete, next to the text that referred to it, referenced in the written text, contained a caption, or was an original item created by the author and not copied from another source (such as cutting and pasting on a computer). For each key factor used with a particular representation, one point was awarded. A total for each mode was calculated, and then a total embeddedness score for all the instances of modes in the text was calculated. The average embeddedness score was found by dividing this total embeddedness score by the number of modal representations other than text.

Two overall or “grand total” embeddedness scores were calculated for each writing sample. The first was termed “grand total (raw)” or GTR, and was calculated by adding the text score, the modal score, and the total embeddedness score together. The second overall score was called “grand total (average)” or GTA and was found by adding the text score, the modal representation score, and the average embeddedness score together. A sample of this rubric is included in Appendix D.

Although all writing scores were considered important for assessing student writing, the GTA was deemed the most appropriate score for describing degree of embeddedness. The rationale for the importance of this score stems from agreement amongst the researchers collaborating on several multi-modal projects that a well integrated piece of writing in total would have a well written text component, complemented by a varied use of multiple modes that are equally well integrated together. The overall embeddedness score reflected this position. A high number of modal representations or one particularly high embeddedness score for one representation could skew the GTR score, but a high GTA score would be more reflective of an overall well integrated multi-modal writing piece.

Inter-rater reliability for the embeddedness rubric was obtained by randomly selecting 30 writing samples and correlating the scores of the researcher with scores from an external observer. The external observer in this study was a licensed teacher who had previously been trained to analyze student data from writing-to-learn studies but had not previously worked with multi-modal representations. Pearson's correlation coefficients were .970 for text, .935 for MR, .947 for AEB, .974 for GTR, and .966 for GTA.

End-of-Unit Assessments

The end of unit assessments were teacher created assessments consisting of multiple choice, extended response conceptual questions, fill in the blank questions, and problem questions of different types based on the particular unit of study. Face validity for the end of unit assessments was determined in two different ways. First, consultation among the participating teachers took place to determine if the selected questions accurately reflected the material covered in the coursework employed by the teachers and matched overall curricular goals of each participating school. Secondly, the National Science Teaching Association (NSTA) and the National Science Education Standards (NSES) were consulted to verify that concepts dealt with and assessed were in line with accepted practice. Internal consistency was determined for all multiple choice questions using Cronbach's alpha and was calculated at .945 for site one (unit one), .963 for site one (unit two), .913 for site two (unit one), .966, for site two (unit two), .897 for site three (unit one), .863 for site three (unit two), and .549 for site four (unit one). Inter-rater reliability for assessment of the extended response conceptual questions was accomplished in an identical manner as the reliability assessment for the embeddedness rubric. Pearson's correlation coefficients for each are shown in table 3.

Table 3. Correlation Coefficients for Extended Response Inter-Rater Reliability Measures

Site	Unit	Assessment	r	N
1	1	Extended Response 1	.872	30
1	1	Extended Response 2	.963	30
1	1	Extended Response 3	.923	30
1	2	Extended Response 1	.824	30
1	2	Extended Response 2	.958	30
2	2	Extended Response 1	.957	30
2	2	Extended Response 2	.866	30
2	2	Extended Response 3	.804	30
3	1	Extended Response 1	.973	30
3	1	Extended Response 2	.986	30
3	1	Extended Response 3	.919	30
3	1	Extended Response 4	.955	30
3	1	Extended Response 5	.909	30
3	2	Extended Response 1	.965	30
3	2	Extended Response 2	.992	30
3	2	Extended Response 3	.882	30
3	2	Extended Response 4	.853	30
4	1	Extended Response 1	.919	30
4	1	Extended Response 2	.955	30

Analysis of Variance

Group level analysis included utilizing analysis of variance procedures to test for significant differences between treatment and control classes for each participating group. Analysis of this type was applied to comparison of scores on end of unit assessments. Analysis of variance was also applied to the assessment of writing scores for each unit as well. Specifically, student text scores, modal representation scores, average embeddedness scores, grand total (raw) and grand total (average) scores were compared between treatment and control conditions. Analysis of variance was also applied to the baseline science competency and baseline embeddedness scores. For site four, a significant difference between treatment and control was noted in the baseline embeddedness measure, therefore analysis of covariance (ANCOVA) was utilized for assessing the writing scores, with baseline embeddedness score as a covariate.

Mertler and Vanetta (2002) indicate three main assumptions that must be taken into account when utilizing analysis of variance procedures. First, an assumption of normality that supposes each variable and combination of variables are normally distributed, is necessary. Secondly, an assumption of linearity that there is a straight line relationship between variables is necessary. Finally, an assumption of homogeneity in which it is assumed that the properties of any one part of the data set is consistent with any other part, is needed. Each of these assumptions was tested for each group prior to analysis. Normality and linearity were tested through visual inspection of data plots. Homogeneity was assessed through Levene's test of equal variance.

Effect Size Calculations

Effect size calculations are gaining popularity, particularly in educational research (Gunel, et. al 2008). The effect size calculation allows for comparison between treatment and control conditions by size of the effect of the experimental condition in terms of standard deviation units. Thalheimer and Cook (2002) point out that effect size calculations allow for comparison across several independent studies. In the research presented here, analysis of the four separate sites will be considered as four separate studies with a related theme. Therefore, effect size calculations will be helpful in determining whether similar effects from treatment were present in the separate cases.

Cohen's d values were calculated for all scores previously mentioned in the analysis of variance overview. Cohen (1992) provides an often used standard for categorizing effect size calculations, with effect sizes between .20 and .49 categorized as small, effect sizes of .50 to .79 as medium, and effect sizes of .80 and greater as large. This categorization scheme is utilized in this study.

Correlation and Regression Analysis

Analysis at the level of the individual students was also undertaken with the data presented here with correlational and regression analysis. Thompson, Diamond, McWilliam, Snyder, and Snyder (2005) point out that correlational research is particularly useful in cases where subjects are not assigned to treatment groups. While these researchers admit that correlational measures cannot indicate causality, they can be used in conjunction with other statistical analyses to strengthen an argument of benefit from an intervention. Thompson et al. apply the term "logic based" analysis to a combination approach in which logic and research backed theory, along with a variety of data analysis techniques are utilized in conjunction to evaluate a particular intervention strategy. Osborne and Waters (2002) note that for measures such as regression analysis to meaningfully contribute to discussion, however, the same three assumptions noted earlier for analysis of variance must be tested. In addition, reliability of measures must be established. These factors will all be considered for each type of analysis.

For purposes of this study, it was reasoned that if the utilization of a greater degree of embeddedness of multiple modes in text was beneficial for conceptual improvement, this benefit would be manifested statistically in positive correlations between student embeddedness measures and test scores, regardless of treatment condition. Further, if emebeddedness is beneficial, the degree of embeddedness should be an effective predictor of overall assessment scores. In particular, regression analysis reported here tested whether student background science knowledge, as measured by the baseline competency measures or student embeddedness score was a better predictor of overall end of unit assessment score.

Summary

The methods and data analysis employed in this study were developed to help provide as accurate an interpretation of the study results as possible. The procedures were developed within the philosophical framework posited in this chapter, and after consultation with the research literature reported in the previous chapter. Further, the methods utilized were influenced by the pilot study run by the researcher that preceded this study. The results that were obtained from these procedures will be reported in the next chapter, followed by a discussion of the meaning of the results.

CHAPTER FOUR

RESULTS

This chapter will focus on the statistical results that were obtained from the methods outlined in the previous chapter and lead to the discussion in the next chapter. One main intention of this research was to test whether benefit from the embedding of multiple modes in writing could be shown in different geographic locations. To this end, results will be reported in two main sections. First, each site of the four studied will be viewed as an individual case and the results from the particular case presented. Following these individual cases, a summary of the findings from all four sites will be presented. Discussion of these results will take place in the following chapter.

Statistical Assumptions

The normality assumption for all tests was analyzed using visual inspection of graphical plots, as well as Kolmogorov-Smirnov tests. These indicated that the normality assumption was valid for all tests utilized. Linearity was assessed using Q-Q plots and examination of these from SPSS data indicated that patterns resembled linearity for each test. Finally, homogeneity was assessed with Levene's test for equal variance. When variances were found to be significantly different, adjusted means were utilized.

Common Abbreviations Utilized

Several of the data tables in this section will report scores for sub-sections of end of unit assessments. While the end of unit assessments were not all identical, several types of questions were utilized in most of the assessments. Table 4 lists the abbreviations that will be utilized throughout this chapter for the subsections of assessments analyzed. Reference to this table will aid understanding of the tables throughout this chapter.

Table 4. Common Abbreviations Utilized in Data Tables

Abbreviation	Explanation of Question Type
MC	Multiple Choice
ER	Extended Response Conceptual Questions
AS	Atomic Structure Question
EC	Electron Configuration Questions
COMP	Completion Questions (Fill in the Blank)
PROB	Problem Sets
LS	Lewis Structure Diagrams

Site One Results

Equivalency of Groups

Site one consisted of two separate chemistry classes taught by the same teacher. One class served as the control group while the other received the embeddedness lesson and served as the treatment group. Equivalency of the two classes was assessed using ANOVA procedures applied to the results of the baseline science competency test. No significant difference was noted between the two groups at the alpha = .05 level. Table 5 summarizes the results. All assessment scores reported in this chapter are percentages.

Table 5. Site One: Mean Baseline Scores by Treatment Condition

Condition	Baseline	SD	N
Treatment	52.38	11.67	18
Control	46.36	12.07	19
TOTAL	49.29	12.09	37

Note: No Significant Difference ($p > .05$)

Site one did not partake in baseline embeddedness testing due to a time constraint issue in the participating school district.

Unit One Results

Unit one for the first site involved the concept of atomic structure. All students in the study from both control and treatment classes were assessed with an identical end of unit exam consisting of 40 multiple choice questions (MC), a question over atomic structure that asked students to fill in a table with the appropriate number of subatomic particles using information from the periodic table (AS), and three extended response conceptual questions (ER1, ER2, ER3). Extended response question 1 asked students to discuss the current understanding of the structure of an atom and evidence for this model, extended response question 2 asked students to compare isotopes, ions, and neutral atoms of different elements, and extended response question 3 asked students to describe average atomic mass and how it is determined. Analysis of variance was employed to test for differences in these scores, as well as the overall test score (TOTAL) based on treatment. No significant differences at the alpha = .05 level were detected for any of the scores. Unit one results are summarized in Table 6.

Table 6. Site One: Unit One Mean Assessment Scores by Treatment Condition

Condition	MC	SD	AS	SD	ER1	SD	ER2	SD
Treatment	78.48	11.18	90.97	15.34	57.78	13.53	82.32	25.62
Control	72.76	12.71	91.44	10.66	48.42	16.75	77.99	17.22
TOTAL	75.54	12.18	91.22	12.97	52.97	15.79	80.10	21.51

Condition	ER3	SD	TOT	SD	N
Treatment	60.00	24.73	79.07	11.37	18
Control	47.36	20.23	74.16	10.76	19
TOTAL	53.51	23.12	76.55	11.19	37

All students also participated in a writing assignment for unit one in which they wrote a magazine article about the structure of the atom and the history of our understanding of the atom. It was the intent of the researcher to require all students, regardless of treatment condition, to use at least one mode other than text in their writing, but this requirement was not communicated to site one students for the unit one writing task. As a result, only seven total students (five from the treatment group and two from the control group) spontaneously utilized modes other than text. Limited analysis was applied to the writing samples due to this deviation from the intended procedure. Results, however, were collected for mean scores by treatment for the writing subcategories described in the methods section: text (TEXT), modal representations (MR), average embeddedness score (AEB), grand total-raw (GTR), and grand total-average (GTA). No significant differences were noted between treatment groups on any of these scores at the alpha = .05 level. Table 7 summarizes these results.

Table 7. Site One: Unit One Mean Writing Scores by Treatment Condition

Condition	Text	SD	MR	SD	AEB	SD	GTR	SD	GTA	SD	N
Treatment	10.11	.963	2.17	4.31	.611	1.04	7.50	13.92	5.61	9.65	18
Control	10.47	.841	1.16	3.48	.263	.806	2.84	8.67	2.47	7.43	19
TOTAL	10.28	1.64	1.64	3.89	.432	.929	5.11	11.61	4.00	8.61	37

Effect size data was also collected for analysis at the group level. Effect size differences were calculated between treatment and control groups for all categories analyzed in the previous section in reference to the end of unit assessment and writing scores. Thalheimer and Cook (2002) utilize a categorization scheme in which effect sizes with Cohen's d below .20 are termed negligible, those between .20 and .49 are small, those between .50 and .79 are medium, and those of .80 and greater are large. This scheme is used throughout the results reported for this research. These categories are reported along with Cohen's d in Table 8 for assessment from unit 1 and in Table 9 for writing scores from unit 1. All Cohen's d are reported in relation to the treatment group compared to the control, therefore a negative value indicates a greater mean score for the control group.

Table 8. Site One: Effect Size Calculations by Treatment Condition for Unit One Assessment

Question Type	Cohen's d	Effect Size
Multiple Choice	.11	negligible
Atomic Structure	- .04	negligible
Extended Response Question 1	.63	medium
Extended Response Question 2	.21	small
Extended Response Question 3	.58	medium
Total Test Score	.46	small

Table 9. Site One: Effect Size Calculations by Treatment Condition for Unit One Writing Scores

Writing Score	Cohen's d	Effect Size
Text	- .41	small
Modal Representation	.27	small
Average Embeddedness	.39	small
Grand Total (Raw)	.42	small
Grand Total (Average)	.38	small

Unit Two Results

Site one repeated the cycle of writing task followed by assessment for a second unit. The second unit immediately followed the first and dealt with the concepts of electron configuration and the periodic table. The end of unit assessment included multiple choice questions (MC), two extended response conceptual questions (ER1, ER2), and a question in which students were asked to determine electron configurations for several elements (EC). Extended response question 1 dealt with the emission spectrum of an atom and extended response question 2 asked students to defend the arrangement of the periodic table. Mean assessment scores for each section and for overall total score by treatment are shown in table 10 below.

Table 10. Site One: Unit Two Mean Assessment Scores by Treatment Condition

Condition	MC	SD	ER1	SD	ER2	SD
Treatment	89.28*	6.02	42.86*	17.73	47.14*	16.84
Control	83.14	7.68	32.35	10.33	31.76	12.37
TOTAL	85.91	7.54	37.10	14.88	38.71	16.28
Condition	**EC**	**SD**	**TOTAL**	**SD**	**N**	
Treatment	87.95	12.61	77.87*	6.36	14	
Control	87.50	11.48	71.75	6.85	17	
TOTAL	87.70	11.80	74.51	7.22	31	

*p < .05

These scores were also analyzed to determine if significant differences existed between treatment groups. The treatment condition (M = 89.28, SD = 6.02) outperformed the control group (M = 83.14, SD = 7.68) on multiple choice questions (t = 2.49, p = .016). Treatment (M = 42.86, SD = 17.73) also outperformed control (M = 32.35, SD = 10.33) on extended response question 1 (t = 2.060, p = .049). A similar result was detected for extended response question 2 (t = 2.93, p = .007) where treatment

(M = 47.14, SD = 16.84) outperformed control (M = 31.76, SD = 12.37). Finally, on total test score (t = 2.559, p = .016), treatment (M = 77.87, SD = 6.36) scored significantly higher than control (M = 71.75, SD = 6.85). Significant differences are noted in table 11.

Table 11. Site One: Unit Two Significant Differences by Treatment Condition ($\alpha = .05$)

Measure	Higher Score	t	p
Multiple Choice	Treatment	2.439	.016
Extended Response 1	Treatment	2.060	.049
Extended Response 2	Treatment	2.930	.007
Total Score	Treatment	2.559	.016

For this second unit, group one students also participated in a writing task in which they were required to use a mode of representing science outside of text. The writing task involved writing letters to the following year chemistry students about electron configuration and the periodic table. Table 12 summarizes the writing scores by treatment group for this writing task. None of the differences in writing scores were significant at the alpha = .05 level.

Table 12. Site One: Unit Two Mean Writing Scores by Treatment Condition

Condition	Text	SD	MR	SD	AEB	SD	GTR	SD	GTA	SD	N
Treatment	10.14	.86	4.86	1.70	3.26	.60	21.29	4.84	18.26	2.24	14
Control	10.23	.90	5.94	2.63	3.11	1.01	24.53	7.71	19.29	3.48	17
TOTAL	10.19	.87	5.45	2.29	3.18	.84	23.07	6.68	18.82	2.99	31

Effect size calculations were also gathered for unit two assessment and writing scores. Effect size scores are summarized in tables 13 and 14.

Table 13. Site One: Effect Size Calculations by Treatment Condition for Unit Two Assessment

Question Type	Cohen's d	Effect Size
Multiple Choice	.91	large
Extended Response Question 1	.77	large
Extended Response Question 2	1.02	large
Electron Configuration	.04	negligible
Total Score	.95	large

Table 14. Site One: Effect Size Calculations by Treatment Condition for Unit Two Writing Scores

Writing Score	Cohen's d	Effect Size
Text	- .11	negligible
Modal Representation	- .49	small
Average Embeddedness	.18	negligible
Grand Total (Raw)	- .49	small
Grand Total (Average)	- .36	small

Further analysis at the level of the individual student was undertaken to determine if correlations existed between the individual writing scores and the overall total score on the unit 2 assessment. This data is noted below in table 15. None of the correlations measured were significant at alpha = .05.

The final data analysis for group two was to apply regression analysis to determine if baseline science competency scores or grand total (average) emebeddedness score (GTA) was a better predictor of overall test score. Unstandardized and standardized scores were both determined. For the unit 2 assessment, baseline test score was a slightly better predictor of test performance than GTA. Table 16 summarizes the results.

Table 15. Site One: Correlations between Writing Scores and Unit Two Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.133	31
Modal Representations	.032	31
Average Embeddedness	.240	31
Grand Total (Raw)	.085	31
Grand Total (Average)	.131	31

Table 16. Site One: Regression Analysis Predicting Assessment Total for Unit Two

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.088	.112	.146
Grand Total (Average)	.340	.449	.141

Site Two Results

Equivalency of groups

Site two included three separate classes taught by the same teacher. Two of the classes were randomly assigned to form the control group while the third class was the treatment group. All participating students were assessed with the baseline science competency test prior to the initiation of the study interventions. All students also participated in a baseline writing task to assess their level of emebeddedness prior to the writing experiences for the study. ANOVA procedures were applied to these results to test for initial differences between the treatment and control group. No significant difference was detected at the alpha = .05 level for the baseline competency test or for the baseline embeddedness test. Table 17 summarizes the baseline testing data.

Table 17. Site Two: Mean Baseline and Baseline Embeddedness Scores by Treatment Condition

Condition	Baseline	SD	Embeddedness	SD	N
Treatment	43.12	17.32	4.61	1.75	18
Control	49.09	12.42	4.99	1.21	42
TOTAL	47.30	14.19	4.80	1.39	60

Unit One Results

Unit one for site two involved instruction with the concepts of ionic bonding and formula naming rules. All students in both treatment and control conditions were assessed at the conclusion of the unit with fill in the blank completion questions about ionic compound formation (COMP), problems involving formula writing, naming for ionic compounds, percent composition, and empirical formula determination (PROB), and multiple choice questions (MC). Table 18 summarizes these results.

Table 18. Site Two: Unit One Mean Assessment Scores by Treatment Condition

Condition	COMP	SD	PROB	SD	MC	SD
Treatment	58.89	31.04	55.55	23.98	66.66	22.31
Control	65.71	25.49	70.32*	22.49	77.59	15.02
TOTAL	63.67	27.18	65.89	23.74	74.32	17.66

Condition	TOTAL	SD	N
Treatment	61.11	22.31	18
Control	73.04*	15.02	42
TOTAL	69.46	18.18	60

* $p < .05$

Comparison of mean scores indicated that the control group (M= 70.32, SD = 22.49) outperformed the treatment group (M = 55.55, SD = 23.98) on the problem questions (t = 2.075, p = .026). For overall assessment total, Levene's test for equality of variance noted a significant difference (F = 5.547, p = .022). Therefore, adjusted means were utilized for analysis and indicated that control (M = 73.04, SD = 15.02) outperformed treatment (M = 61.11, SD = 22.31) for the overall unit one assessment score (t = 2.075, p = .049).

The writing assignment for unit one required students to write a magazine article describing the characteristics of ionic bonding and how to name ionic compounds. Analysis of writing scores for treatment and control groups is summarized in Table 19. No significant differences at the alpha = .05 level were noted.

Table 19. Site Two: Unit One Mean Writing Scores by Treatment Condition

Condition	Text	SD	MR	SD	AEB	SD	GTR	SD	GTA	SD	N
Treatment	9.22	1.22	3.06	2.82	1.70	1.44	15.56	6.31	13.98	4.14	18
Control	9.21	1.37	1.76	2.26	1.50	1.79	13.04	5.57	12.31	4.09	42
TOTAL	9.22	1.32	2.15	2.49	1.50	1.69	13.80	5.86	12.83	4.14	60

Table 20 summarizes significant differences for site two in unit one.

Table 20. Site Two: Unit One Significant Differences by Treatment Condition (α = .05)

Measure	Higher Score	t	p
Problem Questions	Control	2.075	.026
Overall Test Total	Control	2.075	.049

Effect sizes were also calculated for each of the assessment scores and each of the writing scores previously discussed. The effect size data is summarized in Tables 21 and 22.

Table 21. Site Two: Effect Size Calculations by Treatment Condition for Unit One Assessment

Question Type	Cohen's d	Effect Size
Completion	- .25	small
Problem Questions	- .65	medium
Multiple Choice	- .64	medium
Total Score	- .69	medium

Table 22. Site Two: Effect Size Calculations by Treatment Condition for Unit One Writing Scores

Writing Score	Cohen's d	Effect Size
Text	.01	negligible
Modal Representation	.54	medium
Average Embeddedness	.12	negligible
Grand Total (Raw)	.44	small
Grand Total (Average)	.41	small

Correlation and regression analysis were used to provide data analysis at the level of individual students for unit one. Correlations were calculated between writing scores and unit one test total scores. Text scores had a significant correlation at the p = .01 level with overall test scores. This was the only writing measure that was significantly correlated to overall test score. Table 23 summarizes the correlational data.

Table 23. Site Two: Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.333**	60
Modal Representations	-.035	60
Average Embeddedness	-.013	60
Grand Total (Raw)	.091	60
Grand Total (Average)	.082	60

** = significant at p < .01

Regression analysis compared the strength of prediction for overall unit one test score from the variables of baseline competency score and GTA from the writing scores. For unit one, the baseline competency score was a better predictor of overall unit test performance. Table 24 summarizes the regression data.

Table 24. Site Two: Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.628	.149	.490
Grand Total (Average)	.047	.511	.011

Unit Two Results

In unit two, the students at site two dealt with the concepts of bonding, molecular geometry, and bond energy. All students were assessed with an identical end of unit assessment that consisted of multiple choice questions (MC), three extended response questions (ER1, ER2, ER3), and a question that asked students to draw Lewis structure diagrams (LS). The first extended response question dealt with the difference between dipole forces and hydrogen bonding, the second dealt with differences between ionic and molecular compounds, and the third with bond energy and bond stability. Table 25 summarizes assessment data for this unit.

Table 25. Site Two: Unit Two Mean Assessment Scores by Treatment Condition

Condition	MC	SD	ER1	SD	ER2	SD	ER3	SD
Treatment	73.40	15.38	62.11	29.74	57.89	33.92	66.32	23.14
Control	73.71	11.75	64.55	21.51	62.73	27.14	78.18	23.94
TOTAL	73.62	12.82	63.81	24.06	61.27	29.15	74.60	24.15

Condition	LS	SD	TOTAL	SD	N
Treatment	57.02	16.96	68.78	16.56	19
Control	58.71	21.11	70.81	11.57	44
TOTAL	58.20	19.83	70.20	13.17	63

The writing assignment for unit two at site two involved students writing a letter to junior high students to describe what VSEPR theory is and how the geometry of a particular molecule is determined. Table 26 summarizes these results.

Table 26. Site Two: Unit Two Mean Writing Scores by Treatment Condition

Treatment	Text	SD	MR	SD	AEB	SD	GTR	SD	GTA	SD	N
Treatment	9.53	1.26	5.26*	2.28	3.68*	.94	22.00	7.33	18.36*	2.52	19
Control	9.70	1.07	3.39	2.77	2.64	1.83	19.41	9.85	15.82	4.55	44
TOTAL	9.65	1.12	3.95	2.76	2.95	1.68	20.19	9.18	16.59	4.19	63

* $p < .05$

Significant differences were noted for three writing scores from unit two. The treatment group (M= 5.26, SD = 2.28) outperformed the control group (M = 3.39, SD = 2.77) on modal representation score (t = 2.593, p = .012), the treatment group (M = 3.68, SD = .94) outperformed the control group (M = 2.64, SD = 1.83) on average embeddedness score (t = 2.978, p = .004), and treatment (M = 18.36, SD = 2.52) outperformed control (M = 15.82, SD = 4.55) on grand total–average score (t = 2.837, p = .006). Significant differences are summarized in table 27.

Table 27. Site Two: Unit Two Significant Differences by Treatment Condition ($\alpha = .05$)

Measure	Higher Score	t	p
Modal Representation	Treatment	2.593	.012
Average Embeddedness	Treatment	2.978	.004
Grand Total (Average)	Treatment	2.837	.006

Effect size calculations were also determined for both assessment and writing score differences. Negative scores indicate better performance by the control group.

Table 28. Site Two: Effect Size Calculations by Treatment for Unit Two Assessment

Question Type	Cohen's d	Effect Size
Multiple Choice	-.02	negligible
Extended Response 1	-.10	negligible
Extended Response 2	-.17	negligible
Extended Response 3	-.50	medium
Lewis Structure Diagram	-.08	negligible
Total Test Score	-.15	negligible

Table 29. Site Two: Effect Size Calculations by Treatment for Unit Two Writing Scores

Writing Score	Cohen's d	Effect Size
Text	-.15	negligible
Modal Representation	.71	medium
Average Embeddedness	.64	medium
Grand Total (Raw)	.28	small
Grand Total (Average)	.63	medium

Analysis at the level of the individual student was also carried out to determine if correlations existed between student writing scores and overall test performance. The correlations are summarized in table 30. None of the correlations were significant.

Table 30. Site Two: Correlations between Writing Scores and Unit Two Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.099	63
Modal Representations	.034	63
Average Embeddedness	.177	63
Grand Total (Raw)	.059	63
Grand Total (Average)	.131	63

Linear regression analysis was also utilized to determine if baseline science competency scores or grand total (average) scores were better predictors of end of unit assessment scores. This data is summarized in table 31.

Table 31. Site Two: Regression Analysis Predicting Assessment Total for Unit Two

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.088	.121	.092
Grand Total (Average)	.411	.400	.131

Site Three Results

Equivalency of Groups

All students at site three completed the baseline science competency test and baseline writing activity described in the methods section just prior to initiation of study procedures. Group scores by treatment were compared. No significant differences at alpha = .05 were noted. Table 32 summarizes baseline test results by treatment.

Table 32. Site Three: Mean Baseline and Baseline Emebeddedness Scores by Treatment Condition

Condition	Baseline	SD	Baseline Embeddedness	SD	N
Treatment	47.92	13.55	4.24	3.00	48
Control	47.19	9.51	3.73	3.13	22
TOTAL	47.69	12.36	4.08	3.03	70

Unit One Results

Unit one for site three involved concepts dealing with classification of matter. Scores for the end of unit exam were analyzed to test for differences by treatment. End of unit exam scores analyzed included a total score, as well as component scores from a multiple choice (MC) section and from five separate extended response conceptual questions (ER1, ER2, ER3, ER4, ER5). Extended response question 1 dealt with classifying and separating samples of matter, question 2 involved evidence of chemical reactions and characteristics of chemicals, question 3 dealt with chemical and physical changes, question 4 dealt with determining correct names for matter, and question 5 dealt with the conservation of energy. Table 33 summarizes mean assessment data from unit one.

Table 33. Site Three: Unit One Mean Assessment Scores by Treatment Condition

Condition	MC	SD	ER1	SD	ER2	SD	ER3	SD	ER4	SD	N
Treatment	85.25	11.93	68.54	19.89	66.88	13.71	88.54	13.88	46.88	22.85	48
Control	83.82	12.50	61.36	16.42	65.91	14.36	90.58	14.85	48.86	24.97	22
TOTAL	84.80	12.04	66.29	19.05	66.57	13.82	89.18	14.12	47.50	23.37	70

Condition	ER5	SD	TOT	SD	N
Treatment	61.98	22.47	77.02	10.44	48
Control	59.09	27.33	75.65	9.09	22
TOTAL	61.07	23.94	76.59	9.99	70

Writing scores were analyzed from products that involved students writing a letter to local junior high students about how to classify matter and separate matter. A summary of writing scores is in table 34.

Table 34. Site Three: Unit One Mean Writing Scores by Treatment Condition

Condition	Text	SD	MR	SD	AEB	SD	GTR	SD	GTA	SD	N
Treatment	9.31	1.56	4.27*	2.45	2.78*	1.30	19.19*	6.99	16.37	3.89	48
Control	9.63	2.46	3.00	1.90	2.00	1.52	15.64	6.11	14.63	4.68	22
TOTAL	9.41	1.88	3.87	2.35	2.54	1.41	18.07	6.88	15.82	4.20	70

* $p < .05$

Means for all assessments and writing tasks were compared to determine if significant differences existed. The treatment group (M = 4.27, SE = 2.45) outperformed the control group (M = 3.00, SE = 1.90) on modal representation score (t =2.153, p = .035). Treatment (M = 2.78, SE = 1.30) also outperformed control (M = 2.00, SE = 1.52) on average embeddedness score (t = 2.218, p = .030). Finally, treatment (M = 19.19, SE = 6.99) outperformed control (M = 15.64, SE = 6.11) on grand total (raw) score (t = 2.050, p = .044). Significant differences detected from this data are summarized in table 35.

Table 35. Site Three: Unit One Significant Differences by Treatment Condition (α = .05)

Measure	Higher Score	t	p
Modal Representations	Treatment	2.153	.035
Average Embeddedness	Treatment	2.218	.030
Grand Total (Raw)	Treatment	2.050	.044

Effect size data was also collected for analysis at the group level for unit one. As indicated in the tables, all effect sizes for assessment were either small or negligible and two scores (extended response question 3 and 4) had better performance in the control group. In the writing scores, the mean text scores were higher for the control group, but for all other measures, medium or small effect sizes favored the treatment group. Table 36 and table 37 summarize this data.

Data analysis at the level of the individual student was also undertaken to examine relationships between an individual student's score on the assessment task and the writing task scores. These correlations are presented in table 38. The three component writing scores (text, modal representations, and average emebeddedness score) all yielded significant correlations at alpha = .05. Both total scores (grand total raw and grand total average) yielded significant correlations at alpha = .01.

Table 36. Site Three: Effect Size Calculations by Treatment Condition for Unit One Assessment

Question Type	Cohen's d	Effect Size
Multiple Choice	.12	negligible
Short Answer Question 1	.39	small
Short Answer Question 2	.07	negligible
Short Answer Question 3	-.15	negligible
Short Answer Question 4	-.09	negligible
Short Answer Question 5	.12	negligible
Total Test Score	.14	negligible

Table 37. Site Three: Effect Size Calculations by Treatment Condition for Unit One Writing Scores

Writing Score	Cohen's d	Effect Size
Text	-.17	negligible
Modal Representation	.56	medium
Average Embeddedness	.58	medium
Grand Total (Raw)	.54	medium
Grand Total (Average)	.43	small

Table 38. Site Three: Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.262*	70
Modal Representations	.267*	70
Average Embeddedness	.393**	70
Grand Total (Raw)	.388**	70
Grand Total (Average)	.398**	70

* = significant at $p < .05$

** = significant at $p < .01$

Linear regression analysis was also performed to test whether the individual student's background science knowledge, as measured by the baseline science competency test, or student's degree of emebeddedness, as measured by their grand total-average score was a better predictor of the student total score on the end of unit assessment. Table 39 summarizes this data and indicates that when standardized regression coefficients (beta values) are compared, the baseline competency score was a stronger predictor of the total score on the unit 1 assessment.

Table 39. Site Three: Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.308	.084	.381
Grand Total (Average)	.805	.246	.338

Unit Two Results

Unit two results were obtained for the same analysis as were undertaken for unit one results. Unit two immediately followed unit one and was focused on the core concept of atomic structure. The end of unit assessment for unit 2 included multiple choice questions (MC), four extended response questions (ER1, ER2, ER3, ER4), an electron configuration question (EC), and an atomic structure question (AS) where students indicated the number of subatomic particles in atoms of different elements. Extended response question 1 asked students to discuss important discoveries and relationships that shaped our view of the structure of the atom, question 2 asked students to discuss differences between ions, common isotopes and uncommon isotopes, question 3 asked students to describe how atomic mass is determined and utilized, and question 4 asked students to describe the current view of electron configuration. Table 40 lists summary data for all assessment scores from unit two.

Table 40. Site Three: Unit Two Mean Assessment Scores by Treatment Condition

Condition	MC	SD	ER1	SD	ER2	SD	ER3	SD	ER4	SD
Treatment	73.07*	15.10	62.50	16.31	80.14	19.03	49.58	21.83	59.38	23.98
Control	63.47	17.46	63.18	18.61	72.42	28.31	50.91	23.69	57.72	25.20
TOTAL	70.05	16.38	62.71	16.93	77.71	22.44	50.00	22.26	58.86	24.20

Condition	EC	SD	AS	SD	TOTAL	SD	N
Treatment	84.64	21.14	86.90	16.51	75.50*	12.56	48
Control	76.14	34.05	76.30	28.79	68.02	17.38	22
TOTAL	81.96	25.95	83.57	21.51	73.15	14.55	70

* $p < .05$

The writing task for unit 2 asked students to create a magazine article describing the history of our understanding of atomic structure as well as our current understanding of the structure of the atom. Table 41 lists summary data for writing scores from unit two.

Analysis at the group level detected four significant differences between the treatment and control groups. In the assessment scores, the treatment group (M = 73.07, SD = 15.10) outperformed the control group (M = 63.47, SD = 17.46) on the multiple choice section (t = 2.351, p = .022). Treatment (M = 75.50, SD = 12.56) also outperformed control (M = 68.02, SD = 17.38) on the total assessment score (t = 2.042, p = .045). Significant differences were also noted for two writing scores in unit two. Treatment (M = 10.44, SD = 1.01) outperformed control (M = 9.91, SD = 0.92) on the text score (t = 2.089, p = .040). Treatment (M = 17.97, SD = 3.04) also outperformed control (M = 16.08, SD = 3.09) on the grand total (average) score (t = 2.404, p = .019). Unit two significant differences at the alpha = .05 level are summarized in table 42.

Table 41. Site Three: Unit Two Mean Writing Scores by Treatment Condition

Condition	Text	SD	MR	SD	AEB	SD	GTR	SD	GTA	SD	N
Treatment	10.44*	1.01	5.40	2.35	2.14	.83	20.90	6.03	17.97*	3.04	48
Control	9.91	.92	4.22	2.18	1.94	.95	17.95	5.26	16.08	3.09	22
TOTAL	10.27	1.01	5.03	2.35	2.08	.87	19.97	5.92	17.38	3.16	70

* p < .05

Table 42. Site Three: Unit Two Significant Differences by Treatment Condition (α= .05)

Measure	Higher Score	t	p
Multiple Choice	Treatment	2.351	.022
Total Assessment Score	Treatment	2.042	.045
Text	Treatment	2.089	.040
Grand Total (Average)	Treatment	2.404	.019

Effect size data also indicated higher performance from the treatment group in several areas. With the assessment questions, medium effect size differences were shown for the multiple choice questions, the question on atomic structure, and the overall score.

In all of these cases, treatment outperformed control. For writing scores, medium effect sizes were noted for all score subgroups except average embeddedness which exhibited a small effect size. Again, in all cases treatment outperformed control. Table 43 and table 44 list effect size data for unit two.

Table 43. Site Three: Effect Size Calculations by Treatment Condition for Unit Two Assessment

Question Type	Cohen's d	Effect Size
Multiple Choice	.61	medium
Extended Response Question 1	-.04	negligible
Extended Response Question 2	.35	small
Extended Response Question 3	-.06	negligible
Extended Response Question 4	.07	negligible
Electron Configuration	.33	small
Atomic Structure	.51	medium
Total Score	.53	medium

Table 44. Site Three: Effect Size Calculations by Treatment Condition for Unit Two Writing Scores

Writing Score	Cohen's d	Effect Size
Text	.55	medium
Modal Representation	.52	medium
Average Embeddedness	.23	small
Grand Total (Raw)	.52	medium
Grand Total (Average)	.63	medium

Analysis at the level of the individual student again included correlational analysis between writing scores and the overall assessment score. All correlations were significant at the alpha = .01 level. Table 45 summarizes the correlation data.

Table 45. Site Three: Correlations between Writing Scores and Unit Two Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.332**	70
Modal Representations	.542**	70
Average Embeddedness	.507**	70
Grand Total (Raw)	.569**	70
Grand Total (Average)	.647**	70

** = significant at $p < .01$

Linear regression analysis was used to test whether baseline science competency scores or embeddedness scores were better predictors of student performance on the end of unit assessment. Table 46 summarizes these results that indicate the grand total (average) embeddedness score was a more powerful predictor of student overall assessment score for this unit.

Table 46. Site Three: Regression Analysis Predicting Assessment Total for Unit Two

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.329	.103	.280
Grand Total (Average)	2.795	.403	.607

Individual Analysis Grouped by Level of Achievement

A final analysis undertaken for this site involved analyzing data at the level of the individual students when the overall group was divided by achievement level. The achievement level division was accomplished using the prior year's science scores on the Iowa Test of Educational Development (ITED). The ITED scores were available to the researcher for sites three and four (to be discussed next) only. Student scores on this standardized exam were utilized as a general measure of student science ability. Three categories were created based on this score. Students with an ITED science score of 71 or lower were grouped in the low achievement group, students with ITED scores between 72 and 87 were grouped in the middle achievement group, and students with ITED scores of 88 or greater were grouped in the high achievement group. The cut-off scores utilized were employed to yield three relatively similar sized groups. For each group, the correlation and liner regression analyses described earlier were re-run on data for each achievement level grouping separately. Table 47 lists the correlation data for the low achieving group for unit one and Table 48 lists correlation data for the low achieving group for unit two. Significant correlations were noted for all writing scores and total assessment score except text in unit two. No significant correlations were noted in unit one.

Table 47. Site Three (Low): Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	-.053	23
Modal Representations	.354	23
Average Embeddedness	.228	23
Grand Total (Raw)	.341	23
Grand Total (Average)	.216	23

Table 48. Site Three (Low): Correlations between Writing Scores and Unit Two Assessment Total Score

Writing Score	Pearson Correlation	N
Text	-.209	23
Modal Representations	.708**	23
Average Embeddedness	.613**	23
Grand Total (Raw)	.671**	23
Grand Total (Average)	.636**	23

** = significant at $p < .01$

Table 49 summarizes linear regression analysis for unit one for the low achieving group, while Table 50 summarizes this information for unit two. The results from unit one indicate the baseline competency test was a better predictor of unit test performance than degree of embeddedness, while unit two results indicate the opposite result.

Table 49. Site Three (Low): Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.273	.169	.336
Grand Total (Average)	.306	.393	.162

Table 50. Site Three (Low): Regression Analysis Predicting Assessment Total for Unit Two

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.097	.206	.082
Grand Total (Average)	3.435	.957	.624

The middle achieving group showed significant correlations for all categories except text and modal representation scores in unit one. This data is summarized in table 51 (unit one) and table 52 (unit 2).

Table 51. Site Three (Middle): Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	-.009	27
Modal Representations	.329	27
Average Embeddedness	.580**	27
Grand Total (Raw)	.455*	27
Grand Total (Average)	.471*	27

** = significant at $p < .01$

Table 52. Site Three (Middle): Correlations between Writing Scores and Unit Two Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.674**	27
Modal Representations	.528**	27
Average Embeddedness	.445*	27
Grand Total (Raw)	.564**	27
Grand Total (Average)	.636**	27

* = significant at $p < .05$

** = significant at $p < .01$

Regression analysis, shown in tables 53 and 54 for the middle group, indicated that in both units, the embeddedness score was a better predictor of student overall performance on the end of unit assessment.

Table 53. Site Three (Middle): Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	-.002	.130	-.003
Grand Total (Average)	1.233	.474	.471

Table 54. Site Three (Middle): Regression Analysis Predicting Assessment Total for Unit Two

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.278	.177	.238
Grand Total (Average)	2.336	.531	.664

The only significant correlations noted for the high achieving group were in unit two on the average embeddedness score and the grand total (average score).

Table 55. Site Three (High): Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.227	20
Modal Representations	.250	20
Average Embeddedness	.333	20
Grand Total (Raw)	.394	20
Grand Total (Average)	.385	20

Table 56. Site Three (High): Correlations between Writing Scores and Unit Two Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.378	20
Modal Representations	.328	20
Average Embeddedness	.502*	20
Grand Total (Raw)	.352	20
Grand Total (Average)	.563**	20

* = significant at p < .05

** = significant at p < .01

Regression analysis for the high achieving group indicated that for both units, the grand total (average) score indicating degree of embeddedness was a better predictor of overall assessment performance. Tables 57 and 58 show this data from units one and two.

Table 57. Site Three (High): Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.157	.155	.221
Grand Total (Average)	.591	.338	.380

Table 58. Site Three (High): Regression Analysis Predicting Assessment Total for Unit Two

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.125	.279	.089
Grand Total (Average)	2.573	.907	.571

Site Four Results

Group four consisted of five total classes taught by one teacher. Three of the classes were randomly selected to form the treatment group and two classes were combined to form the control group. Due to practical constraints and scheduling conflicts, group four was only able to complete one unit for the study. The unit dealt with the topic of electron configuration and periodic trends. The writing task for the unit involved students writing to future chemistry students and explaining to them the fundamental concepts associated with these topics.

Equivalency of Groups

All participating students were assessed prior to initiation of study procedures with the same baseline competency test students from groups one, two, and three were given. In addition, group four students were asked to complete a baseline writing task to assess degree of emebeddedness in writing. Table 59 summarizes the data from these tasks by treatment condition.

Table 59. Site Four: Mean Baseline and Baseline Emebeddedness Scores by Treatment Condition

Condition	Baseline	SD	Baseline Embeddedness	SD	N
Treatment	47.71	16.38	4.61	2.20	55
Control	43.21	19.48	3.03	2.47	40
TOTAL	45.81	17.79	3.94	2.43	95

There was no significant difference by treatment in terms of baseline competency scores, however, treatment (M = 4.61, SD = 2.20) outperformed control (M = 3.03, SD = 2.47) on the baseline embeddedness assessment (t = -3.233, p = .002). Due to this significant difference, baseline embeddedness scores were used as a covariate when assessing writing scores for this group.

Unit One Results

The end of unit assessment for group four included a multiple choice section (MC) and two extended response conceptual questions (ER1, ER2). The first extended response question asked students to describe why the periodic table is organized in the manner it is, while the second question asked students to choose three separate elements from three groups in the periodic tables and explain their characteristics and why they have these. Total test assessment score was also collected. Table 60 summarizes the results of the end of unit assessment.

Table 60. Site Four: Unit One Mean Assessment Scores by Treatment Condition

Condition	MC	SD	ER1	SD	
Treatment	83.83*	9.79	44.55	15.49	
Control	78.33	9.85	41.75	14.30	
TOTAL	81.51	10.13	43.37	14.99	
Treatment	ER2	SD	TOT	SD	N
Treatment	74.18*	20.52	76.63*	9.95	55
Control	63.50	21.19	70.77	8.93	40
TOTAL	69.68	21.36	74.16	9.92	95

* $p < .05$

Writing scores for the unit one writing task were also collected and this data is summarized in table 61.

The scores from the assessment data and the writing scores were analyzed to test for significant differences. In the assessment scores, the treatment group (M = 83.83, SD = 9.79) outperformed the control group (M = 78.33, SD = 9.85) on multiple choice

Table 61. Site Four: Unit One Mean Writing Scores by Treatment Condition

Condition	Text	SD	MR	SD	AEB	SD	GTR	SD	GTA	SD	N
Treatment	9.98*	1.35	6.02	3.22	2.67	1.23	23.98*	11.01	18.67*	4.60	55
Control	9.25	1.48	4.70	2.95	2.21	1.34	19.35	9.19	16.16	4.79	40
TOTAL	9.67	1.45	5.46	3.17	2.48	1.29	22.03	10.23	17.89	4.65	95

* p < .05

questions (t = 2.693, p = .008), treatment (M = 74.18, SD = 20.52) outperformed control (M = 63.50, SD = 21.19) on extended response question 2 (t = 2.471, p = .015), and treatment (M = 76.63, SD = 9.95) outperformed control (M = 70.77, SD = 8.93) on overall assessment score (t = 2.956, p = .004). As previously mentioned, due to the significant difference in baseline embeddedness, writing score differences by treatment were analyzed using the baseline embeddedness score as a covariate. The treatment group (M = 9.98, SD = 1.35) outperformed the control group (M = 9.25, SD = 1.48) on text score (F = 5.663, MS = 11.364, p = .019), treatment (M = 23.98, SD = 11.01) outperformed control (M = 19.35, SD = 9.19) on grand total (raw) score (F = 4.528, MS = 483.773, p = .036), and treatment (M = 18.67, SD = 4.60) outperformed control (M = 16.16, SD = 4.79) on grand total (average) score (F = 6.29, MS = 22.11, p = .014). Significant differences on end of unit assessment data are summarized in table 62 and on the writing scores are summarized in table 63.

Table 62. Site Four: Unit One Assessment Significant Differences by Treatment Condition (α = .05)

Measure	Higher Score	t	p
Multiple Choice	Treatment	2.693	.008
Extended Response 2	Treatment	2.471	.015
Total Assessment Score	Treatment	2.956	.004

Table 63. Site Four: Unit One Writing Significant Differences by Treatment Condition ($\alpha = .05$)

Measure	Higher Score	F	p
Text	Treatment	5.663	.014
Grand Total (Raw)	Treatment	4.528	.033
Grand Total (Average)	Treatment	6.290	.012

Effect size differences were also calculated between the treatment and control scores. These differences are summarized in table 64 for assessment scores and table 65 for writing scores. The same categorization scheme utilized throughout this results chapter is used here to differentiate the category of effect size.

Table 64. Site Four: Effect Size Calculations by Treatment Condition for Unit One Assessment

Question Type	Cohen's d	Effect Size
Multiple Choice	.51	medium
Extended Response 1	.19	small
Extended Response 2	.52	medium
Total Test Score	.62	medium

Analysis at the level of the individual student involved correlation and regression calculations. As with the other groups, correlations between each writing score and the overall test total score were calculated. This data is summarized in table 66. All correlations were significant at the alpha = .01 level.

Table 65. Site Four: Effect Size Calculations by Treatment Condition for Unit One Writing Scores

Writing Score	Cohen's d	Effect Size
Text	.52	medium
Modal Representation	.43	small
Average Embeddedness	.36	small
Grand Total (Raw)	.45	small
Grand Total (Average)	.54	medium

Table 66. Site Four: Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.348**	95
Modal Representations	.335**	95
Average Embeddedness	.394**	95
Grand Total (Raw)	.411**	95
Grand Total (Average)	.431**	95

** = significant at $p < .01$

Regression analysis was also utilized to determine if the baseline competency score or the grand total (average) score was a better predictor of overall test score for each student. The summary data in table 67 shows that when standardized (beta) coefficients are compared, the grand total (average) score was a better predictor than the baseline science competency score.

Table 67. Site Four: Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.065	.052	.117
Grand Total (Average)	.872	.193	.423

Individual Analysis Grouped by Level of Achievement

Data was also available for group four that allowed for coding of each student as a low, middle, or high science achiever. Just as with the group three students, the same cut-off scores from the Iowa Test of Educational Development were utilized to group the students. The individual analysis tests (correlation and regression) were then applied to each ability group separately. Table 68 lists the correlation data for the low science achievers and table 69 lists the regression data for this group. Three of the five correlations (modal representations, grand total-raw, and grand total–average) were significant correlations at the alpha = .01 level, while the average embeddedness score correlation to overall test assessment score was significant at the alpha = .05 level. Regression analysis indicated that the grand total (average) score was a better predictor of overall test performance.

Table 68. Site Four (Low): Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.343	33
Modal Representations	.444**	33
Average Embeddedness	.403*	33
Grand Total (Raw)	.500**	33
Grand Total (Average)	.526**	33

* = significant at p < .05

* = significant at p < .01

Table 69. Site Four (Low): Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	-.042	.083	-.081
Grand Total (Average)	1.435	.419	.54

Information for the middle achievement group is summarized in table 70 and table 71. For this group, no significant correlations were noted and the baseline competency score and the grand total (average) score were nearly equal predictors of the overall score.

Table 70. Site Four (Middle): Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.209	33
Modal Representations	.052	33
Average Embeddedness	.010	33
Grand Total (Raw)	.153	33
Grand Total (Average)	.105	33

** = significant at $p < .01$

Table 71. Site Four (Middle): Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.045	.082	.100
Grand Total (Average)	.187	.279	.123

Finally, the high science achievement group data is summarized in table 72 and table 73. For this group, average embeddedness score correlated with overall test performance was a significant correlation at the p = .01 level, while the two grand total scores were significantly correlated with overall test performance at the p = .05 level. With this group, the grand total (average) score was a better predictor of overall test performance.

Table 72. Site Four (High): Correlations between Writing Scores and Unit One Assessment Total Score

Writing Score	Pearson Correlation	N
Text	.349	29
Modal Representations	.326	29
Average Embeddedness	.537**	29
Grand Total (Raw)	.367*	29
Grand Total (Average)	.429*	29

* = significant at p < .05

** = significant at p < .01

Table 73. Site Four (High): Regression Analysis Predicting Assessment Total for Unit One

Variable	Unstandardized	Std. Error	Standardized
Baseline Science Score	.003	.108	.004
Grand Total (Average)	.734	.304	.429

Summary of Data

Practical constraints, scheduling conflicts, and other incidental factors made it impossible to carry out a study where all four groups participating studied the same concepts for the same amount of time, used the same writing tasks and took identical assessments. The differences in groups made it necessary to analyze data from each group individually. While it is not possible to run an overall data analysis, comparing the results already discussed can provide for potential interpretive benefit. Table 74 summarizes all group data discussed in this chapter, noting where significant differences in assessment scores and writing scores emerged, as well as where significant correlations between writing scores and the overall assessment score were present. Table 75 summarizes the practical differences between groups where significant differences were detected. For assessment scores, practical differences in terms of points are indicated, along with overall point totals for the questions that produced significant differences. Practical writing score differences are all indicative of numerical differences. For example, a 2.0 difference in MR would indicate that one group utilized 2 more modal representations on average when compared to the other group. Finally, Table 76 is a summary of effect size calculations for all sites. This summary of data, along with the individual group information will provide the framework for the discussion in the following chapter.

Table 74. Summary of Results from all Sites

	Assessment Results		Writing Score Results		Correlation Results	
	Unit One	Unit Two	Unit One	Unit Two	Unit One	Unit Two
Group One	MC AS ER1 ER2 ER3 TOT	MC ***(T) ER1 *** (T) ER2 *** (T) EC TOT ***(T)	TEXT MR AEB GTR GTA	TEXT MR AEB GTR GTA	TEXT (N/A) MR (N/A) AEB (N/A) GTR (N/A) GTA (N/A)	TEXT MR AEB GTR GTA
	Unit One	Unit Two	Unit One	Unit Two	Unit One	Unit Two
Group Two	COMP PROB***(C) MC TOT *** (C)	MC ER1 ER2 ER3 LS TOT	TEXT MR AEB GTR GTA	TEXT MR ***(T) AEB ***(T) GTR GTA ***(T)	TEXT *** MR AEB GTR GTA	TEXT MR AEB GTR GTA
	Unit One	Unit Two	Unit One	Unit Two	Unit One	Unit Two
Group Three	MC ER1 ER2 ER3 ER4 ER5 TOT	MC ***(T) ER1 ER2 ER3 ER4 EC AS TOT ***(T)	TEXT MR ***(T) AEB ***(T) GTR *** (T) GTA	TEXT ***(T) MR AEB GTR GTA *** (T)	TEXT *** MR *** AEB &&& GTR &&& GTA &&&	TEXT &&& MR &&& AEB &&& GTR &&& GTA &&&
	Unit One		Unit One		Unit One	
Group Four	MC ***(T) ER1 ER2 ***(T) TOT ***(T)		TEXT ***(T) MR AEB GTR ***(T) GTA ***(T)		TEXT &&& MR &&& AEB &&& GTR &&& GTA &&&	

MC = Multiple Choice; ER1, 2, … = Extended Response Question 1, 2, …; AS = Atomic Structure;

EC = Electron Configuration; COMP = Completion Questions; PROB = Problem Set;

LS = Lewis Structure TOT = Total Score

TEXT = Text Score; MR = Modal Representation Score; AEB = Average Embeddedness Score;

GTR = Grand Total Raw Score; GTA = Grand Total Average Score

*** = Significant at $p < .05$ &&& = Significant at $p < .01$

Table 75. Summary of Practical Differences by Treatment Condition

Site	Unit	Category	Practical Difference	Total Possible	Higher Group
1	2	MC	1.80	30	Treatment
1	2	ER 1	0.50	5	Treatment
1	2	ER 2	0.80	5	Treatment
1	2	TOTAL	2.70	45	Treatment
2	1	PROB	2.25	15	Control
2	1	TOTAL	3.84	32	Control
2	2	MR	1.90	-	Treatment
2	2	AEB	1.00	-	Treatment
2	2	GTA	2.50	-	Treatment
3	1	MR	1.30	-	Treatment
3	1	AEB	0.78	-	Treatment
3	1	GTR	3.50	-	Treatment
3	2	MC	2.70	27	Treatment
3	2	TOTAL	6.86	98	Treatment
3	2	TEXT	0.50	-	Treatment
3	2	GTA	2.00	-	Treatment
4	1	MC	2.40	48	Treatment
4	1	ER 2	1.10	10	Treatment
4	1	TOTAL	4.08	68	Treatment
4	1	TEXT	0.70	-	Treatment
4	1	GTR	4.50	-	Treatment
4	1	GTA	2.50	-	Treatment

Table 76. Summary of Effect Size Data

	Assessment Results		**Writing Score Results**	
	Unit One	Unit Two	Unit One	Unit Two
Group One	MC .11 (N) AS -.04 (N) ER1 .63 (M) ER2 .21 (S) ER3 .58 (M) TOT .46 (S)	MC .91 (L) ER1 .77 (L) ER2 1.02 (L) EC .04 (N) TOT .95 (L)	TEXT .41 (S) MR .27 (S) AEB .39 (S) GTR .42 (S) GTA .38 (S)	TEXT -.11 (N) MR -.49 (S) AEB -.18 (N) GTR -.49 (S) GTA -.36 (S)
	Unit One	Unit Two	Unit One	Unit Two
Group Two	COMP -.25 (S) PROB -.65 (M) MC -.64 (M) TOT -.69 (M)	MC -.02 (N) ER1 -.10 (N) ER2 -.17 (N) ER3 -.50 (M) LS -.08 (N) TOT -.15 (N)	TEXT .01 (N) MR .54 (M) AEB .12 (N) GTR .44 (S) GTA .41 (S)	TEXT -.15 (N) MR .71 (M) AEB .64 (M) GTR . 28 (S) GTA .63 (M)
	Unit One	Unit Two	Unit One	Unit Two
Group Three	MC .12 (N) ER1 .39 (S) ER2 .07 (N) ER3 -.15 (S) ER4 -.09 (N) ER5 .12 (N) TOT .14 (N)	MC .61 (M) ER1 -.04 (N) ER2 .35 (S) ER3 -.06 (N) ER4 .07 (N) EC .33 (S) AS .51 (M) TOT .53 (M)	TEXT -.17 (N) MR .56 (M) AEB .58 (M) GTR .54 (M) GTA .43 (S)	TEXT .55 (M) MR .52 (M) AEB .23 (S) GTR .52 (M) GTA) .63 (M)
	Unit One		Unit One	
Group Four	MC .51 (M) ER1 .19 (S) ER2 .52 (M) TOT .62 (M)		TEXT .52 (M) MR .43 (S) AEB .36 (S) GTR .45 (S) GTA .54 (M)	

MC = Multiple Choice; ER1, 2, … = Extended Response Question 1, 2, …; AS = Atomic Structure;

EC = Electron Configuration; COMP = Completion Questions; PROB = Problem Set;

LS = Lewis Structure TOT = Total Score

TEXT = Text Score; MR = Modal Representation Score; AEB = Average Embeddedness Score;

GTR = Grand Total Raw Score; GTA = Grand Total Average Score

L = Large Effect M = Medium Effect S = Small Effect N = Negligible Effect

CHAPTER FIVE

DISCUSSION OF RESULTS

The study that has been presented in the preceding chapters was designed to answer the three specific research questions listed below:

1. Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to a greater degree of embeddedness in student writing?
2. Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to greater conceptual understanding as measured by end of unit assessments?
3. Can correlations between degree of embeddedness in writing and student performance be detected in a variety of classroom settings?

These questions were formulated to explore the potential benefit of utilizing multi-modal writing tasks in actual classroom settings to improve science conceptual understanding. The research literature reviewed in chapter two suggested an overlap between these multi-modal writing opportunities and characteristics of effective science learning environments. The research literature also referenced the idea of student attainment of a "multi-modal competency" as a characteristic separate from but potentially related to student attainment of science conceptual understanding. In the context of this study, development of a multi-modal competency referred to skill at recognizing and utilizing appropriate applications of different representations of science understanding as part of an integrated and coherent description of a concept. Ideally, successful attainment by students of multi-modal competency would also lead to a greater student conceptual understanding. Again, the literature reviewed hinted that this potential relationship could become a reality in well-designed tasks.

The possibility of a link between multi-modal competency and conceptual gains arising from a review of the literature along with results of pilot studies directed by the researcher led to the study presented here exploring the dual ideas of multi-modal competency and science conceptual understanding. The research presented here expands current efforts to investigate this link by specifically focusing on the idea of "embeddedness" in multi-modal writing tasks. In essence, the research discussed here sought to improve understanding of whether or not encouraging the embeddedness of different modes of representation in student written text had an impact on either student multi-modal competency (as measured through assessment of the multi-modal writing tasks) or student conceptual understanding (as measured with typical end of unit assessments). Student embeddedness was encouraged through instruction focused on recognition and implementation of strategies used to integrate alternative modes with text in common science sources. The investigations took place in different classrooms, with different teachers leading the classes, and in different geographical locations. In the previous chapter, all data and results collected were presented. In this chapter, the results presented in the preceding chapter will be discussed in three main sections. First, a summary of the answers to the research questions will be presented. Second, a discussion of the results from each specific site will be offered. Finally, three main overarching ideas emerging from the answers to the research questions and the discussion of each site will be discussed as both a potential explanation for the results and as suggestions of areas of interest for further study.

Answers to Research Questions

Research Question 1

Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to a greater degree of embeddedness in student writing?

The first research question was concerned with whether or not participation in embeddedness encouraging activities led to measurable differences in the writing products created by the students who had experienced these activities. Recall that for each site, the treatment students participated in a lesson specifically designed to highlight examples of effective integration of text and other modes of representing science, emphasize techniques that could be utilized to accomplish this embeddedness, and create a checklist that could be used to assess the embeddedness in a product. This instruction dealt with the idea of multi-modal competency by focusing on the issue of how to combine different modes to appropriately and effectively communicate about science concepts. It is important to note that the instruction associated with these lessons and the checklists created to assess embeddedness were not designed to focus on content or conceptual understanding, but rather to focus on techniques deemed to be important by students for creating effective multi-modal science communication. One measure of the effect of these lessons was comparison of the writing tasks created by the students who received the embeddedness instruction and those who did not. Statistically, this analysis was carried out through comparison of group means by treatment condition and calculation of effect size differences by treatment condition.

A total of seven separate writing tasks across all sites were compared for significant differences at the $p = .05$ level in the five categories discussed in chapter three: text (TEXT), modal representations (MR), average embeddedness score (AEB), grand total – raw (GTR), and grand total – average (GTA). On four of the seven tasks, significant differences indicating the treatment group outperformed the control group in at least one writing score category were noted. The seven writing score comparisons involved a total of thirty-five category comparisons (five per task) and eleven of these thirty-five indicated significantly better performance by the treatment group.

Three of the five categories (AEB, GTR, GTA) specifically took into account embedding techniques and higher scores in these categories were interpreted by the researcher as indicating a greater degree of embeddedness. These three categories accounted for twenty-one of the total category measures, and on seven of these twenty-one, treatment groups significantly outperformed control groups. Each of the four particular cases where at least one significant difference was noted included at least one of these three specific embeddedness measures being significantly better in the treatment group. There were no category comparisons in which the control group significantly outperformed the treatment group. It is important to note that only one site, site four, had significant differences in degree of embeddedness prior to the initiation of study procedures and in this group, the baseline embeddedness scores were used as a covariate in analysis of variance. Further, no other differences in instructional practices between treatment and control groups other than the embeddedness lessons were reported by any of the teachers at any site. Therefore, in four of seven cases, students participating in embeddedness encouraging activities also had significantly higher scores on at least one writing category related to embeddedness.

Effect size calculations are helpful when comparing situations that are not equivalent because they measure differences in standard deviation units. In this study, effect size differences are particularly useful due to the fact the particular units of study, writing tasks, and end of unit assessments were all unique to the particular site. Overall, the same thirty-five writing category comparisons described in the comparison of group means were also analyzed with effect size calculations (five categories in each of the seven writing assignment cases). Twenty-eight of the effect size differences favored the treatment group. Of the seven categories in which effect size differences indicated better performance by the control group, four were for the text scores that measured characteristics of the written text alone. Only two of the seven categories indicating better performance by the control group were found in the three categories the researcher has argued are specifically linked to embeddednes, and none of these were for the average embeddedness score. Therefore, of twenty-one total category comparisons linked to embeddedness, nineteen showed better performance by the treatment condition. In the nineteen differences favoring the treatment group, seven were classified as medium and ten were classified as small, with two classified as negligible.

In summary, while not all writing task scores showed significant performance differences by treatment condition, the significant differences that did exist were all cases in which treatment outperformed control. Further, the majority of effect size differences in writing categories were also cases of treatment outperforming control. Treatment scores for AEB were higher than control scores at all sites and for all units, even though this did not always result in a significant difference. Potential reasons for both the increased occurrence of higher performance by the treatment group and the lack of this higher performance in all cases will be discussed in detail later in this chapter.

Research Question 2

Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to greater conceptual understanding as measured by end of unit assessments?

The second research question refers to the second main idea of this study, the link between encouraging embeddedness in multi-modal writing tasks and student science conceptual understanding. Similar to the first research question, analysis relevant to this question involved comparison of means and effect size calculations. For this question, the scores for comparison came from end of unit assessments developed by the instructors participating in the study. The instructors were asked to prepare end of unit assessments that included multiple choice questions as well as extended response conceptual questions. The instructors and the researcher worked together to insure face validity of the questions as discussed in the methods chapter. Although there was some overlap of concepts covered at different sites, practical constraints made it impossible to administer identical tests to all students at all sites.

One important clarification is necessary before discussing the results related to this question. Careful reading of the wording of the question highlights a critical meaning intended by the researcher. The question is intended to explore the relationship between participation in the embeddedness encouraging activities and conceptual understanding, not the link between greater embeddedness in writing and science conceptual understanding. The idea driving the question is that even though a student's participation in the lessons highlighting examples of embeddedness and strategies to encourage embeddedness may not necessarily lead to greater embeddedness in writing, the participation itself can positively benefit student conceptual understanding. The cognitive consideration of how different modes are integrated to communicate a message in a source of information a student consults, or the consideration of how to utilize multiple modes in a piece of writing a student constructs could provide opportunity for the cycling and feedback initiation discussed in chapter two that leads to improved conceptual understanding, even if it does not necessarily lead to measurable differences in writing. The results from the first research question indicate that many, but not all groups of students did show greater embeddedness when they participated in the embeddedness lessons designed for this study. These differences, however, may not be a requirement for improved conceptual understanding.

As with the writing scores, seven separate end of unit assessments provided data relevant to this question. The comparisons of each end of unit assessment were done by section, as well as for overall test total. The total number of sections for each assessment varied. Four of the seven assessments had at least one section in which significant differences were noted between treatment and control groups. Three of these four cases involved greater performance by the treatment group. For the second unit test at site one, treatment outperformed control on multiple choice, extended response questions 1 and 2, and overall total score. For the second unit test at site three, treatment significantly outperformed control on multiple choice and overall total. For site four, only one unit test was assessed and treatment outperformed control on multiple choice, extended response question 2 and overall test total. The only instance in which control significantly outperformed treatment was for unit one at site two where control had significantly higher scores for the problem set and the overall test total.

Effect size differences were also calculated for each test section and overall total. Forty total effect size calculations were determined. Twenty-five of the effect size calculations showed higher performance by the treatment group. Of the fifteen differences that favored the control group, ten came from site two. All five of the remaining differences favoring the control group were categorized as negligible. In contrast, the twenty-five measures in which effect size differences favored the treatment group included seven categorized as negligible, six categorized as small, eight categorized as medium, and four categorized as large. In sum, twenty-three effect size differences were categorized as greater than negligible and eighteen of these were cases of treatment outperforming control.

Again, the end of unit assessment scores did not indicate an across the board benefit for the treatment group. There were some sections of assessments in which control groups outperformed treatment groups and some where differences between the groups were negligible. However, as with the writing scores, the majority of cases where significant differences occurred and where effect size differences were larger than negligible, were cases in which groups receiving embeddedness instruction outperformed groups who did not receive the instruction.

Research Question 3

Can correlations between degree of embeddedness in writing and student performance be detected in a variety of classroom settings?

In the previous research question, analysis at the level of the group was the focus. The third research question aimed at assessing the relationship of embedding multi-modal representations in text with conceptual understanding at the level of the individual student. This question stemmed from findings in the pilot study leading to this research. In the earlier study, significant positive correlations between scores for embeddedness and student performance on end of unit assessments were detected. While it is understood that a positive correlation is not indicative of a causal relationship, it is evidence that a positive relationship exists between the variables.

The researcher recognizes that other potential rationales for a positive correlation between degree of embeddedness and test performance exist and that other, underlying factors related to both embeddedness in writing and test performance may explain the relationship. For example, students who embed information in text better may have a learning style more in line with a style beneficial for learning science and therefore may do better on science assessments because of the learning style. It is also possible that some cognitive factor associated to writing is linked to both embeddedness and test performance. Therefore, it is accepted that positive correlations do not necessarily indicate a specific benefit for individual students arising from the use of multi-modal tasks. However, the correlations can provide one piece of evidence in a “logic-based” research frame to at least suggest a relationship between embeddedness and conceptual understanding. The continual detection of this relationship in a variety of settings with a variety of students of varying abilities (including writing skill) would be greater evidence of this relationship.

For all sites, bivariate correlations were calculated for the relationships between each of the writing score variables and the overall test score for every student. The only exception to this was for unit one at site one. Correlations were not calculated for this site because very low numbers of modes other than text were utilized by students. Therefore, little meaningful data for this type of calculation was obtained. The particular characteristics of this site will be further addressed in the next section. Excluding this case left six total cases for which correlations were determined. The six cases involved one in which bonding and formula naming was the topic of instruction, one in which classification of matter was the topic of instruction, one in which atomic structure was the topic of instruction, one in which molecular geometry was the topic of instruction, and two in which electron configuration was the topic of instruction.

A total of thirty correlations (five in each of the six cases) were collected. One set of correlations was calculated for site one (unit 2) and site four (unit 1). At sites two and three, correlations were calculated for two separate units. Of the thirty total correlations, only two correlations were negative. Both negative correlations were from site two, unit one and included the correlation between modal representation score and overall test score and average embeddedness score and total test score. All other correlations were positive. Further, three of the correlations were significant at the $p = .05$ level and thirteen were significant at the $p = .01$ level. Eighteen of the thirty total correlations were between measures determined to be specifically related to degree of embeddedness (AEB, GTR, GTA) and overall test performance. Of these scores, seventeen were positive correlations and nine were significant correlations at the $p = .01$ level.

In addition to the correlations calculated, linear regression techniques were applied to student scores. Specifically, the ability of the grand total (average) or GTA score was compared to that of the baseline science competency score for predicting overall test performance. The same six cases for which correlation analysis was undertaken were utilized for the linear regression calculations. For site one, unit two, the standardized regression coefficient (beta coefficients) were nearly equal, with the baseline score having a beta of .146 and the GTA having a beta of .141. For site two, unit one, the beta coefficient for the baseline score was greater (.490) than the beta coefficient for the GTA score (.011), but for unit two, the GTA (.131) was slightly higher than the baseline score (.092). The beta coefficient for baseline scores (.381) for site three, unit one was higher than that for GTA (.338), but this trend was also reversed in unit two where the beta coefficient for baseline score was .280 and GTA was .607. For site four, the baseline beta coefficient was -.081 and the GTA was a much higher .545.

In summary, the vast majority of correlation scores indicated a positive relationship between measures of embeddedness and overall test performance. Many of these positive correlations were deemed significant, supporting the assertion that degree of embeddedness in writing tasks and overall conceptual understanding are positively related. The regression data was more varied, but indicated that in most cases, the degree of embeddedness is nearly equal or greater in ability to predict evaluation performance compared to a students baseline science competency. In all cases in which two consecutive units were assessed, the GTA score was a better predictor of test performance on the second unit.

Discussion of Results by Sites

As previously mentioned, the characteristics of each of the testing sites were quite varied. In addition, the topics studied during the research procedures, as well as the time for each topic, the evaluations used for assessment, and the writing tasks themselves were different at each site. Analysis of the findings from each group separately is instructive both in terms of further explaining the general findings mentioned previously and in terms of identifying key overall ideas emerging from this research. This analysis by site will be presented in this section.

Site One

Site one was the smallest of the four sites in terms of number of participants. At this site, the treatment group consisted of 18 students while the control group had 19 students. Unit one at this site covered material dealing with atomic structure while unit two dealt with electron configuration and the periodic table. Students were assessed with the baseline competency test and there was no difference in baseline science competency by treatment group. The baseline embeddedness measure was not administered. No significant differences were detected between treatment and control on any end of unit assessment measures or writing measures for unit one. For unit two, the treatment group scored significantly higher on the multiple choice, extended response 1 and 2 questions, and overall test performance, but no significant differences in writing scores were detected. Analysis at the level of individual students was only applied to unit two results, due to factors to be discussed shortly. All correlations between writing scores and overall test performance were positive, but none were significant at the $p = .05$ level. The strongest positive correlation (.240) was between average emebeddedness score and overall test performance. Linear regression analysis indicated that baseline competency score and GTA score were nearly equal predictors of overall test performance for unit two.

Several factors related to implementation of the study procedures were likely contributing factors to the results for this group. Personal correspondence between the teacher at each site and the researcher was utilized to ascertain the degree of match between the intended procedures for the embeddedness lessons and the actual implementation of these procedures. It was the intent of the researcher to have all teachers provide a two-day lesson to treatment groups in which three main activities, took place. First, students were to highlight and discuss examples of multi-modal use and integration in a variety of sources of science information. Ideally, this would be followed by a discussion about strategies for integrating multiple modes of representation into text, culminating in the production of a class generated checklist to assess any source of information for degree of embeddedness. Finally, students were to be given an opportunity to practice embedding different modes with text using the strategies from their checklist. The following excerpt from correspondence with the teacher at site one indicates the characteristics of the actual embeddedness encouraging lessons at this site:

> We spent between 40-50 minutes total on the idea of embeddedness. I had the students look up examples in their textbook and report to the class what they had found. We then made the checklist, using the one provided as an example.

This description indicates that while students were given opportunities to find examples and discuss strategies, they were not given an opportunity to practice the utilization of the strategies for embeddedness. While this was not a major deviation from the intended lesson, this may have impacted the degree to which these strategies were initially grasped and effectively implemented.

A second, more critical characteristic of implementation at site one that potentially impacted results was a significant deviation from the overall study procedure. As discussed in chapter three, all students were to utilize at least one mode other than text in their writing products. This requirement was not in place for the unit one writing task at site one as indicated by the following correspondence from the participating teacher following receipt of the writing samples from unit one by the researcher:

> I did not require students from either treatment or control to utilize modes other than text. I indicated that they could do this if desired, but they were not required to. I will ask them to do this on the second writing.

This deviation from the intended procedure was the reason no correlation or linear regression analysis was done for unit one at site one. Analysis of the writing samples produced indicated that one obvious impact was that students did not spontaneously utilize modes other than text. Only seven total students used any mode other than text. Five of these students were in the treatment group and two in the control group. While this may provide some evidence that participation in the embeddedness lessons made students more aware of, and then more likely to utilize other modes, due to the very limited number of examples, it is not strong evidence. The overall result of very limited modal use, however, provides stronger evidence that students will not spontaneously utilize modes other than text unless specifically instructed to do so. Therefore, encouraging students to embed modes is likely a multi-step process, in which one essential step is the requirement of students, at least in initial writing, to use modes other than text, rather than expecting them to do this on their own.

Another potential factor impacting results from site one was the way feedback for the writing samples was obtained. As a requirement of the study, teachers were asked to have each student provide a rough draft of their multi-modal writing product for analysis by a reviewer other than the teacher. Suggested outside audiences included junior high students, parents, or peers at the high school not enrolled in chemistry. The feedback was to be done before the end of unit assessment so that students could utilize the feedback to improve their written products, revise, and then be assessed for conceptual understanding by the end of unit exam. This general process was utilized by students at site one, however, the reviewers consisted of members of the other treatment group. The control group critiqued the papers from the treatment group and vice versa. One implication of this procedure may have been that through review of the treatment group papers, control students may have picked up embeddedness techniques that were being utilized and then applied them to their own products. The results of the second unit, in which end of unit performance in several areas was significantly better for the treatment condition, but no significant differences in writing scores was detected may be a reflection of this exchange of drafts between groups. This may have in essence neutralized any differences in writing or embeddedness scores that could have existed from receiving differentiated instruction. As mentioned in the initial section of this chapter, the participation itself in embeddedness encouraging activities may be beneficial for conceptual understanding, even if the writing produced by the students does not reflect a difference in utilization of embeddedness techniques. Student performance on unit two indicating benefit for the treatment group in terms of conceptual understanding, but no benefit in terms of multi-modal competency provides evidence to support this contention.

Finally, site one results also indicate some cumulative benefit for the treatment group. The unit two results, measuring student conceptual understanding on an assessment after two consecutive units where the cycle of producing written products and then being assessed was repeated, did show some differentiation between treatment and control not present in the first unit. As mentioned before, the utilization of embedding strategies to produce a well integrated multi-modal text may involve several steps. These steps may take time to develop in students and the attainment of a "multi-modal competency" may not be possible for many students after only one exposure to the technique and one opportunity to employ it. If the multi-modal competency attainment takes time, then logically any conceptual benefit resulting from embedded multi-modal tasks would not immediately appear. The particular implementation characteristics at site one discussed previously may have delayed this development even more. This may also partially explain why there are positive correlations between embeddedness measures and overall test performance, but no significant correlations. Perhaps, further opportunity to utilize these techniques would have strengthened the relationship between these factors.

Site Two

Site two involved sixty total participants split into a control group of forty-two students and a treatment group of eighteen students. While the control group did have a higher mean baseline competency score, no significant difference existed between the groups. The same situation existed for the baseline embeddedness scores. At site two, students studied the concepts of ionic bonding and formula naming for their first unit. The control group scored significantly higher on the problem set questions and the total test score. No significant differences existed for writing scores. Text scores were positively correlated with overall test performance and this correlation was significant at the p = .01 level. No other correlations were significant, and the only two correlations in the entire study that were negative were found at this site for unit one. These were average embeddedness score and modal representations. Linear regression determined that for this site, the baseline science competency score was a better predictor of overall test performance (beta = .490) than GTA (beta = .011).

For the second unit, the situation at site two changed slightly. Results from this unit indicated that while the control group still outperformed the treatment group on raw assessment scores, there were no significant differences. Further, the writing task analysis indicated that the treatment condition performed significantly better than control on three writing measures: modal representations, average embeddedness score, and grand total (average). There were still no significant correlations between writing scores and test performance, but all correlations were positive, unlike the first unit.

The initial correlation results from unit one at site two indicate that a relationship between quality of text produced and student conceptual understanding is present. Much of the work with multi-modal writing tasks is predicated on the premise that non-traditional writing tasks, regardless of what modes are used, are beneficial. The results from unit one indicate that this was the case, as a better written product was significantly correlated in a positive way with overall test performance. This finding indicates that even if multi-modal competency (including use of embeddedness) is not accomplished immediately, multi-modal writing tasks can still be beneficial for students due to the writing aspect of the task.

In terms of the impact of the embeddedness encouraging lessons, degree of embeddedness in writing, and student assessment performance, implementation may have again been a major factor. As the following correspondence with the teacher at site two indicates, the level of implementation at this site was even less than at site one:

> For the test classes, I showed 2 examples. Text without any pics, diagrams, etc., just text and then text with diagrams, etc.. We discussed how much the embedded items (and I did not call them embedded items) helped to make the topic clearer, and how that could hold true in their communication of not only chem but any topic to others. Then each test class came up with their own checklist, together as a class with me leading the discussion and helping to clarify their brainstorming.

From this correspondence, it appears that the exploration of examples of integration of modes with text was limited and that students were not given the opportunity to practice embedding modes in text. The characteristics of the checklists created by the student groups may have also impacted the opportunity for benefit.

The checklists that were utilized by the treatment group at site two included a total of eleven items. Nine of the items were simply types of modes and student writers were asked to respond as to whether or not they utilized that type of mode. One item assessed whether color was used or not. Only one item referred to any sort of integration, asking whether the mode was "placed with the text it goes with". The checklists at this site emphasized use of modes to a much greater degree than embeddedness or integration of modes. This may have been a reflection of the instructor's view that embedding and integration of multiple modes is a natural consequence of using different modes that the teacher asserted already took place in the classroom:

> I use everything in my class, however, so it was not new to anyone, just not emphasized that I use several forms of embeddedness.

The results from site two indicate that students did not naturally develop skill in integrating the modes with text immediately. There did appear to be, however, improvement in terms of multi-modal competency for the second unit and a greater degree of this competency in the treatment group. Students in the treatment group performed significantly better on three of the five writing measures, including two of the three specific embeddedness measures. Further, correlations were positive for this second experience with overall test performance, indicating that repeated opportunities for all students may have led to an increase in beneficial multi-modal use. The correlation data for unit one indicated that both modal representation score and average embeddedness score were negatively correlated with overall test performance. In light of the description of implementation at this site, it seems likely that students perceived very little emphasis on embeddedness skills on the initial writing task and any use of modes other than text and any integration of those modes was incidental and not purposeful. It would seem unlikely that any sort of beneficial cognitive consideration of how the modes and the text work together that could then lead to a better conceptual understanding took place. However, repeated exposure, and perhaps more emphasis on modal use for the second unit may have overcome some of these deficiencies. Again, results at this site indicate the likelihood that the use of multi-modal writing tasks will improve student understanding is increased through effective implementation and multiple opportunities. It cannot be assumed that students will naturally develop or employ these strategies just because they are required to use multiple modes or because multiple modes are utilized in a classroom.

Site Three

Site three included students taught by the researcher during two consecutive units. The first unit involved the topic of classification of matter, while the second unit dealt with atomic structure. The treatment and control groups were equivalent in terms of baseline science competency and in terms of baseline embeddedness ability prior to implementation of study procedures. For unit one, no significant differences in performance on end of unit assessments were noted, but the treatment group had significantly better modal representation scores (MR), average embeddedness scores (AEB), and grand total-raw (GTR) scores. For unit two, the treatment condition showed significantly greater performance on multiple choice questions and overall test total score. Writing scores were significantly better for the treatment group for text and for grand total-average (GTA). The correlations between all writing scores and test performance for both units were all positive, with the text and modal representation correlations for unit one significant at the $p = .05$ level and all other correlations significant at the $p = .01$ level.

The fact that the researcher was also the teacher for these groups insured a high degree of fidelity for research procedures. In addition, the researcher / teacher had led classes in similar embeddedness encouraging activities as a part of the pilot study leading up to this study. The researcher also designed the measures that were used to assess embeddedness making it much more likely that characteristics measured on the embeddedness rubric were emphasized in the embeddedness encouraging lessons at this site. The researcher recognizes the differential experiences of the researcher / teacher at this site likely impacted the results and makes comparison of this site with others somewhat precarious. However, comparison of the results from unit one and unit two at this site only indicate a major factor impacting these results is the idea of a cumulative effect mentioned earlier.

The results from site three indicate some conceptual benefit associated with participating in embeddedness encouraging activities developed in the second of the two units. One possible explanation for this result relates to the topics studied. Different concepts or topics may lend themselves to the technique of using multi-modal writing as a method of improving conceptual understanding better than others. It is possible that in this case, some characteristic of the concepts involved with atomic structure make it more likely that embedding multi-modal representations in text is more helpful in developing an appropriate conceptual understanding than with the concepts associated with classification of matter.

However, another possible explanation of the findings at this site is that participating in more than one multi-modal writing task is more beneficial for students than a single exposure to these types of tasks. Improvement in student multi-modal competency may be gained through repeated use, and this may in turn lead to an increase in the conceptual benefit associated with the use of these types of tasks. A student's first experience with applying strategies to embed modes into text may not be as beneficial conceptually as the second or third. This contention is supported to some degree by the fact that at this site, differences appeared in the first unit favoring the treatment classes in terms of embeddedness and writing measures, but indication of conceptual benefit did not appear until the second unit. The correlation data may also back this assertion of cumulative benefit as well, as all correlations were stronger for the second unit than for the first, although other factors may have also impacted improvement. Regression analysis may also support this contention, as unlike in the first unit, for the second unit, GTA was a better predictor of overall test performance.

Another interesting question arises from the data at site three related to this idea of cumulative benefit. The data showed that while the treatment group performed significantly better on unit one MR, AEB, and GTR scores, this greater degree of performance was not repeated in unit two. In unit two, the treatment group was significantly higher on text scores and GTA score. While this could be interpreted as evidence against cumulative benefit by pointing out the benefit in terms of multi-modal competency for the treatment group seemed to decrease, an alternative explanation is that the repeated exposure to using multi-modal tasks was beneficial for students in the control group in terms their multi-modal competency. Even though the control students had not received specific training about embeddedness and did not have checklists to assess their writing with, they still produced two consecutive multi-modal writing tasks. It is conceivable that the control students improved their ability to utilize and integrate different modes simply through participation in the multi-modal writing tasks. This improvement from the control group did not completely alleviate differences between control and treatment on all writing and embeddedness measures, but it may have contributed to lessening the differences between the groups. The stronger correlations found in unit two for all could be indicative of improved performance by both the control and treatment groups in terms of multi-modal competency. In essence, the cumulative benefit extended to both groups.

The situation at site three in which the participating students were members of the researcher's classes allowed for analysis of data in a way not possible at the other sites due to access to prior year standardized testing data. This data included composite science scores for students on the Iowa Test of Educational Development (ITED). The ITED data provided a grouping variable in which students could be categorized by science ability level. The particular grouping procedures were discussed in the previous chapter and resulted in three categories of students based on science achievement: low, middle, and high. The analysis at the level of the individual students (correlation and regression) were then re-applied and analyzed by achievement level. The results may indicate differential situations existing in terms of relationships between writing scores and assessment performance for students of different ability levels.

For the low achieving students, two major themes emerged. First, in both unit one and two, text was negatively correlated with assessment performance. However, all other writing scores were positively correlated. Secondly, the trend of cumulative effect was quite obvious for the low achieving students. The positive correlations all increased dramatically for the second unit and all positive correlations for the second unit were significant at the $p < .01$ level. The regression analysis mirrored this trend, as baseline competency score was a better predictor of assessment performance in unit one, while GTA was a better predictor for unit two. This data may indicate that low achieving students need time to employ strategies for embeddedness and for experiencing the conceptual benefits associated with embedding multiple modes, but if given opportunities to produce more than one of these types of products the connection between degree of embeddedness and conceptual performance may increase. This connection may be noted even if the written text, which may prove consistently difficult for low achieving students, is not particularly strong.

For the middle-achieving students, a similar trend compared to the low achieving students was noted, but the impact of the multi-modal task may have been more immediate. Consistent with the low achieving group, the middle group had a negative text correlation for the first unit. All other correlations were positive, with GTR and GTA being significant correlations at the $p < .05$ level and AEB being significantly positive at the $p < .01$ level. Thus, a stronger positive relationship between writing scores outside of text and achievement was noticed in unit one for this group. For the second unit, all correlations were positive and significant at the $p < .01$ level except the AEB which was positive and significant at $p < .05$. Interestingly, although it exhibited a significant positive correlation, the only correlation that was weaker for the second unit was the AEB. This could possibly be a factor of the type of assignment for this unit or the characteristics of the conceptual matter. The linear regression data supported the contention that the middle achieving group realized benefit from multi-modal usage earlier than the lower achieving group, as the GTA was a better predictor of assessment performance on the first unit. This remained true for the second unit as well.

Finally, the data for the high achieving students reinforced the support for a cumulative effect in that all correlation measures except GTR improved from unit one to unit two. In addition, while regression analysis indicated that GTA was a better predictor of overall assessment performance on both units, the difference between the predictive ability of GTA and baseline competency score was more pronounced in unit two. The interesting aspect of this data set was that while all correlations were positive, the only significant correlations were between AEB and test performance and between GTA and test performance on unit two. This result could indicate that for higher level students, the use of multi-modal tasks is not as critical in improving conceptual understanding as for lower achieving students. The higher level students may be more likely to perform well on assessments regardless of the particular type of instruction utilized. However, it may be that a differentiating factor among high achievers may be their ability to embed modes effectively (as measured by the AEB and GTA score). Higher achieving students may be more cognitively able to improve conceptual understanding if they choose to utilize embeddedness techniques than if they do not and relative to other high achieving students, level of use of this characteristic may impact performance.

Site Four

Site four consisted of the largest sample size. This site was at the same school as site three, but consisted of classes taught by a different teacher. Site four only completed one cycle of writing and testing and it dealt with the concepts of the periodic table and periodic trends. Overall, site four produced the most dramatic differences between treatment and control groups. The treatment group outperformed the control group on three of four assessment measures and three of the five writing measures, after only one round of writing and testing. All correlations were positive and significant at the $p < .01$ level. Regression analysis indicated that the grand total (average) score was almost four times better at predicting the overall assessment score than was the baseline science competency score.

The major idea emerging from data analysis at site four was that striking differences between treatment and control could be attained after only one unit of study. The data from site three and from site four provided the strongest evidence of differential achievement between the treatment and control groups. It could be argued that the gains discussed from site three data were due more to the fact that the researcher was teaching the classes than from any particular benefit associated with the strategies and techniques explored. However, in site four with a different teacher, even stronger evidence exists of a benefit for both multi-modal competency and for science conceptual understanding emerging from participation in embeddedness encouraging activities. The proximity of this particular teacher to the researcher allowed for extended dialogue about implementation issues and to a greater fidelity to research procedures than with site one or two. In general, site four presents evidence that when a high degree of implementation was attained, a benefit from participation in embeddedness activities was present. In addition, site four dealt with a different topic than had been considered at site three, indicating the benefits from these tasks may not be limited to particular topics.

In addition, site four provided some additional evidence in terms of differential impact for differing ability levels. Again, low achieving students showed positive correlations between writing scores and assessment performance, with text showing the weakest relationship. The regression analysis indicated a much greater predictive ability of GTA compared to baseline competency for the low achieving students. As before, the average embeddedness score showed the strongest correlation for the high achieving students, possibly supporting the contention that ability to embed appropriately is critical in determining relative performance among high achieving students. The linear regression data supported this evidence. The middle achieving students at site four showed all positive correlations but had no significant correlations and had little difference in predictive ability between GTA and baseline competency score in the regression analysis. At site four, it appeared that for the middle level students, the multi-modal tasks had less relationship with test performance. Overall, site four data would again indicate different degree of impact from participation in embeddedness activities and from utilizing embeddedness techniques on student performance based on level of science ability.

Ideas Emerging from the Data

Consideration of the overall research results, as well as the results by specific sites leads to three main ideas emerging from this study. These ideas provide a framework to study and advance the understanding of the use of multi-modal writing tasks as a part of an effective science classroom, as well as characteristics for practitioners utilizing these tasks to consider. The key issues identified by the researcher are that the level of implementation is critical, that greater benefit may result as a part of a cumulative process as opposed to a one-time use, and that differential benefit may result for students of different ability levels. Each of these issues will be discussed separately and an attempt will be made to link each issue to the theoretical discussion in chapter two.

Implementation

In general, implementation as discussed here refers to the way student learning about embedding multiple modes of representing science in written text is encouraged. Three main factors are important in accomplishing this implementation. First, students must be given a chance to experience the different modes utilized in sources of science information. Secondly, students must be given opportunity to explore and evaluate different ways that the multiple modes, including text, can be integrated to create an accurate, thorough, and coherent description of a particular science concept. Finally, the students must be given an opportunity to create their own multi-modal products and evaluate the effectiveness of their own products in terms of communicating about a science concept. In the particular case of this study, the embeddedness encouraging lessons the treatment groups received, followed by the multi-modal writing tasks and the self evaluation of the writing tasks were intended to provide these integral factors.

The practical realities, however, of the different sites and the differential interpretations of the intended study procedures in terms of the embeddedness encouraging lessons resulted in a situation in which the impact of different levels of implementation could be assessed. The results discussed previously indicated that lower levels of implementation of the procedures associated with the embeddedness encouraging lessons resulted in little difference in multi-modal use, integration of multiple modes in text, or performance on end of unit assessment between treatment and control. In one case, there was actually greater performance by the control group. Conversely, when the degree of implementation was high, the treatment groups not only had greater integration of alternative modes in their writing but also had better performance on the end of unit assessments. It appeared that well implemented embeddedness encouraging lessons improved both student multi-modal competency and student conceptual understanding of the topics written about. Exploration of specific factors related to implementation provides some explanation for the results encountered in this study, as well as links to the theoretical ideas presented in chapter two.

One characteristic of implementation is the general idea of emphasis. Recall from chapter two the discussion on the interactive nature of an effective science classroom. In that discussion, it was pointed out that with study of any particular concept, a vast array of actions and characteristics are interacting to create the specific learning environment. Further, the ideas consistent with an interactive constructivist view of learning would posit that a multitude of personal cognitive actions for each individual student and social learning factors based on the complex interactions of a cooperative learning environment are present in every classroom. The obvious conclusion is that with so many complex and sometimes competing factors present, if a particular idea or concept is not emphasized, it is difficult to assume students will spontaneously realize its importance and either understand or apply it appropriately. Actions associated with a lower level of implementation in terms of embeddedness encouraging activities would result in less emphasis on this as a critical issue or concept and would likely make it difficult for students to recognize the importance of utilizing the skills related to it. Less utilization of embeddedness techniques would lead to little or no benefit from the techniques in terms of conceptual development. The classes in which multiple modes were not required of student written products or where the embeddedness encouraging lessons were minimal demonstrated this relationship between less emphasis, less embeddedness, and less conceptual benefit.

Another aspect of implementation with these multi-modal tasks revolves around the characteristics of the assignments given to students. Even if effective instruction and sound learning opportunities are provided for students to develop an awareness of how embedding and integrating text with multiple modes can result in enhanced communication of scientific ideas, if the multi-modal writing tasks are not designed effectively, benefit for students may not result.

One possible example of this from this particular study deals with the feedback. In one particular case, feedback came from members of the other treatment condition. Students were aware that the "audience" for the multi-modal product was students taking the same class they were. This conception of audience may have made it more likely that the student authors engaged in knowledge telling activity, described by Bereiter and Scardamalia (1987) as simple recall of information already learned. If the student authors perceived no need to effectively communicate their understanding to their audience because the student authors assumed the audience already understood the material, they may have deemed it sufficient to re-tell what had already been learned. Bereiter and Scardamalia would assert that this would not lead to the more cognitively demanding knowledge transforming writing or the knowledge constituting writing proposed by Galbraith (1999). In addition, the translation ideas proposed in chapter two to result from multi-modal consideration would not be necessary. This lesser cognitive demand would be less likely to lead to improved conceptual understanding for the author. Students may also have been less likely to feel the need to create more effective communication of the science topic through utilization of the techniques associated with embedding multiple modes, even though these techniques may have been recognized in the evaluation of sources of information as a part of the embeddedness encouraging techniques. This may explain why the treatment classes who received feedback from their peers in a different section of the same course showed better assessment scores (they had realized some benefit from the consideration of multiple modes in the embeddedness lessons) but did not have a higher degree of embeddedness in their writing.

Finally, just as research negates the assertion of a strong text hypothesis that all writing tasks are beneficial for student conceptual development or improvement due simply to the fact that the students are engaged in writing, the results of this study would indicate that simply requiring students to produce a multi-modal writing product does not automatically lead to benefit. The one site in which control scores were better on end of unit assessments was the site in which the teacher expressly stated that embedding multiple modes was a normal part of the class and students had been exposed to it on a consistent basis. The assumption that students would spontaneously acquire the ability to integrate writing with text based on the use of multiple modes in the classroom can be deemed a faulty assumption based on the results. Further, when students were not specifically required to utilize other modes besides text, they typically did not use many modes other than text and they did not spontaneously utilize embedding techniques. Evidence here suggests in order for students to create effectively embedded multi-modal writing tasks and ultimately improve their understanding of the associated science content, a high degree of focused, consistent, and explicit implementation is necessary.

Cumulative Effect

Evidence from this study indicates that the benefit associated with multi-modal writing tasks may be cumulative in the sense that the construction of multiple multi-modal writing tasks over time may be more beneficial than a one-time usage. Multiple student attempts will likely increase both multi-modal competency and the conceptual development associated with the use of these tasks. Several cases from this study provide evidence to back this assertion. In terms of raw scores, all writing scores improved for both treatment and control at sites one, two and three from unit one to unit two except the AEB scores for site three (recall site four did not have a second unit). Sites one and three both exhibited a greater degree of conceptual benefit for the treatment group after the second consecutive unit of study. In addition, for sites two and three (the sites where correlations were calculated for two consecutive units) correlations between writing scores and end of unit assessment were greater on eight of ten measures on the second unit and regression analysis indicated greater predictive power of GTA scores for the second unit assessment score. This cumulative benefit characteristic may be related to several theoretical assertions related to cognition and multi-modal activities.

Cognitive models were offered in chapter two that attempted to describe the cognition associated with both recognizing and utilizing multiple representations of a similar concept. While the particulars of each model differed, there were common features. First, all models asserted that text representations and visual representations were dealt with in separate neurological locations. When dealing with information presented in multiple ways, the text information and the visual information must first be dealt with individually and then information from the different modes must be integrated. In creating a multi-modal writing project, this may entail at minimum a consideration of the text to be written, then the appropriate alternative mode to use and its structure, and finally how the two can be integrated. This final step is the step involving the embeddedness techniques associated with the embeddedness encouraging lessons. In initial experiences with creating multi-modal products, students may struggle with any or all of these three aspects. The theoretical information in chapter two also highlighted several proposed models of cognition associated with writing tasks alone, indicating a cognitive load associated with the text production itself. Particularly if using alternative writing tasks in the science classroom is a new experience for students, it is likely that a single experience is not sufficient to allow students to deal with the task expectations in a way that maximizes learning potential. It is much more likely that multiple opportunities to attempt to create effective text, choose appropriate modes, and apply appropriate integrating techniques, are needed in order to afford students the practice necessary to maximize benefit. A first attempt at this type of activity may be too cognitively demanding to immediately result in effective multi-modal competency or to fully gain conceptual benefit.

Another cognitive action associated with utilizing multi-modal writing tasks is the idea of translation. As discussed in chapter two, translation has been offered as both a characteristic of writing-to-learn activities and of activities involving multiple representations. In terms of writing activities, “translation” has been described as a process in which student authors must change their descriptions of concepts being dealt with from the language of the instructor or science source they consult to “everyday” language for their writing. As previously mentioned, this initial translation is significantly impacted by the audience addressed. Later, the student must translate back to the language of the instructor for assessment purposes. In relation to multi-modal tasks, the idea of translation has been associated with the cognitive action of students translating the meaning of different representations of similar concepts in different modes into their own understanding. Both of these translation activities (text and multi-modal) are instrumental in an overall process in which the student assesses their own conception of a particular science idea, determines where this conception is inadequate, and then attempts to develop a more accurate conception. It is this process of personal cognition that ultimately leads to conceptual development. Like the ideas discussed in the previous paragraph, these translation processes are cognitively demanding. In instances where students are attempting to deal with brand new conceptual material in the classroom, as well as utilize a new technique such as creating a multi-modal writing product, it is likely initial attempts may not automatically initiate this translation action, but later attempts following cumulative opportunities to deal with these tasks may have a greater chance of doing so.

Differential Impact for Different Ability Levels

Evidence from this study also indicated that the relationship between embedding multiple modes in text and performance on assessments may be realized in a differential manner depending upon the science ability level of the student. The data to support this contention is limited, as ability grouping based on an outside variable not utilized in other analyses was only available for sites three and four. However, as was previously discussed in the analysis by sites, results indicated different outcomes for different ability level groups. At both sites, low achieving students had low or negative correlations between text and test performance, but positive correlations with all other writing measures and test performance. Further, the low achieving students had dramatic improvement as a group between the first and second unit. For high achieving students, the strongest correlations were between average embeddedness score and total test score and between grand total (average) and total test score, raising the possibility of these attributes being critical differentiating factors for these students. While the middle ability students did not show as much consistency between site three and four, the results from this group were somewhat different from the other ability level groups, supporting the claim of differential impact based on science ability level.

The researcher is aware of the possibility that some underlying factor is associated with both degree of embeddedness and test performance and may better explain the relationships shown in the correlation data. However, rationales related to the factors considered in this study may also explain the data. One possible explanation for the differential results by ability level, particularly those at the lower level, is related to the arguments offered for the cumulative effect. Lower achieving students my have more difficulty dealing cognitively with representations from both the text perspective and the alternative modes perspective. The strategy for many may have been to neglect one or the other as they created their multi-modal product. The data from this study would indicate the text descriptions received less focus than the modal representations especially in initial attempts. The results analyzed for low achieving students were both from sites with higher levels of implementation, therefore the greater emphasis on the techniques for embedding modes in text may have been perceived as an indication that they should focus on modes other than text to a greater degree. In addition, the dramatic changes from the first to second unit may indicate that the cumulative effect idea is quite prevalent with the low achieving students.

In terms of the high achieving students, the average embeddedness score appeared to be the most critical factor in differentiating achievement on the assessments relative to other high achieving students. It is more likely that the high achieving students were able to balance the text demands and the demands of selecting appropriate modes better than the other ability level students. The final factor discussed previously, the ability to effectively integrate or embed the modes with the text, may have been the factor that conceptually set some of these higher achieving students from others. This may support the earlier contention that the process of learning to effectively use multi-modal representations progresses in stages and the integration factor is difficult to accomplish in initial attempts, even for higher ability students. However, if the embeddedness techniques are utilized effectively, conceptual benefit is attained. This was also the correlation that showed the most improvement from unit one to unit two, indicating this may have been where the higher-level students focused their attention as they produced their second consecutive multi-modal product.

A potential overall explanation for the differentiation in terms of ability levels stems from the cognitive load ideas related to cognition posited in chapter two. The

Cognitive Load Theory and the Integrated Cognitive Load and Generative Model of Cognition asserted that as students consider multiple representations, there is a limit placed on their cognitive abilities by the size of their working memory. Three types of information are processed according these ideas: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. Intrinsic cognitive load is a fundamental characteristic of the topic and concepts considered. This is termed “intrinsic” because it is relatively stable and not affected by instructional practices. Extraneous cognitive load is cognitive load as a result of instruction that does not necessarily help with the attainment of conceptual understanding, but nonetheless takes up working memory capacity. Finally, germane cognitive load is cognitive load that is a result of instruction and does help the student attain conceptual understanding. The key issue relative to learning is the ratio of extraneous to germane cognitive load once the capacity requirement for the working memory is determined.

In terms of ability level differences, it is likely that for different ability levels, different “typical” capacities of working memory exist. Logically, higher achieving students, with greater science conceptual understanding, would need less working memory capacity to deal with the intrinsic cognitive load, leaving more working capacity to deal with issues associated with the production of the multiple modes and integrating these with text. In addition, the ratio of extraneous to germane cognitive load would be impacted by what items are perceived by the individual students to be extraneous and germane. Therefore, if a lower achieving student were to focus on using modes instead of text, to that student text is extraneous, while to other students it might be germane information. The differential impacts on different ability groups of these multi-modal tasks may result from differential allocation of cognitive load in working memory, perhaps based on different decisions about what is germane to the task that may be manifested in differential conceptual outcomes.

Summary

The results discussed here support a link between effective instruction dealing with the idea of embedding multiple modes in text, effectively embedding multiple modes in written text, and student conceptual attainment. However, this conceptual attainment is not a guaranteed aspect of participating in lessons that encourage this embeddedness or of creating a multi-modal product. In order to maximize the potential for benefit, effective implementation of lessons encouraging embeddedness and production of effective assignments calling for creating of multi-modal products is needed. In addition, students must be given multiple opportunities to experience creating multi-modal products. Finally, the science ability level of the student may affect what aspects of the product they will focus their attention on and may impact how beneficial the activity is for an individual. In the next chapter, implications arising from these findings will be discussed as a way to encourage further research.

CHAPTER SIX

SUMMARY AND CONCLUSIONS

In this chapter, a summary of the research undertaken for this study will be presented along with a discussion of limitations to the study and implications from the study for further research. Specifically, the research questions and design, along with the results will be reviewed. Following that discussion, limitations of the study will be addressed. Finally, implications arising from the study in terms of pedagogy and research will be suggested.

Research Questions

The following questions guided the research presented here:

1. Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to a greater degree of embeddedness in student writing?
2. Does encouraging students to embed multiple modes of representing science information with text in writing tasks lead to greater conceptual understanding as measured by end of unit assessments?
3. Can correlations between degree of embeddedness in writing and student performance be detected in a variety of classroom settings?

Research Design

A quasi-experimental quantitative design was utilized to study the impact of embedding multiple modes of representing science information in student writing. Four groups participated, with each group being comprised of existing general chemistry classes taught be a specific instructor. The number of actual classes in each group varied from two to five, but in all cases teachers randomly assigned some of the classes to the treatment group and the others to the control group. All students were assessed prior to the first unit of study utilized in this research with a baseline science competency exam. Students in three of the four groups were also assessed for baseline degree of embeddedness in their writing prior to initiation of the study.

Following baseline assessment procedures, treatment students participated in a lesson designed to specifically highlight the use of multiple modes of representing science information and strategies to integrate the modes with written text. Control groups did not participate in this lesson. All students at each unique site were assigned an identical writing task and all students in each group were assessed with identical end of unit exams. Analysis of student performance tested for differences due to treatment for student performance on the end of unit exams, as well as differences due to treatment in the degree of embeddedness employed in student writing. This provided the quantitative group data for the study. Additionally, analysis at the level of individual students was undertaken to compute correlations between degree of embeddedness in writing and student performance on exams, as well as prediction of student performance on exams from degree of emebeddedness and from baseline science competency. In three of the four groups, this writing and exam cycle was completed a second time for the next consecutive unit of study.

Results

- Significant differences were noted between some groups of students who participated in embeddedness encouraging activities and groups who did not in terms of degree of embeddedness in writing. Cases where teacher implementation of the embeddedness encouraging activities was high had a greater chance of these differences occurring, as did the second consecutive unit.
- Significant differences were noted between some groups of students who participated in emebddedness encouraging activities and groups who did not in

terms of conceptual understanding as measured by end of unit assessments. Again, higher levels of implementation and multiple opportunities increased the chances for these differences.

- Positive correlations were found in a vast majority of cases between writing measures of embeddedness and student performance on end of unit assessments.

Limitations

The researcher recognizes several limitations to the quantitative aspects of the study. First, the statistical assumption of random assignment was violated at both the level of the teachers involved and the students. Teachers were approached by the researcher and agreed to participate. It is likely that the participating teachers' willingness to participate indicates an interest in the utilization of writing tasks in the science classroom. This interest on the part of the teachers may have been manifested in their science classrooms valuing and employing writing differentially than many science classrooms. The results obtained here may have been influenced by these classroom environments and may not be able to be generalized to all science classrooms, particularly those with instructors who utilize little or no written work.

Likewise, student assignment was not random, but was a function of pre-existing classes. In many secondary settings, class assignment may be influenced by scheduling factors, including extra-curricular activities and other participation opportunities for students. Therefore, utilizing intact classrooms does present the potential for a group effect confounding the data analysis. While baseline assessments were undertaken to test the equivalency of groups in terms of science conceptual understanding and degree of embeddedness in student writing, the researcher recognizes that not all potential covariates could be identified and controlled for in analysis and therefore highlights this lack of random assignment as a limitation.

The practical implementation of the intended research design presented a series of limitations to the study. The original research design involved all classes from all groups participating in identical writing assignments and end of unit assessments for identical conceptual units of study. This would have led to a 2 X 2 fixed effects design that would have allowed for analysis of the data for differences by treatment, by teacher, or by interaction between these two variables. Practical developments in the course of the research, including schedules of the participating schools and teachers made it impossible for this congruency among groups to be attained, resulting in analysis taking place on a case by case basis. This case by case analysis did allow the researcher to note trends that existed regardless of the topics used, the timing of the topics, or the specifics of the writing assignments. However, the researcher also recognizes that by having different topics, different writing assignments, and different assessments used in the different groups, the ability to provide any sort of summary data comparing groups was significantly limited.

Finally, the embeddedness rubric utilized to assess student writing samples provides a limitation to this study. This rubric has been developed by the researcher through consultation with other researchers studying multi-modal writing tasks and refined through experiences in pilot studies. The refinement of this tool is ongoing. Each opportunity to utilize the assessment tool provides input regarding ways to improve the assessment device. It is highly unlikely that all characteristics of embeddedness have been accounted for by the rubric. In addition, the optimum combination of the subsections to meaningfully represent a quantitative measure of student embeddedness is also the subject of continued debate. While the use of this rubric has some justification based in the theoretical literature and past use, the researcher realizes that complete justification of this tool has not been attained.

Implications

The implications arising from this study can broadly be categorized into those important for teachers and those important for researchers. Ideally, collaboration among teachers and researchers will result in a combined approach to implementing effective, research based practices in classroom settings to realize the benefits suggested in this study.

For teachers, the main implication is the realization of the importance of connecting student production of alternative modes of communicating science information with text. It appears that simply requesting that students utilize a mode other than text, or offering students the opportunity to utilize modes other than text does not automatically lead to a situation in which students develop a better understanding of material. From a pedagogical standpoint, teachers who wish to utilize multi-modal writing tasks must realize that to achieve greater benefit, time and effort must be devoted to allowing student exploration of how multiple modes can be integrated with and embedded in text. This consideration does not naturally occur for students. Purposeful instructional practices are needed to explicitly point out occurrences of effective embeddedness in common science sources, as well as opportunity to discuss and describe why the strategies employed and reviewed are effective. Realization of this integral process for learning necessitates not only a greater allowance of time for these tasks in classrooms, but also for greater planning on the part of the instructor.

Connected to this realization of the importance of explicit ties between modes and integration in text are two other issues. First, ideally this realization will result in teachers recognizing naturally occurring opportunities to point out connectedness between different modes. One very logical source for these connections is in textbooks. Textbook use can be enhanced if teachers see textbooks as not only a source of content, but also as a source of information about how science information can be effectively communicated. The textbook (or other science source information) can become a model for students of strategies to effectively integrate multiple modes for an outside audience as they prepare their own written multi-modal products. Secondly, the recognition of integration of modes can be used by teachers as a way to encourage students with different learning styles and different learning abilities to explore concepts in modes more appropriate for them. The recognition of differential presentations of material in sources already present in the classroom can provide a mode of differentiated instruction not currently utilized.

The results of this study also indicate that when teachers utilize multi-modal tasks, they must approach the tasks with flexibility. It appears from the results discussed here that the benefit from students creating multi-modal tasks may take some time to develop. Students may need multiple exposures to these types of tasks before true conceptual benefit is realized. In addition, different ability level students may be impacted differently and on a different time frame. Teachers of classrooms with a wide range of ability levels must understand that certain aspects of the tasks may be more or less germane to some students. This may result in teachers assessing the outcomes of the tasks in different ways for different students.

Several further questions for researchers arise from this investigation as well. One of the most important areas of further research may be the link between these types of writing tasks and assessment. In many cases throughout this study, the type of assessment question combined with the multi-modal experience may have impacted the performance of students. For example, there were cases where treatment groups outperformed control groups on multiple choice questions. The multiple choice questions in which treatment groups performed better tended to have many instances where questions referred to diagrams, charts, or pictures. One question arising from this is whether students who experience opportunities to explore multiple modes and integrating them in text are better prepared to deal with particular types of assessment questions.

Factors related to the characteristics of the student products may provide interesting avenues for further research as well. One characteristic considered a positive embeddedness technique is the production by a student of an original alternative mode, as opposed to simply copying a pre-existing mode from another source. It is likely that different cognitive activity is associated with these different techniques for attaining multiple modes and therefore, would lead to different conceptual outcomes. In addition, most students when using an alternative mode, link the mode to a single idea or concept. For example, in the writing products on atomic structure, students were asked to discuss three main ideas: the history of our understanding of the atom, the subatomic particles in an atom, and the link between information in the periodic table and atomic structure. Most students when utilizing a mode other than text used a mode that referred to only one of these main ideas. Further research may begin to determine if the ability to link one particular mode to more than one main idea or concept is more conceptually beneficial than a one to one match between concept and mode. Neither of these connections was specifically dealt with in this study, but could certainly provide impetus for further study.

Finally, the differential outcomes experienced by the different ability level students raises some important questions for further research. The most obvious question is whether different ability level students necessitate different type of instruction to most effectively encourage embeddedness in their multi-modal products, and to attain conceptual benefit. The level of cognitive engagement necessitated by the consideration of the many facets of using and integrating multiple modes related to the cognitive abilities of different students may be a determining factor in the type of instruction that would most effectively benefit different types of students. In addition, the number and types of different modes required of students may be variables impacting student performance as well. Lower ability students may not be as able to develop a multi-modal competency if they are required to use too many modes or are given open-ended requirements instead of having more focused instructions on the requirements.

Conversely, higher ability students may be limited in their ability to fully develop their multi-modal competency and their abilities to integrate and embed modes in text if they are given too focused requirements. As previously suggested, assessment of multi-modal tasks may also need to be differentiated for different students, and the assessment link is one that not only needs to be considered by teachers, but also by researchers.

APPENDIX A

SAMPLE EMBEDDEDNESS LESSON

OVERVIEW

This activity has been designed to help students engage with the idea of how best to communicate about science information. Specifically, students will have the opportunity to explore different ways to integrate multiple modes of representing science concepts. One aspect that is emphasized is that there is a difference between simply using several different modes to communicate information about a particular topic and using different modes in a way that they are integrated together to communicate an understanding about a topic. The general goal of the activity is to help students recognize ways that sources of information about science integrate different modes and compare the effectiveness of these different strategies. Ultimately, students will be asked to develop an evaluation tool that they can use to analyze their own ability to embed multiple modes of representation into a written task. The particular lesson presented here was used during a unit in which the concept of classification of matter was being discussed. The topic and the sources that demonstrate the different modes could be manipulated to fit any other particular situation.

Big Questions:

- How can you effectively communicate your understanding about science concepts by using different modes of representation?
- How can you effectively integrate these modes together to help your audience better understand your ideas?
- What are the most important factors in determining what modes to use and how to combine them?

Outline of Activity:

Day One

Read Wikipedia article about States of Matter & respond individually to the following questions:

- ? = What are the strengths and weaknesses of this article in terms of communicating to you what the big idea is
- ? = What about the article was difficult for you
- ? = How could you improve the effectiveness of communicating this information in writing
-

Individual students create a list of ALL the possible modes they can think of to communicate understanding in a written format (pictures, graphs, tables, diagrams, math equations, etc.)

- Share the list with a partner
- Create a master list on board

Examine the textbook sections on states of matter from several different general chemistry textbooks with groups of 3-4 and answer the following on chart paper:

- ? = What modes does your text use?
- ? = How does your text tie them together?
- IS THIS EFFECTIVE – Why or Why not?

Class Discussion

- Share responses to previous questions & discuss
- Determine factors that would influence the modes to use

  E.g. – Audience, Technology Available for presentation, topic

Show Website (or other sample of topic from a unique source)

(www.chem.purdue.edu/gchelp/atoms/states.html)

What audience is this aimed at? Why is it effective or not?

Group Activity

- Groups of 3-4 students select one of a list of possible audiences (3rd graders, Junior High Students, High School Students, College Students, College professors, parents) and a method of presenting (book, overhead, website, other).
- Groups create a plan for a well integrated, multi-modal presentation about Classification of Matter that demonstrates appropriate consideration of factors we discussed on overhead transparency or poster

Day Two

Present the Transparencies / Posters

- Class members keep track of characteristics that made them well integrated and informative, characteristics that made them difficult to understand as they view other presentations

Groups create a checklist to determine if a particular written piece is a well integrated, multi-modal representation of the science information

- Display Checklists on Overhead
- Present
- Create a Class Checklist we can all live with

ASSIGN: Find a page in textbook and one other source about same topic and assess with our checklist

APPENDIX B

SAMPLE STUDENT CHECKLIST

TYPES OF MODES USED (Note amount of each):

______ Text (Required)	______ Pictures
______ Graph	______ Table
______ List	______ Diagram
______ Math	______ Animation

INTEGRATING TEXT WITH MODES OF REPRESENTATION:

______ Captions on Pictures

______ Texts refers to pictures (e.g. See figure 1)

______ Alternative modes (other than text) spread throughout

______ Bold, italics, underline words important to text & modes

______ Define key terms displayed in modes in text

AUDIENCE CONSIDERATIONS

______ Vocabulary appropriate for audience

______ Pace & "Density" of product appropriate for audience

______ Alternative modes appropriate for audience

COMMENTS:

APPENDIX C

SAMPLE STUDENT WRITING TASK

Dear Chemists-

Hello current chemists and former junior high scientists. I am writing you today with a request. Recently, we encountered a problem in our junior high science department. Some hoodlums from another Junior High broke into our storeroom and, in an attempt to sabotage our science program, mixed several items in our chemical storeroom together. The scientific saboteurs got away and we now have several mixtures of substances in our storeroom. Fortunately, the juvenile delinquents who perpetrated this crime left the empty bottles next to the mixtures they made. Therefore, we know what substances have been mixed, but we have several questions:

1) How should we attempt to separate the mixtures that have been created? What techniques are available to us and in what situations are each appropriate and why?

2) How can we tell if there has been a chemical change that has taken place when the items were mixed? What evidence would indicate that the chemistry of the items has changed? Why would these characteristics indicate chemical changes?

3) How could we arrange our stockroom so that we have an organization system based on chemical properties? Why would this be better than a system based on physical properties?

If you could please write us back and explain what you think we should do, we would appreciate it.

Thanks for the help,

Dr. Whitinok & The North Central Scientists

APPENDIX D

RESEARCHER GENERATED RUBRIC

TEXT	**All(3)**	**Most(2)**	**Some(1)**	**None(0)**
Grammatically Correct	______	______	______	______
Accurate	______	______	______	______
Covered Required Topics	______	______	______	______
Thorough	______	______	______	______

OVERALL TEXT SCORE ______

MODAL REPRESENTATIONS:

Number of DIFFERENT Modes Used (other than text) ______(a)

____ Picture ____ Graph ____ Table ____ List ____ Diagram ____ Math

Number of TOTAL Modal Representations ______(b)

Number of INAPPROPRIATE Representations ______(c)

Number of TOPICS Related to Modal Representations ______(d)

OVERALL SCORE = (b – c) + a + d = ______

INDIVIDUAL MODAL REPRESENTATIONS:

KEY: (N) = Next to Text (R) = Referred to in text (A) = Accurate

(C) = Complete (CA) = Caption (O) = Original

1) TYPE________________

N____ R____ A____ C____ CA____ O____ TOTAL ______

2) TYPE________________

N____ R____ A____ C____ CA____ O____ TOTAL ______

3) TYPE________________

N____ R____ A____ C____ CA____ O____ TOTAL _____

4) TYPE________________

N____ R____ A____ C____ CA____ O____ TOTAL ______

5) TYPE________________

N____ R____ A____ C____ CA____ O____ TOTAL ______

6) TYPE

N____ R____ A____ C____ CA____ O____ TOTAL ______

OVERALL SCORE_______ # of MODES______ AVG. EB SCORE ______

GRAND TOTAL (RAW)_______ GRAND TOTAL (AVG) _______

REFERENCES

Alters, B. J. (1997). Whose nature of science? *Journal of Research in Science Teaching, 34*(1), 39-55.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy: Project 2061*. New York: Oxford University Press.

Appleton, K. (1993). Using theory to guide practice: Teaching science from a constructivist perspective. *School Science and Mathematics, 93*, 269-274.

Ausubel, D. P. (1967). *Learning theory and classroom practice*. Toronto, ON: Ontario Institute for Studies in Education.

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

Bernsen, N. (1993). *A research agenda for modality theory.* Unpublished manuscript.

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). In Bransford J., Brown A. L. and Cocking R. R. (Eds.), *How people learn: Brain, mind, experiences and school* (1st ed.). Washington, D.C.: National Academy Press.

Bybee, R. W. (1995). Achieving science literacy. *The Science Teacher, 62*(7), 28-33.

Byrnes, J. P., & Fox, N. B. (1998). The educational relevance of research in cognitive neuroscience. *Educational Psychology Review, 10*(3), 297-341.

Cavagnetto, A. (2006). Setting the question for inquiry: The effects of whole class vs. small group on student achievement in elementary science. University of Iowa).

Cleminson, A. (1990). Establishing an epistemological base for science teaching in light of contemporary notions of the nature of science and how children learn science. *Journal of Research in Science Teaching, 27*, 429-445.

Cobb, P. (1994). Where is the mind? constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7), 13-20.

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

Duffy, T., & Cunningham, D. (1996). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 170-198). New York: Macmillan.

Duschl, R. A. (1985). Science education and philosophy of science. *School Science and Mathematics, 87*, 541-555.

Duschl, R. A. (1988). Abandoning the scientistic legacy of science education. *Science Education, 72*, 51-62.

Eliam, B., & Poyas, Y. (2008). Learning with multiple representations: Extending multimedia learning beyond the lab. *Learning and Instruction, 18*, 368-378.

Eliot, C. (1889). *Educational reform*. New York: Century.

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

Ertmer, P., & Newby, T. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performanc Improvement Quarterly, 6*, 50-72.

Eve, R. A., & Dunn, D. (1990). Psychic powers, astrology, and creationism in the classroom? evidence of psuedoscientific beliefs among high school biology and life science teachers. *The American Biology Teacher, 52*, 10-21.

Flower, L., & Hayes, J. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication, 31*, 21-32.

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education, 92*, 404-423.

Galbraith, D. (1999). Writing as a knowledge-constituting process. In D. Galbraith, & M. Torrance (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 139-159). Amsterdam: Amsterdam University Press.

Galbraith, D., & Torrance, M. (1999). *Knowing what to write: Conceptual processes in text production*. Amsterdam: Amsterdam University Press.

Gallagher, J. (1971). A broader base for science teaching. *Science Education*, *55*, 329-338.

Gowin, D. B. (1981). *Educating*. Ithaca, NY: Cornell University Press.

Gunel, M., Hand, B., & Gunduz, S. (2006). Comparing student understanding of quantum physics when embedding multimodal representations into two different writing formats: Presentation format vs. summary report format. *Science Education, 90*(6), 1092-1112.

Hand, B., Prain, V., & Yore, L. (2001). Sequential writing tasks' influence on science learning. In G. Rijlaarsdam, P. Tynjala, L. Mason & K. Lonka (Eds.), *Studies in writing volume 7, writing as a learning tool:Integrating theory into practice* (pp. 105-129). Dordrecth, the Netherlands: Kluwer Academic Publishers.

Henriques, L. (1997). *A study to define and verify an interactive constructive elementary school science teaching.* Unpublished University of Iowa, Iowa City, IA.

Hildebrand, G. (2004). *Hybrid writing genres: A link between pleasure and engagement.* Unpublished manuscript.

Hodson, D. (1986). Rethinking the role and status of observation in science education. *Journal of Curriculum Studies, 18*, 381-396.

Hofstein, A., & Yager, R. E. (1982). Societal issues as organizers for science education in the 80s. *School Science and Mathematics*, *82*, 539-547.

Holliday, W., Yore, L., & Alvermann, D. :. (1994). The reading-science writing-learning connection: Breakthroughs, barriers, and promises. *Journal of Research in Science Teaching*, *31*, 877-893.

Hurd, P. (1970). *New directionsin teaching secondary school science*. Chicago: Rand & McNally.

Hurd, P. (1998). Scientific literacy: New minds for a changing world. *Science Education, 82*, 407-416.

Hurd, P. D. (1958). Science literacy: Its meaning for american schools. *Educational Leadership, 16*(1), 13-16.

Johnson, R. L., & Peeples, E. E. (1987). The role of scientific understanding in college: Student acceptance of evolution. *The American Biology Teacher, 49*, 93-98.

Kelly, G. J., & Green, J. (1998). The social nature of knowing: Toward a sociocultural perspective on conceptual change and knowledge construction. In B. Guzzetti, & C. Hynd (Eds.), *Perspectives on conceptual change: Multiple ways to understand knowing and learning in a complex world* (pp. 145-181). Mahwah, NJ: Erlbaum.

Keys, C. W. (1999). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Science Education, 83*, 115-130.

Kimball, M. E. (1967). Understanding the nature of science: A comparison of scientists and science teachers. *Journal of Research in Science Teaching, 5*, 110-120.

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

Kromhout, R., & Good, R. (1983). Beware of societal issues as organizers for science education. *School Science and Mathematics, 83*, 647-650.

Lederman, N. G., & O'Malley, M. (1990). Students' perceptions of tentativeness in science: Development, use, and sources of change. *Science Education*, *74*, 225-239.

Lemke, J. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex Publishing.

Matthews, M. R. (1994). *Science teaching: The role, history and philosophy of science*. New York: Routledge.

Mayer, R. (1996). Learners as information processors: Legacies and limitations of educational psychology's second metaphor. *Educational Psychologist*, *31*(3/4), 151-161.

Mayer, V. (1997). Global science literacy: An earth system view. *Journal of Research in Science Teaching, 34*(2), 101-105.

McCarthy, S., & Raphael, T. (1992). Alternative research perspectives. In J. A. Irwin, & M. A. Doyle (Eds.), *Reading / writing connections: Learning from research* (pp. 2-30). Newark, NJ: IRA.

McComas, W. (1996). Ten myths of science: Reexamining what we think we know about the nature of science. *School Science and Mathematics, 96*(1), 10-16.

McComas, W. (2004). Keys to teaching the nature of science: Focusing on NOS in the science classroom. *Science Teacher, 71*(9), 24-27.

McCurdy, R. (1958). Toward a population literate in science. *The Science Teacher, 25*, 366-408.

Miller, J. D. (1996). Science literacy for effective citizenship. In R. E. Yager (Ed.), *Science / technology / society as reform in science education* (). Albany, NY: SUNY.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, D.C.: U.S. Government Printing Office.

National Research Council. (1996). *National science education standards*. Washington, D.C.: National Academy Press.

National Science Teachers Association. (1971). NSTA position statement on school science education for the 70's. *The Science Teacher, 38*, 46-51.

National Society for the Study of Education. (1960). *Rethinking science education: Fifty-fifth yearbook of the NSSE*. Chicago: University of Chicago Press.

Norris, C. (1997). *Against relativism: Philosophy of science, deconstruction and critical theory*. Malden, MA: Blackwell Publishers.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to science literacy. *Science Education, 87*(2), 224-240.

Novak, J. D. (1977). *A theory of education*. Ithaca, NY: Cornell University Press.

Osborne, J. (1996). Beyond constructivism. *Science Education, 80*(1), 53-82.

Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher, 24*(7), 5-12.

Pineda, L., & Garza, G. (2000). A model for multimodal reference resolution. *Computational Linguistics, 26*(2), 139-193.

Prain, V. (2006). Learning from writing in secondary science: Some theoretical and practical implications. *International Journal of Science Education, 28*(2), 179-201.

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

Prawat, R. S. (1999). Cognitive theory at the crossroads: Head fitting, head splitting, or somewhere in between? *Human Development, 42*, 59-77.

President's Scientific Research Board. (1947). *Science and public policy (vol. 1 and 4)*. Washington, D.C.: Government Printing Office.

Reveles, J., Cordova, R., & Kelly, G. (2004). Science literacy and academic identity formulation. *Journal of Research in Science Teaching, 41*(10), 1111-1144.

Rowell, P. A. (1997). Learning in school science: The promises and practices of writing. *Studies in Science Education, 30*, 19-56.

Ruddell, R., & Unrau, N. (1994). *Theoretical models and processes of reading*. Newark, DE: International Reading Association.

Rutherford, F., & Ahlgren, A. (1990). *Science for all americans*. New York: Oxford University Press.

Ryan, A. G., & Aikenhead, G. S. (1992). Students' preconseptions about epistemology of science. *Science Education, 76*, 559-580.

Shamos, M. (1995). *The myth of science literacy*. New Brunswick, NJ: Rutgers University Press.

Stanley, W., & Brickhouse, N. (2001). Teaching science: The multicultural question revisited. *Science Education, 28*, 35-49.

Stenhouse, D. (1985). *Active philosophy in education and science*. London: George Allen & Unwin.

von Glasersfeld, E. (1989). Cognition, construction of knowledge and teaching. *Synthese, 80*, 121-140.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wallace, C., Hand, B., & Prain, V. (2004). *Writing and learning in the science classroom*. Dordrecht, The Netherlands: Kluwer Academic Press.

Yager, R. E. (1991). The constructivist learning model. *The Science Teacher, 67*(1), 44-45.

Yore, L., Bisanz, G., & Hand, B. (2003). Examining the literacy component of science literacy: 25 years of language and science research. *International Journal of Science Education, 25*, 689-725.

Yore, L., & Shymansky, J. (1997). *Constructivism: Implications for teaching, teacher education, and research-breakthroughs, barriers, and promises.* Unpublished manuscript.

Yore, L., & Treagust, D. (2006). Current realities and future possibilities: Language and science literacy - empowering research and informing instruction. *International Journal of Science Education, 28*(2), 291-314.

Zahorik, J. (1995). *Constructivist teaching. fastback 390*. Bloomington, IN: Phi Delta Kappa.