EXAMINING THE INFLUENCE OF INSTRUCTIONAL STRATEGY ON STUDENT LEARNING AND SELF-EFFICACY IN SCIENCE

Carolyn Hushman

Follow this and additional works at: https://digitalrepository.unm.edu/educ_ifce_etds

Recommended Citation
Carolyn Hushman
Candidate

Educational Psychology: Individual, Family, and Community Education
Department

This dissertation is approved, and it is acceptable in quality and form for publication

Approved by the Dissertation Committee:

Scott C. Marley, Chairperson

David Atencio

Terri Flowerday

James P. Selig
EXAMINING THE INFLUENCE OF INSTRUCTIONAL STRATEGY ON STUDENT LEARNING AND SELF-EFFICACY IN SCIENCE

BY

CAROLYN HUSHMAN

B.S. Science Education, University of Oklahoma, 2000
M.A. Earth Science, University of Northern Colorado, 2004
M.S. Physical Education: Outdoor Education, University of Northern Colorado, 2004

DISSERTATION

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
Educational Psychology

The University of New Mexico
Albuquerque, New Mexico

December, 2011
DEDICATION

To Anni and Kate,
your wonder and curiosity leave me in awe.
ACKNOWLEDGEMENTS

First, I would like to thank my advisor and chair, Dr. Scott C. Marley. He generously shared his time, offering feedback on my ideas and helping me become a better writer in the process. He always had a positive word, or a good story, to help me stay the course when times got rough. Dr. Marley has helped me to become a better researcher and scholar.

I would also like to thank my committee members for their time spent reading and offering their thoughts. Dr. Terri Flowerday has been instrumental to my success as a graduate student. Her courses were my introduction to the field, and her encouragement helped me finish the program. Dr. James P. Selig has been a calming presence throughout this process. His clear explanations and willingness to share his perspective have been instrumental to my success. Dr. David Atencio has been unwavering in his enthusiasm and support for my project. His positive energy has been contagious.

I would also like to take this space to thank my fellow graduate students. While working together in Dr. Roxana Moreno’s lab, Sara Abercrombie, Dana Booker, and Cara Farnell and I spent countless hours discussing salient aspects of research. Their friendship and conversation helped me prepare for this project. Julie Sanchez was always a shoulder to lean on as we figured out how to be mothers and graduate students. Finally, I would like Kira Carbonneau. Words are not enough to express my appreciation for her insightful ability to talk through research issues when needed or knowing when a good workout was the only solution.

My family has also been greatly supportive during my dissertation and graduate career. My parents were always willing to entertain and play with the girls when I needed
more time. My daughters, Anni and Kate, were always there to remind me life was much
closer than this project. They have taught me to be more balanced in my approach to
work. My husband, Glenn, provided encouragement, a sounding board, and countless
cups of tea as I completed the writing. I look forward to continuing our careers together.

Finally, I would like to thank Dr. Roxana Moreno and Dr. Gloria Napper-Owen,
both who have been great mentors throughout my graduate school experience. Dr.
Napper-Owen has always offered sound career advice and unyielding support for me as I
learned to balance graduate school and parenthood. I feel fortunate to have her as a model
of what it means to be both a successful researcher and passionate educator. Dr. Moreno
also was unyielding in her support of me. She offered sound advice and insight as Glenn
and I began our journey to start two careers in academia. As my advisor she challenged
me to question research, to work harder, and spend time doing what I loved outside of
graduate school. She is missed.
EXAMINING THE INFLUENCES OF INSTRUCTIONAL STRATEGY ON
STUDENT LEARNING AND SELF-EFFICACY IN SCIENCE

by

Carolyn Hushman

B.S. Science Education, University of Oklahoma, 2000
M.A. Earth Science, University of Northern Colorado, 2004
M.S. Physical Education: Outdoor Education, University of Northern Colorado, 2004

DOCTOR OF PHILOSOPHY EDUCATIONAL PSYCHOLOGY

ABSTRACT

This study investigated whether the level of instructional guidance affected student learning and science self-efficacy when nine- and ten- year old children learn to design unconfounded experiments using control of variable strategies (CVS). Specifically, the goal of this study was to replicate and extend prior research that examines the impact of the level of guidance on ability to successfully design scientifically credible experiments that isolate the relationship between two variables (Klahr & Nigam, 2004). Sixty children, who were enrolled in a summer sports program, were randomly assigned in equal numbers to one of the three following conditions: 1) guided instruction, where children received instruction with examples and explanations; 2) direct instruction, where children received instruction through lecture with examples; and 3) discovery learning, where children received instruction through methods of self discovery. Before starting the experiment, participants completed a prior knowledge pretest and a science self-efficacy measure. After receiving a common introduction to scientists’ use of experiments,
children received their assigned instructional treatment and completed the outcomes. The outcomes included: 1) design of experiments; 2) a recall measure about designing unconfounded experiments using CVS; 3) an application measure about applying CVS to scientific experiments; 4) an evaluation measure using CVS to evaluate a scientific experiment; and, 5) science self-efficacy. Results were analyzed using one-way ANOVAs and repeated-measures ANOVA. Children who received guided instruction designed a greater percentage of experiments correctly and had greater changes in self-efficacy relative to direct instruction and discovery learning. Children receiving direct instruction and guided instruction performed better than children who learned through discovery on all other learning measures. Results were interpreted in terms of theoretical and educational implications.
# TABLE OF CONTENTS

LIST OF FIGURES ............................................................................................................. xi

LIST OF TABLES ............................................................................................................... xii

CHAPTER 1: INTRODUCTION ............................................................................................... 1
  What Does a Scientifically Literate American Know? ..................................................... 3
  What Are the Appropriate Ways to Instruct Science Within our Schools to Develop
  Scientifically Literate Individuals? ................................................................................. 6

CHAPTER 2: REVIEW OF LITERATURE ................................................................................. 9
  Overview of the Review of Literature ............................................................................ 9
  Instructional Strategies .................................................................................................. 9
  Self-Efficacy in Science Education .................................................................................. 28
  The Present Study ......................................................................................................... 34

CHAPTER 3: METHODS ...................................................................................................... 37
  Study Overview .............................................................................................................. 37
  Pilot Study ...................................................................................................................... 37
  Primary Study ............................................................................................................... 39

CHAPTER 4: RESULTS ....................................................................................................... 49
  Analysis Overview ........................................................................................................ 49
  Preliminary Analyses .................................................................................................... 49
  Primary Analyses ......................................................................................................... 52

CHAPTER 5: DISCUSSION .................................................................................................. 58
  Study Overview ............................................................................................................ 58
Theoretical Implications ........................................................................................................... 62
Educational Implications .......................................................................................................... 63
Strengths and Limitations ......................................................................................................... 65
Future Directions ....................................................................................................................... 67
Conclusion .................................................................................................................................. 70

References .................................................................................................................................... 71

Appendix A: Session One Script for Pilot and Primary Study .................................................. 81
Appendix B: Direct Instruction Script, Pilot and Primary Study .................................................. 82
Appendix C: Guided Instruction Script, Pilot and Primary Study ................................................. 84
Appendix D: Discovery Script, Pilot and Primary Study .............................................................. 86
Appendix E: Introduction Session Two Script, Pilot ................................................................. 87
Appendix F: Instructions for Designing Experiments and Learning Outcomes, Pilot .... 88
Appendix G: Study Evaluation of Science Experiments, Pilot ................................................... 89
Appendix H: Demographic Questions, Pilot and Primary Study ................................................. 92
Appendix I: Pretest, Pilot and Primary Study ............................................................................ 93
Appendix J: Pretest Item Analysis ............................................................................................ 95
Appendix K: Science Self-Efficacy Questions, Pilot and Primary Study ................................. 96
Appendix L: Science Self-Efficacy Item Analysis ...................................................................... 98
Appendix N: Recall of CVS, Pilot and Primary Study .............................................................. 101
Appendix O: Recall Item Analysis ........................................................................................... 102
Appendix E: Application of CVS, Pilot and Primary Study............................................. 103
Appendix Q: Item Analysis-Application ........................................................................... 105
Appendix R: Evaluation of Science Experiments, Primary Study ................................. 106
Appendix S: Evaluation Item Analysis............................................................................ 107
Appendix T: Introduction Session Two Script, Primary Study ...................................... 108
Appendix U: Analysis of Pretest..................................................................................... 109
Appendix V: Analysis of Time......................................................................................... 110
Appendix W: Analysis of Designed Experiments ........................................................... 111
Appendix X: Analysis of Learning Outcomes................................................................. 112
Appendix Y: Analysis of Changes in Science Self-Efficacy.......................................... 113
LIST OF FIGURES

Figure 1. Image of experimental ramp ................................................................. 50

Figure 2. Mean science self-efficacy at pretest and posttest by condition............. 56
LIST OF TABLES

Table 1. Demographic information for the primary study ........................................ 41
Table 2. Descriptive statistics for time in the pilot and main study ............................ 50
Table 3. Inter-item correlation matrix for outcome measures .................................. 52
Table 4. Means by condition on learning outcome measures ................................. 53
CHAPTER 1
INTRODUCTION

According to the 2009 Trends in International Mathematics and Science Study (TIMSS) fourth and eighth graders in the United States rank in the 72\textsuperscript{nd} and 74\textsuperscript{th} percentiles, respectively, in international testing on science knowledge and reasoning. In addition, scores on these tests have not changed significantly in twenty years of data collection (Gonzales et al, 2009). The National Assessment of Educational Progress (NAEP) 2009 results indicate 76\% of New Mexico fourth graders and 79\% of eighth graders are below proficient in science knowledge and reasoning. New Mexico ranks 42\textsuperscript{nd} and 41\textsuperscript{st} for fourth and eighth graders, respectively, out of the 45 states participating in this assessment program. Within the state of New Mexico there are differences in science reasoning among different groups. Students who are eligible for the national school lunch program—an indication of social economic status—15\% of students are considered proficient compared to 43\% who are not eligible for these services. Only 15\% of Hispanic students are considered proficient while eight percent of Native American students and 48\% of White students are proficient (NAEP, 2009). Many experts credit the poor academic achievement of U.S. students in science on falling academic standards in science and a lack of clear and achievable goals for science curriculum (e.g. Bybee, 1997; DeBoer, 2000; National Center for Improving Student Learning and Achievement in Mathematics and Science, 1999).

Appropriate goals for science education curriculum in the United States have been debated since the inclusion of science content in the national curriculum in the early 19\textsuperscript{th} century (DeBoer, 2000). These debates have centered on whether an individual who has a
science education from this country should be able to: 1) demonstrate the attainment of science knowledge to pursue careers in science; or 2) have the ability to understand and participate in the public forum concerning science-based social issues. For example, when the Russian satellite Sputnik first orbited the earth in 1957, fears sparked a reform that was led by leading scientists in the areas of medicine, engineering, and chemistry, to refocus the aims of science education on doing science and thinking scientifically (Lazarowitz & Tamir, 1994). This reform lead to the beginning of elementary school science and secondary classrooms focused on the science laboratory instead of science textbook.

A few years later in response to further criticisms of science education, Robert Carlton (1963) the president of the National Science Teacher’s Association (NSTA) asked leading scientists and educational experts what the goals of science education should be. Experts most frequent response focused on attaining science discipline specific facts and their relationship to the real world. Carlton’s report led to a decade of reforms that were centered on exposing American students to science facts at the expense of connecting science to daily experiences (DeBoer, 2000). This emphasis on facts is in contrast to the 1970s where the focus of science became its role in producing citizens who could economically be part of a modern technologically advanced society. This switch is evidenced by NSTA stating in Science-Technology-Society: Science Education for the 1980s the goal of science education is to “develop scientifically literate individuals who understand how science, technology, and society influence one another and who are able to use this knowledge in their everyday decision-making" (NSTA, 1982, p. 250).
At the root of the science curriculum debate, are the two central questions: 1) What does a scientifically literate American know? and 2) What are the appropriate ways to instruct science within our schools to develop scientifically literate individuals (AAAS, 1989; Bybee, 1997; DeBoer, 2000; Dewey, 1966; Hurd, 1958, 1970; McCurdy, 1958; Schamos, 1995)? These two questions will be explored in further detail in the following paragraphs.

**What Does a Scientifically Literate American Know?**

The definition of science literacy has evolved from focusing on what science knowledge an individual has to the individual being able to use science to address societal issues (Hurd, 1958, 1970). In 1958, the idea of a scientifically literate citizen focused on the ability to understand how the world was changing in terms of science and technology (Rockerfeller Brother’s Fund, 1958). The emphasis in the science education curriculum became the production of citizens who could keep the country competitive in the global economic market and through the military. The seventies did not solely focus the concern on a workforce, but on a society that could apply science to evaluate data to solve global problems (Hurd, 1970). These problems included, but were not limited to, access to clean water and food for all people, more efficient use of natural resources, and advancements in medical treatment and care.

Today the definition of science sets a balance between what we know and what we can do with an emphasis on doing. The set of science standards that serve as the framework for most state science standards, *National Science Education Standards*, merges science content with the needs of society through this definition of science literacy:
Scientific literacy means that a person can ask, and, or determine answers to questions derived from curiosity about everyday experiences. It means that a person has the ability to describe, explain, and predict natural phenomena. Scientific literacy entails being able to read with understanding articles about science in the popular press and to engage in social conversation about the validity of the conclusions. Scientific literacy implies that a person can identify scientific issues underlying national and local decisions and express positions that are scientifically and technologically informed. A literate citizen should be able to evaluate the quality of scientific information on the basis of its source and the methods used to generate it. Scientific literacy also implies the capacity to pose and evaluate arguments based on evidence and to apply conclusions from such arguments appropriately. (The National Research Council, 1996, p. 22).

Another leader in the national science curriculum debate, the American Association for the Advancement of Science (AAAS), also stresses the importance of being able to evaluate scientifically produced data in terms of being trustworthy in their 1993 publication *Science for all Americans; Project 2061*. Central to the ability to evaluate experimental data is the understanding of designing unconfounded experiments. Unconfounded experiments isolate all variables and manipulate one to learn about the relationship between the independent and dependent variables (Klahr & Nigam, 2004). For example, a student experimenting to understand the relationship between the length of a ramp and the distance a ball rolls would need to ensure that everything is the same
between two ramps except the length of the ramp. Specifically, the AAAS benchmarks highlight the strategies used to design unconfounded experiments for third- through fifth-grade and sixth- through eighth-grade children. In New Mexico, designing unconfounded experiments to test a hypothesis and using the data to inform knowledge is a central tenet in standard one of the New Mexico Science Content Standards (NM Standards, 2003) for second- through eighth-grade students.

Teaching children to design unconfounded experiments is more than explaining terminology. Teachers must help students develop strategies for applying the terminology in an experimental setting. For example, if a student is asked to design an experiment for investigating how surface features affect the distance a ball rolls off a ramp, and they set two ramps up with carpet on one ramp and wood on another ramp but the ramps are of different length, they have failed to control all variables that could affect the ball distance. In this scenario, the student needs to know more than the independent, control and dependent variables mean. They need to understand the relationships between these variables. Definitional knowledge will not help students understand that the different ramp lengths and the different surface textures have confounded the conclusions that can be drawn in regards to what affects the distance the ball rolls. The strategies a student uses to design unconfounded experiments have two names in the literature, isolation of variables strategies (IVS) or control of variable strategies (CVS: Pressley & McCormick, 1995). The ability to devise unconfounded experiments is a critical skill in scientific literacy, but the best ways to develop this skill is still being investigated.
What Are the Appropriate Ways to Instruct Science Within our Schools to Develop Scientifically Literate Individuals?

While the understanding of CVS is important to modern science literacy another aspect of the national conversation over science literacy focuses on what constitutes effective science instruction. The central question is the role the teacher plays while improving students’ inquiry skills. One end of this spectrum involves students participating in “hands-on” activities as the teacher facilitates the learning by offering minimal amounts of guidance. The opposite end of the spectrum has the teacher explicitly guiding the learner through the tasks by lecturing (Banilower, Cohen, Pasley & Weiss, 2010). The following paragraphs will discuss three common levels of guidance in the science instruction literature.

For the purposes of this study, instruction will be defined as the delivery of academic information to develop thinking skills (Pressley & McCormick, 1995). Guidance during instruction includes the strategies educators use to support the learning of specific content. For example, if a teacher is delivering explicit instruction on designing unconfounded experiments they will give a detailed and organized presentation of the types of variables and their relationship to each other. The teacher may even explain several examples and non-examples in order to highlight the application of important strategies. The student receives a high amount of guidance in the sense they were told exactly what they needed to know in order to successfully design an experiment.

When thinking about instructional guidance three competing views exist. First, are learning theorists who advocate that teachers provide no instructional guidance. These
theorists propose that students learn best when allowed to construct essential knowledge from their own experiences. Instructional strategies of this nature are often referred to as discovery learning (Bruner, 1961; Inhelder & Piaget, 1958). Second, are researchers who argue that learners should be presented with explicit detailed instructions that fully explain a concept, also known as direct or explicit instruction (Pressley & McCormick, 1995). Finally, there are learning experts who support the use of guided instruction. These theorists suggest that students learn from the construction of knowledge based in experiences and that this learning is mediated by teacher guidance (Mayer, 2004).

Children are unlikely to learn and gain mastery of designing unconfounded experiments if they are not motivated to do so in the first place. Kesidou and Roseman (2002) evaluated middle school science programs and concluded that student interest or value for science must be engaged so that learning can take place. One aspect of individuals’ motivations is the judgments they make regarding their capabilities to be successful in specific situations (Bandura, 1997). For example, a student with high self-efficacy in science will view the challenge of designing unconfounded experiments as something to be mastered where as a student with low self-efficacy is more likely to see the activity as something to be avoided. While research comparing self-efficacy and levels of instructional guidance is lacking, theory suggests that any level of instructional guidance that offers pathways to success should increase a student’s self-efficacy (Ackerman, Kyllonen, & Roberts, 1999).

The present study examined whether the level of instructional guidance provided before a laboratory experience positively impacted fourth- and fifth-grade children’s learning and science self-efficacy. Specifically, the goal of this study was to compare
three levels of instructional support: discovery learning, guided instruction, and direct instruction. The instructional strategies were compared within the context of designing unconfounded experiments using control of variable strategies. The outcomes of interest were children’s abilities to design experiments, recall information concerning CVS, apply their understandings to evaluate an experiment using CVS, and the impact on the science self-efficacy were assessed. The following chapters contain: 1) a review of literature associated with the role of instructional guidance on learning in science, science self-efficacy, and the research hypotheses guiding this study; 2) the methodology used to address the hypotheses; 3) the results of both preliminary and primary analyses; and 4) a discussion situating the findings of this study within the current literature, limitations of this study, and directions for future research.
CHAPTER 2
REVIEW OF LITERATURE

Overview of the Review of Literature

The present study examined whether the instructional strategy provided before a laboratory experience positively impacted children’s learning and science self-efficacy. The review of literature examines: 1) level of instructional guidance; 2) self-efficacy in relation to science education; and 3) the gaps in the literature that the present study was designed to address and the research hypotheses guiding this study.

Instructional Strategies

In 1961, Bruner called for more instruction to be delivered through instructional strategies that capitalize on discovery learning. Bruner’s article stimulated a flurry of systematic research aimed at comparing many forms of instruction including discovery methods of learning, guided instruction in various forms, and direct instruction (Mayer, 2004). This section introduces each of these different levels of instructional guidance, the early research designed to investigate instructional strategies to teach problem solving, how instructional strategies have been applied to teaching scientific reasoning, and finally studies focused specifically on scientific reasoning strategies used to design unconfounded experiments.

Levels of instructional guidance. Instruction can mean many things, but for the purposes of this study instruction will be the delivery of academic information to develop thinking skills or content knowledge (Pressley & McCormick, 1995). The strategies educators use to deliver instruction impacts the way students learn. Three types of
instructional strategies of interest to the present discussion are: 1) direct instruction; 2) discovery; and 3) guided instruction.

Direct instruction involves the presentation of detailed information presented in a highly-structured manner that emphasizes academic concepts, explanations for demonstrations, or use of a strategy to solve a problem (Case, 1956; Pressley & McCormick, 1995). Direct instruction is thought to be effective because important aspects of the task or concept are highlighted, examples are used to illustrate the task or concept, and the when and where to use the concept or skill is provided. These three aspects of direct instruction allow students to actively build upon existing schema in efficient and effective ways (Pressley, Harris, & Marks, 1992). For example, chemistry teachers often teach students how to identify an unknown concentration of an acid using a base with a solution of known concentration through a process called acid-base titration. The teacher could use direct instruction by beginning a lesson with a demonstration of the use of the buret and explanation of this piece of equipment, followed by explaining the relationships between the hydrogen concentration, the unknown acid, the known base, and color changes at the equivalence point. Finally, this teacher could end the lesson with how to use the volumetric readings of the amount of base added to reach the equivalence point in order to calculate the concentration of the unknown acid.

Discovery learning deemphasizes the role of instruction and is based on Piagetian theories of intellectual development (Brainerd, 1978). Piaget (1970) emphasized that learning is a spontaneous processes that occurs from a student’s interactions with the external environment and not through interactions with adults or others. This line of instruction is often supported by the perspective that it is necessary for an instructional
strategy to promote motivation by cultivating natural curiosity. When natural curiosity is present, deeper and more meaningful learning is expected due to learner engagement (Bruner, 1961). In the science classroom mentioned above a teacher using discovery learning would give students the task of determining the concentration of the acid and support their discovery by providing resources and equipment such as the buret and scientific manuals.

Guided instructional strategies are more explicit than discovery teaching and more student driven than direct instruction. Often, guided instruction involves using questions, hints, feedback, or modeling that lead students to ways to solve the problem (Mayer, 2004; Pressley & McCormick, 1995). Guided instruction is expected to be an effective strategy because the relevant information is emphasized for the student while engagement in the activity is high (Shulman & Keisler, 1966). For example, the science teacher using guided instruction might give students the task of identifying the concentration of the acid but may supply them with a worksheet with a series of questions that will help them develop the strategy of using a buret with a base with a known concentration to calculate the volume of base needed to neutralize the acid.

Researchers have studied the use of instructional strategies in many different contexts, with diverse samples and many different research methodologies. Scientifically-based research comparing the just-described instructional strategies originated with the teaching of problem solving skills.

**The Early Research: Teaching Problem Solving Skills.** This section focuses on early experimental research where instructional strategy is the manipulated variable. All
the studies discussed here used direct instruction, discovery and guided instruction as the conditions of interest and problem solving as the skill being taught.

In an early study that explored teaching students to solve logic word problems Craig (1956) used word problems that involved choosing the word that did not belong in a list of five words to investigate the amount of instruction needed for learning to occur. Two treatments were examined with 106 sophomores and juniors in college. Half of the students were assigned to receive ‘directed instruction’ and were given guidance (e.g. pick the word with the odd initial sound) along with the reasoning for why the answer was correct afterwards. The students in the comparison condition were given no hints or answers. The students who received guidance during learning solved more word problems correctly than those with no guidance. In addition, those receiving guidance were able to transfer their abilities to other types of word problems, solving more of them correctly than those with no guidance. Using Cohen’s $d$ (1988) to categorize magnitude of effect, standardized differences larger than 0.2 are considered small effects, those around 0.5 are considered medium effects and those larger than 0.8 are considered large effects. The difference between conditions in this study was medium sized.

Kittel (1957) used similar study materials, but his subjects were 132 sixth-grade students. The children received one of three treatments: 1) discovery, where students were given no hints; 2) guided instruction, where students were given hints and explanations; and 3) an expository group, where students were given the rule and correct answer. Kittel screened for reading ability before assigning children to each treatment group so that each group had equal numbers of ‘low’, ‘middle’, and ‘high’ ability readers. Analysis was done on immediate retention, delayed retention (students were tested 3 days
later) and transfer to solving similar word problems. The results of this study indicated a small difference in the guided instruction group performance over the discovery group on all tasks. The guided instruction performed significantly better than the expository with medium differences observed on all tasks. Kittle explained his results in terms of guided instruction allowing students to use mental resources on relevant information instead of wasting resources on failed strategies (discovery) and not being active (direct).

In a study incorporating logic to solve number problems, Gagne and Brown (1961) gave 96 college students problems involving computing the sum of a series of numbers and filling in the missing number. Three treatments were contrasted: a discovery treatment, where participants received no instruction in solving the problems, a guided instruction treatment, where participants received hints and explanations for reasoning; and, a direct instruction treatment, where participants were told the correct answer. The guided discovery treatment solved more problems correctly and solved more transfer problems than the other groups with small- to medium differences observed. The authors were careful to note that the guided discovery treatment required more time to learn and solve the problems and the direct instruction treatment group mastered the materials more quickly.

Shulman and Keisler (1966) summarized the early research on the type of instruction and learning to apply problem solving skills. They concluded that guided instruction allowed learner involvement in constructing new knowledge, assured the appropriateness of the knowledge, and facilitated subsequent integration with existing knowledge. They proposed that the failure of discovery methods of instruction occurs when students do not learn the rule or principle under study therefore they are not
constructing appropriate knowledge or making sense of the new information. Shulman and Keisler argued that this meant learners could not integrate new knowledge into their existing schemas. And, furthermore, direct instruction interfered with learning by not allowing mental activity on the part of the learner. They also acknowledged the research was predominantly in acquiring problem-solving skills and emphasized the need of future research to examine instruction in other content areas.

**Instruction and Scientific Reasoning.** Science educators have been studying the development of scientific reasoning for many decades. Piaget’s theory of intellectual development is often the theoretical basis for studies involving scientific reasoning (e.g. Kuhn & Phelps, 1982; Kuhn, Schauble, & Garcia-Mila, 1992; Schauble, 1990). According to Inhelder and Piaget (1958; 1970), cognitive development and scientific reasoning occur in a certain sequence as children explore, manipulate and make sense of their environment by actively constructing new and more elaborate mental structures as needed. In order for learning to take place the learner must be involved in active discovery and mature enough to receive the knowledge. Piaget emphasized that children cannot be instructed to master tasks in higher stages and Piaget emphasized that learning induced by external forces produces temporary or little change in children’s ability to think logically (Brainerd, 1978; Piaget, 1970).

According to Piaget (1970) there are four stages of cognitive development: the sensorimotor stage (age 0-2); the preoperational stage (2-7); the concrete operational stage (7-11); and, the formal operational stage (11+). Children in the sensorimotor stage are dependent on their senses and motor ability to interact with their environment to engage in learning, while those in the preoperational stage are able to engage in
rudimentary thinking and reasoning use symbols such as language and numbers (Piaget, 1970). Young children in these stages tend to be illogical in their thinking and unable to use evidence to support their conclusions; an important aspect of scientific reasoning (Kuhn & Phelps, 1982). A child in the preoperational stage will not be able to demonstrate logical thinking even with the support of concrete objects.

Concrete operational children are expected to be more advanced in their reasoning when supported with concrete objects (Piaget, 1970). For example, children in this stage often struggle with conservation tasks. These tasks involve changing dimensions of an object involving liquid, number, weight, volume and length. A classical example of concrete operational thought involves a child seeing two rows of five apples with the apples an inch apart in one row and four inches apart in the second row. A child who has mastered conservation of numbers will say the rows have equal amounts of apples even though one row is longer while a child who has not mastered this task will conclude the longer row has more apples. Once a child has mastered all reasoning skills associated with the concrete operational stage they are said to be ready for formal operational thought.

The formal operational stage is focused on gaining the strategies needed to reason in hypothetical situations with multiple dimensions (Inhelder & Piaget, 1958). A classical formal operational task involves asking a child to experiment with a pendulum and discover how variables could affect the speed of the swinging object. Common variables that can be manipulated include the length of the string, height of the release point, and force of the release. Children in the formal operational stage can think abstractly and systematically solve problems with multiple factors. Children who have mastered formal
operation tasks are said to be capable of being scientifically literate (Pressely & McCormick, 1995).

Based on these theoretical notions, researchers began testing Piaget’s hypothesis that scientific reasoning could not be taught to concrete operational children. These studies examined whether children who are in an earlier stage of development can master skills at the next stage with instructional support on tasks associated with higher-level scientific reasoning. The results of these studies suggested that with support younger children can master skills typical of older children. (Kuhn, 1988, 1995; Kuhn & Phelps, 1982; Kuhn, Schauble, & Garcia-Mila, 1992; Schauble, 1990, 1996). The following studies are examples of studies investigating the role of instruction in teaching scientific reasoning with preoperational students learning to conserve or teaching concrete operational students formal operational tasks involving inductive and deductive reasoning.

In one of the first studies examining young children who were considered novices in a context that required more complex thinking than just following a rule, Gelman (1969) demonstrated the impact of guidance. Gelman studied 110 five-year olds over a three-month period in one of the first studies teaching preoperational children to conserve with objects involving liquids, height, weight, and mass. One of the conservation tasks involved in this study was the ability to see two amounts of liquid in differently shaped containers as equal. Teachers directed student attention to the appropriate aspects of the task (equal starting amounts and comparing how tall and wide the containers were) and offered feedback in the form of explanation when the students made errors (guided instruction). Gelman’s results suggested that with guided instruction and practice children
who originally failed Piaget’s tests of conservation could be taught conservation of length, mass, numbers and liquid volume. His study demonstrated scientific reasoning skills could be taught. In terms of evaluating Inhelder and Piaget’s (1958) claim, this study was confounded as it was unclear if the repeated interaction with the task that led to the mastery of conservation or if instructional support played a role.

Other studies addressed this issue by comparing forms of guided instruction with discovery learning to teach scientific reasoning. In general, these studies have found that guided instruction results in improved scientific reasoning relative to discovery learning on recall and transfer tasks involving conservation (Beilin, 1965; Brainerd, 1972; May & Tisshaw, 1975; Wallach & Sprott, 1964). One of the critical differences in these studies concerns the operationalization of guided instruction. For example, May and Tisshaw (1975) studied first graders learning about conservation of numbers. Guided instruction was operationalized as students receiving explanations for their errors. Students who received guided instruction outperformed those students who received no instruction on number conservation tasks, as well as conservation tasks involving liquids. The analysis indicated medium-sized differences between the conditions.

Wallach and Sprott (1964) and Beilin (1965) operationalized guided instruction as the modeling of reasoning skills. Wallach and Sprott (1964) worked with kindergarten and Beilin (1965) worked with second-grade students, to increase learning and understanding in conservation of length. Teachers working with the guided instruction group would perform conservation tasks involving length. The children were then asked questions that highlighted the reasoning processes involved in conservation problems involving length as students performed similar tasks. Students in the discovery conditions
were encouraged to explore the material without interactions with the instructor. Both sets of researchers tested students on problems similar to those used during instruction. Beilin (1965) also asked students to attempt conservation of mass problems. The results indicated medium-sized differences between the conditions on performing tasks similar to those used in the instruction that favored the guided instruction condition for both studies. Beilin (1965) also found medium sized effects in favor of guided instruction when asking participants to transfer their skills to conservation of mass problems.

Brainerd (1972) operationalized guided instruction as responding to third-grade students with “you are right” or “you are wrong” and found that those who received this feedback performed better than those who did not have feedback on performing tasks involving conservation of length. In his discovery condition, children were encouraged to explore the material as well. Brainerd demonstrated that this minimal amount of guidance was enough to produce a small difference in learning in the outcome measure involving performing conservation of length. Brainerd explained this finding as a result of the feedback allowing students to efficiently identify strategies that lead to the correct answers in the guided instruction, while children in the discovery condition were not being able to identify correct reasoning strategies.

These studies indicate that scientific reasoning, in terms of learning conservation, can be taught and that the amount of guidance offered during instruction facilitates learning. The studies did not address if this is only a phenomenon associated with learning conservation tasks or if other tasks involving scientific reasoning can be improved with guided instruction. Research has been done on instructional strategies being used to assist children expected to be in the concrete operational stages learn tasks
that require formal operations. The following studies investigated concrete operational students engaged in formal operational tasks.

In a study with fourth and fifth graders, Kuhn & Angelev (1976) investigated the role of instruction in the development of formal operational thought. Over the course of 15 weeks students were exposed to classical formal operational problems involving scientific thinking. For example, students were given five odorless and colorless chemicals and were asked to isolate which two chemicals produced the color yellow when mixed together. One condition was only given the problem and equipment to solve (discovery). The other condition was given explicit instruction on how to solve the problem (direct instruction). No statistical differences were observed between conditions; with both conditions producing students who could solve formal operational problems at the end of the study. The authors used their findings to support their hypothesis that formal operational tasks could be taught to students considered to be in the concrete operational phase.

To further investigate why there was a lack of difference between instructional conditions in earlier studies Kuhn and Ho (1980) investigated differences between discovery and direct instruction. Their participants were fourth- and fifth graders who engaged in either self-directed study (discovery) or direct instruction while learning scientific reasoning skills involving the evaluation evidence. The sample consisted of students who demonstrated low formal operational reasoning on a pretest. The experimental condition had students working in pairs on formal operational problems of their choice. The control condition was assigned the same problems and given direct instructions on how to solve them. At the end of three sessions, both groups demonstrated
gains in formal operational reasoning skills, but medium-sized differences in favor of the discovery group were observed on progress towards solving problems involving formal operational tasks and transfer problems.

Studies in the area of acquiring scientific reasoning skills do not solely investigate the influence of instruction on mastery of tasks defined only on the child’s ability to perform the task, but they also study the processes involved in the development of the ability to perform the task. For example, Kuhn, Schauble, and Garcia-Mila (1992) studied the ways children revise their strategies when engaged in self-directed exploration involving scientific reasoning. In two studies involving 50 sixth graders, students were exposed to the same experimental design task over a period of three months while researchers made detailed observations regarding the change students reasoning undergoes when generating new knowledge about designing unconfounded experiments. Over time, all students’ scientific reasoning matured, as did their understanding of the experimental process and their use of data to support and refine their strategy use. This study was important in establishing the importance of time and repeated exposure in discovery forms of learning in order to develop strategies used to design unconfounded experiments.

The reviewed studies establish empirical support for the idea that scientific reasoning can be taught through instruction and that scientific reasoning is not dependent on student age. Furthermore, throughout these studies a consistent finding was that allowing students to discover scientific reasoning principles can lead to deeper understandings (Kuhn, & Angelev, 1976; Kuhn & Ho, 1980; Kuhn, Schauble, & Garcia-Mila, 1992). What is inconsistent in this literature is how guided instruction is
operationalized. However, when described as guided instruction the instructional intervention tends to produce improved scientific learning skills (Beilin, 1965; Brainerd, 1972; May & Tisshaw, 1975; Wallach & Sprott, 1964). Studies involving instruction and the development of scientific reasoning have focused on more than just conservation and deductive and inductive reasoning. Studies have also focused on how students develop the ability to design unconfounded experiments.

**Instruction in Control of Variable Strategy.** For an experiment to be considered unconfounded all relevant variables must be taken into account. Unconfounded experiments isolate relevant variables and systematically change one variable of interest at a time to learn about the relationship between the variable of interest (i.e., independent variable) and dependent variable (Klahr & Nigam, 2004). Confounded experiments are those where two or more variables are being manipulated at a time, which results in the inability to make a clear conclusion concerning the relationship between the independent and dependent variable. For example, if a student is asked to design an experiment for investigating how the ball type affects the distance a ball rolls off a ramp, and they set two ramps up with a golf ball on one ramp and a tennis ball on another ramp but the ramps are of different length, they have failed to control all variables that could affect the distance the ball rolls. In this scenario, the different ball types and the different ramp lengths have confounded the conclusions that can be drawn. This student would have to set the ramps with equal lengths to make a clear conclusion regarding how the type of ball affects the distance the ball rolls off the ramp.

Klahr and associates have developed a framework for classifying the complexity of the scientific inquiry process that has been utilized to aid in studying strategies used to
design unconfounded experiments (Klahr, 2000; Klahr & Carver, 1995; Klahr & Dunbar, 1989; Zimmerman, 2000). This framework breaks the scientific inquiry process into three phases of inquiry: 1) hypothesis generation; 2) experimental design; and, 3) evidence evaluation. The ability to design unconfounded experiments greatly impacts the student’s ability to evaluate evidence generated from the experiment. The ability to generate unconfounded experiments is dependent on the successful use of control of variables strategies (CVS) and is a key aspect of the “experimental design” phase of developing scientific reasoning. CVS are the strategies used when designing experiments that involve: 1) identifying relevant variables; 2) controlling all identified variables while changing one factor at a time; and, 3) thinking about different combinations of possible relationships (Klahr & Li, 2005). Children who are adept at creating unconfounded experiments, identifying and correcting confounded experiments and making inferences from experimental outcomes are considered to have mastered CVS. This skill is also part of the national science standards (NRC, 1996) and part of the national benchmarks for grades 2nd through eighth grade (AAAS, 1993). Furthermore, the ability to design unconfounded experiments and use control of variable strategies has been identified as a key characteristic of a student who has formal operational abilities (Inhelder & Piaget, 1958). The following studies investigate how children develop this skill and if instruction can aide in the process.

In response to Piaget’s (1970) theories involving intellectual development of scientific reason, Case (1974) demonstrated that with guided instruction seven- and eight-year-olds could be taught CVS even though they had not mastered Piaget’s conservation of weight tasks and were not in the formal operations phase. In this study with 52
children, two conditions were contrasted: discovery, where children received no instruction but explored the laboratory equipment by designing experiments; and, guided instruction, where students were given examples and non-examples of unconfounded experiments. Using questioning to guide students to use CVS, facilitators allowed students to practice using CVS strategies to explain why examples were or were not confounded. Stronger performances in designing unconfounded experiments were found with students who received guided instruction with a medium-sized difference between conditions. This study provided evidence that students could master CVS with guidance even though they were younger than the expected age described by Piaget.

To study the processes of developing CVS, Kuhn and Phelps (1982) worked with 22 fourth- and fifth graders over the course of 13 sessions. In each session, using discovery methods, students were shown four clear chemicals being mixed together and the mixture turning red. They were then asked to figure out which chemical turned the mixture red. During the sessions, the students were asked to generate hypotheses, design experiments and use their results to make inferences concerning the responsible chemical. The students were interviewed throughout the session about their thought processes in regards to why they were doing what they were doing. The researchers concluded that the development of using CVS strategies was a slow process involving the extinguishing of old strategies with the replacement of effective new strategies, not just the learning of new strategies. The development of CVS was also not a linear process. Success in one session in drawing clear conclusions due to their use of CVS did not mean the strategies would be used in the next session.
Once it was established that younger children could be taught CVS Chen and Klahr (1999) examined the best ways to teach the use of CVS. With 87 seven- to ten-year-old students, they asked students to design and evaluate studies using CVS. Students were divided into the three following instructional conditions to learn to design and evaluate studies: 1) no instruction and probing questions (guided instruction); 2) no instruction or probing questions (discovery); and 3) instruction with no probing questions (direct instruction). Medium sized differences were observed in favor of direct instruction improving student ability to design unconfounded experiments and evaluate other experiments using CVS over the other two conditions. Furthermore, the guided instruction condition did not significantly improve students’ abilities to design or evaluate experiments relative to discovery learning. The authors explained this finding by suggesting the direct instruction emphasized key aspects of using CVS while guided instruction and discovery did not focus learner attention on the relevant skills needed.

In response to the growing number of researchers who theorized that discovery methods of instruction produced deeper and more meaningful learning, Klahr and Nigam (2004) examined fourth- and fifth-grade children’s abilities to use CVS to design experiments after direct instruction or a period of discovery. Children receiving direct instruction were in an instructor-centered environment where the pace, explanation, and examples were controlled by the instructor. Instruction included definitions and relationships between independent, dependent, and control variables. The instructor then designed two experiments, one confounded and the other unconfounded, and asked participants if ‘they can tell’ if the variable of interest affects an outcome. Asking participants a question was a way to cognitively engage participants to ensure equal
amounts of activity in both conditions (Klahr & Nigam, 2004). The instructor then explained what made each experiment confounded or not. In the pure discovery treatment, students controlled the learning environment without feedback or intervention from an instructor. The instructor only provided the learning goal if designing unconfounded experiments. The researchers found that children who were taught using direct instruction reached mastery of the content in greater numbers than students taught using discovery methods. In addition, students receiving direct instruction performed comparably on measures demonstrating deep understanding and transfer of the knowledge as those in the discovery groups. Large condition-related differences were found in ability to apply CVS concepts in favor of direct instruction relative to discovery on a measure of ability to design experiments.

Kuhn and Dean (2005) argued that discovery methods of instruction take longer than a single session and result in deeper understanding of CVS in response to the above study. Investigating the issue of time, the researchers studied 30 sixth-grade students with a study that had three conditions: 1) discovery, where children received no instruction on the designing of experiments through a software program on earthquakes but had practice with program; 2) guided instruction, where children received the suggestion that they focus on the effect of one variable during each session involving practicing with the program; 3) direct instruction, where children completed the initial session and outcome assessments. Students all worked in homogenous pairs in terms of gender and age. Students in the first two conditions completed 12 thirty-minute sessions. Students in all three conditions completed outcomes that required using CVS to evaluate if an experiment was confounded or not. All three groups demonstrated improved ability to
use CVS. Participants in the discovery group took more sessions to display CVS than the other two groups, but were able to transfer the skill to novel situations better at the end of the experimental phase with small condition-related differences. Kuhn and Dean argued that students who are allowed to learn through discovery methods developed CVS in a more meaningful way and are more likely to use the strategy in novel situations.

In response to the applicability of previous research to classroom settings Klahr and Li (2005) designed a study to add external validity to their claims for direct instruction being the best way to teach CVS. With intact middle school science classes two instructional techniques were compared. The experimental group received four class periods of direct instruction on CVS. The comparison group did not receive instruction, but was exposed to the CVS content through the normal science curriculum. Outcome measures included: worksheets where students were asked how they would design an experiment; the design of actual experiments working in dyads; and, achievement on standardized tests of science knowledge. In both groups, students gained mastery of CVS, but students made greater gains of mastery in the direct instruction condition.

To counter Klahr and Nigram (2004), Dean and Kuhn (2007) investigated the mastery of CVS with three groups of 15 fourth-grade students. One group participated in twelve 30-minute sessions over 10 weeks on practice problems without instruction (discovery). A second group received the addition of direct instruction on the same problems at the end of the initial session (direct instruction+discovery). The third group received direct instruction, but no practice sessions (direct instruction). Significant differences were found between knowledge and ability to use CVS after the initial session in favor of those students who received direct instruction. Assessments were then
given at the end of the final session, one week after the final session and seven weeks after the final session. Results demonstrated that students who received the direct instruction plus discovery outperformed the other two conditions on knowledge and transfer activities at the end of the sessions. At the seven-week assessment, the discovery only group outperformed the other two groups on knowledge and transfer activities. The authors concluded that the results suggest discovery learning leads to deeper and more meaningful knowledge acquisition.

This body of literature supports the notion that scientific reasoning skills can be taught and demonstrates that instructional strategy employed can impact learning. In this literature, using direct instruction and discovery learning as the only comparisons does not produce consistent evidence supporting the use of one instructional strategy (Dean & Kuhn, 2007; Klahr & Nigam, 2004; Klahr & Li, 2005). Some studies indicated that instruction that allows discovery of key principles may lead to learning that is more readily transferred to new situations and is more durable over time (Dean & Kuhn, 2007; Kuhn & Dean, 2005). While studies by Klahr and associates have indicated that direct instruction leads to more robust learning in a single session and that the resultant skill mastery is transferable to other outcomes (Klahr & Nigam, 2004; Klahr & Li, 2005). The difference could be associated with the length of treatment. Klahr and associates studied the effects of instructional strategies on learning in a single one-hour session. Dean and Kuhn deliver their treatments over multiple sessions.

In terms of guided instruction, evidence is inconclusive for three reasons. First, operationalization of guided instruction is very different across studies and produced conflicting evidence of the value of the instruction for learning and transfer (Case, 1974;
Chen & Klahr, 1999; Kuhn & Dean, 2005). Second, there were different comparison treatments within studies. Some studies compared guided instruction to discovery learning (Case, 1974) and other studies compared direct instruction to discovery learning (Chen & Klahr, 1999; Kuhn & Dean, 2005). Finally, the length of the treatment varies by study with Dean and Kuhn delivering their treatments over multiple sessions and Case (1974) and Chen and Klahr (1999) employ a single one-hour session.

While there are many studies exploring learning outcomes associated with different instructional strategies none could be found that investigate the role of student motivation. This is especially surprising, considering the benefits of discovery learning are expected to be due to students’ natural curiosities promoting engagement. The following section examines the role of motivation, specifically science self-efficacy, in learning science topics.

**Self-Efficacy in Science Education**

Self-efficacy is the judgment an individual makes regarding their ability to perform various tasks (Bandura, 1977, 1982). This judgment is domain and task specific. For example, an individual with high self-efficacy on the football field might not have the same high self-efficacy on the basketball court. Or, a student with high self-efficacy in reading fiction may not have the same self-efficacy when asked to read expository text. This section discusses the following: 1) Bandura’s theory of self-efficacy; 2) the established positive relationship between science self-efficacy and achieving science literacy in college, high school and middle school; and 3) factors that impact science self-efficacy, specifically prior success in high school and middle school.
According to Bandura (1982), individual perceptions of self-efficacy in a domain influence behavior in that domain. These perceptions influence which tasks a learner engages in so that an individual with low self-efficacy might avoid a task while one with high self-efficacy will engage in the task. Individuals with high self-efficacy are also expected to set challenging goals for themselves, initially engage with the material in a more meaningful manner, persistent in the face of challenge or negative feedback and use a greater number of metacognitive and cognitive strategies when learning (Bandura, 1993; Gungoren & Sungur, 2009; Hoy, 2004; Linnenbrink & Pintrich, 2003; Pajares & Schunk, 2001).

**Science Self-Efficacy and Achievement.** Motivation is a broad term that encompasses many components. In terms of science, a motivated science student has an internal drive that arouses, directs and sustains learning behavior (Bryan, Glynn & Kittleson, 2010). With data collected through observing students in three science teachers’ classrooms over the course of two years, Sanfeliz and Stalzer (2003) concluded that science motivation for students is mainly driven by three components: 1) intrinsic motivation; 2) self-determination; and 3) self-efficacy. These three components do not act independently of each other. For example, highly motivated students in science will most likely enjoy learning science partially due to their belief that they have the ability to learn science. They will also report a higher sense of control over their learning.

Self-efficacy in science is theoretically expected to be a powerful predictor of future success in becoming scientifically literate. McMillan & Forsyth (1991), using their experiences teaching college level biology, theorized that science self-efficacy might be the most important key to success in the science classroom. They predict that students
with high science self-efficacy are more willing to take and persevere through science courses that are difficult because they are more willing to engage in the complex tasks and thinking that are needed for success. They advocate for science experiences in the K-12 setting that allow for building student’s science self-efficacy, but give no specific recommendations for impacting science self-efficacy in the science classroom.

Empirical studies support science self-efficacy as being positively associated with achieving science literacy and science-related choices throughout schooling and future careers. For example, the science self-efficacy of college students has been associated with how long students study science in their educations and decisions to pursue careers in the field of science (Andrew, 1998; Hackett & Betz, 1981; Gwilliam & Betz, 2001; Lent, Brown, & Larkin, 1984; Luzzo, Hasper, Albert, Bibby, & Martinelli, 1999). In a study involving 205 undergraduates from an eastern U.S. university, Hackett and Betz (1981) asked students about their perceived capabilities to successfully complete the degree requirements and pursue careers in twenty occupations in the sciences and math areas. They found that those with the highest self-efficacy in a domain went on to finish degrees and pursue careers in those areas. Using Cohen’s (1988) guidelines for interpreting Pearson’s $r$ a small relationship is an $r$ less than .10; medium is .30; and large is anything over .50. These researchers found a small positive relationship between science self-efficacy and future choices related to science. They also used their results to argue for science and math experiences in the K-12 setting that allow the building of science self-efficacy.

The positive relationship between science self-efficacy and student science literacy achievement has also been established with high school and middle school
populations. In high school, science self-efficacy has been positively correlated with science literacy achievement and involvement with science activities inside and outside of school (Bryan, Glynn & Kittleson, 2010; Kupermintz, 2002; Lau & Roeser, 2002). Kupermintz (2002) surveyed 491 high school science students with a battery of motivational measures in an attempt to understand the effect of classroom factors (e.g., resources in the school and use of technology), individual motivation (e.g., self-efficacy, and self-determination), and personal characteristics (parental interest and mental ability) had on achievement in science. The strongest motivational predictor of science literacy achievement was student science self-efficacy with a small relationship being observed.

Lau and Roeser (2002) sampled 491 10th and 11th grade high school students to examine how motivational and cognitive factors predict student achievement in science. They defined achievement by standardized test scores and their intention to pursue future science studies. The findings indicated that students’ cognitive abilities were medium-sized positive predictors of performance on achievement tests. However, science self-efficacy was the strongest predictor of student commitment to continue studying science and pursuing science-related careers with a small relationship.

Bryan, Glynn and Kittleson (2010) examined the intrinsic motivation, self-efficacy, and self-determination of 910 high school students who were ages 14 through 16 and enrolled in an introductory science course. A subset of students from the original sample responded to surveys, wrote essays concerning their motivations, and participated in interviews. The researchers found that all three factors of motivation were significant predictors of achievement. Relative to the two other predictors, self-efficacy had the strongest predictive value.
Science self-efficacy has also been associated with science literacy achievement at the middle school level (Britner & Pajares, 2001; Pajares, Britner, & Valiante, 2000). Britner and Pajares (2001) surveyed 262 middle school students enrolled in science courses to examine the relationships between factors of motivation, gender and race. They also examined the relationship between science self-efficacy and achievement in science class. They found science self-efficacy to have a large positive relationship with self-efficacy and science achievement for female students only. Pajares, Britner, and Valiante (2000) investigated the relationship between achievement goals, motivational constructs and gender in the area of science achievement with 281 middle school students enrolled in science courses. They found small positive relationships between achievement goals, self-efficacy and achievement in science.

Sources of Science Self-Efficacy. It is generally agreed that self-efficacy is important to the achievement of science literacy. How to increase a student’s self-efficacy is less understood. According to theory, self-efficacy originates from four sources: 1) previous successes within the domain; 2) observation of success in a peer with perceived similar capabilities; 3) verbal persuasion from an authority; and 4) and their physiological response to the task (Bandura, 1982). Researchers have reported significant correlations ranging from .20 to .78 between theses four sources of self-efficacy and overall self-efficacy (Britner, 2008; Britner & Pajares, 2006; Hampton, 1998; Gungoren & Sungur, 2009; Karaaslan & Sungur, 2011; Klassen, 2004; Usher & Pajares, 2006). In general, there is a positive relationship between prior successes and science self-efficacy.

While most of the prior research examining sources of self-efficacy involve math, a few researchers have investigated the sources of science self-efficacy. Britner and Pajares
(2006) surveyed 319 middle school science students with the aim of isolating the relationship of the four sources of self-efficacy with science self-efficacy and achievement in science. Their results indicate medium-sized relationships between science self-efficacy and science literacy achievement. Of the four sources of self-efficacy, only previous mastery experiences were significant predictors of science self-efficacy with small relationships observed. Observing successful peers, persuasion, and their physiological response were not significant predictors of science self-efficacy.

Recently, Britner (2008) studied 502 high school science students for the degree that each of the four sources of self-efficacy correlated with self-efficacy in science. Her results indicate gender differences. For male students direct past experience was the only significant predictor of current self-efficacy in science with a medium-sized relationship. For female students, past experience, vicarious experiences, social persuasion and physiological state were all small significant predictors of science self-efficacy. Persuasion from an authority figure was the strongest positive predictor for females. Science self-efficacy was also a significant predictor of achievement within science courses.

Previous success with science tasks has been demonstrated to correlate with student self-efficacy in science when studied in isolation (Gungoren & Sungur, 2009; Karaaslan & Sungur, 2011). In a study with 900 sixth-, seventh-, and eighth grade students from Turkey, Gungoren and Sungur (2009) found that grade level had a significant relationship with student self-efficacy in regards to science with large effects. Using the Motivated Strategies for Learning Questionnaire (Pintrich, Smith, Garcia, and McKeachie, 1991), these researchers found that sixth grade students had the highest
levels of science self-efficacy and that eighth graders had the lowest levels of science self-efficacy. They also found that science self-efficacy was positively correlated with grades in science across all grade levels. In interviews with a subset of their sample children reported basing their judgment of their capabilities on previous experiences with science. These findings were further supported by Karaaslan & Sungur (2011) using similar procedures and instruments with a population of 518 children in a rural area of Turkey.

Research supports self-efficacy as a positive predictor of achievement in science and that prior mastery experiences impacts student science self-efficacy. Missing from the literature are studies comparing instructional practices within the classroom and instructional impact on the self-efficacy of students. Bryan, Glynn and Kittleson (2010) found that students felt collaborative activities enhanced their motivation, but this study was based on self-reports and was correlational. Also, this study did not investigate changes in motivation after collaborative activities. Lacking from the literature are examinations of the effects of instructional practices on science self-efficacy. Results from previous research suggest that instruction which increases the success of a student should enhance their self-efficacy positively while less effective instructional strategies should result in relatively lower student self-efficacy.

The Present Study

The present study addresses the following three limitations identified in the literature. First, the delivery of instruction literature illustrates conflicting findings between the application of discovery, guided instruction and direct instruction. Considerable debate exists in the operationalization of guided instruction and
instructional strategy instructions differ across studies substantially. When guided instruction is present, favorable learning occurs relative to direct and discovery learning strategies. However, contemporary studies in teaching the use of CVS to design unconfounded experiments focus on direct versus discovery only (Dean & Kuhn, 2007; Klahr & Nigam, 2004; Klahr & Li, 2005; Kuhn & Dean, 2005). Second, in terms of science self-efficacy most of the research involves older students. Science instruction begins in elementary school, but no research was found that addressed elementary-age students’ science self-efficacy. Also, no research was found that directly examined changes in science self-efficacy that can be attributed to classroom experiences or instruction in science.

The literature on the level of instructional guidance and science self-efficacy suggests two things. First, guided instruction may be a more efficacious means of teaching CVS, but has not been included as condition in many of the contemporary studies. Second, it suggests that self-efficacy is an important factor in student achievement in science and that prior success impacts an individual’s self-efficacy. Together, these findings suggest that guided instruction that leads to greater student success on learning outcomes should positively impact their self-efficacy in science.

From these observations found in the literature, the following hypotheses were made:

H1. There will be instructional strategy-related differences in the total number of experiments designed and percentage of unconfounded experiments designed, with:

a. Children who receive guided instruction designing a greater number of experiments than those who receive direct instruction.
b. Children who receive guided instruction designing a higher percentage of unconfounded experiments than those who receive direct instruction.

c. Children who receive direct instruction designing a greater number of experiments than those receiving discovery learning.

d. Children who receive direct instruction designing a higher percentage of unconfounded experiments than those receiving discovery learning.

H2. There will be instructional strategy-related differences on learning outcomes, with:

a. Children who receive guided instruction having higher scores on measures of recall, application and evaluation of control of variable strategies than those who receive direct instruction.

b. Children who receive direct instruction having higher scores on measures of recall, application and evaluation of control of variable strategies than those receiving discovery learning.

H3. There will be instructional strategy-related differences in change in science self-efficacy, with:

a. Children who receive guided instruction will have greater changes in science self-efficacy than those who receive direct instruction.

b. Children who receive direct instruction will have greater changes in science self-efficacy than those receiving discovery learning.
CHAPTER 3

METHODS

Study Overview

In order to address the research hypotheses, two studies were conducted with nine- and ten-year-old participants assigned to one of three instructional conditions. The first study was a pilot study designed to develop the materials and the second study was the primary study employed to test the research hypotheses. In both studies, the instructional conditions were as follows: 1) guided instruction, where children received instruction with examples and explanations; 2) direct instruction, where children received instruction through lecture and examples; and 3) discovery learning, where children received instruction through methods of discovery. Each condition received instruction on how to design experiments using control of variable strategies (CVS) and after instruction designed their own experiments. Before instruction, prior knowledge of the scientific method and science self-efficacy were measured. After instruction, participants were measured on their knowledge of CVS with the following measures: 1) number of unconfounded experiments designed; 2) recall of definitions associated with designing unconfounded experiments; 3) application of CVS; 4) evaluation of experiments using CVS; and 5) change in their self-efficacy in regards to science tasks. The resultant data were analyzed for condition-related differences. The following sections describe the completed pilot and primary studies, respectively. The University of New Mexico’s Human Research Protections Office approved all research procedures.

Pilot Study
The pilot study was designed to develop, test and refine the materials for the main dissertation study. Participants consisted of six nine- and ten-year olds enrolled in one of two summer camps offered through the University of New Mexico. These camps were the Summer Youth Sports Program (SYSP) and the Youth Sport and Fitness Program (YSFP). SYSP is funded by the University of New Mexico’s College of Education and participants must qualify for participation by demonstrating proof of low household incomes. YSFP is funded through the University of New Mexico’s Continuing Education and participants must pay a fee to participate in one of two three-week sessions. Four of the pilot study participants were from the SYSP camp and two were from YSFP.

Initial contact with all of the prospective participants occurred at the same time within each camp and included discussing the benefits and risks associated with participation in the study and handing out the consent form to be given to their parents. Consent forms were collected during drop-off times for each camp the next day. Child assent was collected during the first session. The sample of six included three nine-year-olds, three females, five Hispanics, five fourth graders, and all students were most comfortable speaking English.

The pilot study was designed to determine the developmental appropriateness of the teaching materials and outcome measures as well as the session lengths. Scripts for the session one for the pilot can be found in Appendix A. Scripts for the treatments in both the pilot and main study can be found in Appendix B for direct instruction, Appendix C for guided instruction, and Appendix D for the discovery treatment. Four students completed the pilot study, adjustments to scripts and materials were made, and the final two students completed the study with the modified materials and scripts. This
occurred during the third week of both the camps over the course of two days. Revisions to the materials and scripts included instructions for the use of the equipment, time given to design their own experiments, and the number of items in the outcome measures. The initial group of participants struggled with using the ramp, so changes were made in order to encourage students to use the ramp during the instructional phase instead of just asking if they had any questions at the end. See Appendix E for the original instructions. The time to design experiments was also increased because the initial students designed an average of three experiments in the 10 minutes with a range of 0-4 experiments designed. By adding five minutes the second group was able to design six experiments. See Appendix F for the original wording. Finally, the outcome measures were modified to shorten the length of the session. Sessions for the initial group averaged 48 minutes and for the second group 42 minutes. Changes to the material included dropping the second science fair poster evaluation and two open-ended questions within the evaluation measure. The open-ended questions were dropped because five of the participants left the items blank during the study. See Appendix G for the original measure.

**Primary Study**

**Sample Size.** Sample sizes were calculated using a type I error rate of .05 and power of .80. Due to effect sizes reported in the literature ranging from .21 to 1.72 (Klahr & Nigam, 2004), potential sample sizes were calculated using $d$s of small (0.2), medium (0.5) and large (0.8) (Cohen, 1988). With these effect sizes, a type I error rate of .05 and one-tailed test each condition would need 50, 30 and 20 participants to have power of .80. The sample for this study was selected based on the assumption of a large effect size on all outcomes of interest. This resulted in a study of 60 participants with 20
participants in each treatment level (guided instruction vs. direct instruction vs. pure
discovery).

Participants. Participants were 60 nine- and ten-year-olds enrolled in the SYSP
and YSFP summer camps. Initial contact procedures are described in the pilot study
section. In addition, further recruitment procedures were added that included contact with
parents through direct conversation by the primary investigator as they dropped off or
picked up their child and through distributions of consent materials to children during
camp time. Consent forms signed by the parents were collected at the beginning of each
camp day. Of the 60 participants, 19 (32%) were from the SYSP camp and 41 (68%)
were from the YSFP camp.

The total sample included 39 (65%) males and 33 (55%) 10-year-olds with a
mean age of 9.55 (SD = .50). Thirty-eight (63%) participants reported starting fourth
grade, 20 (33%) reported starting fifth grade, and 2 (3%) reported starting third grade in
the next school year. The sample was composed of 33 (55%) children who reported being
Hispanic, 17 (28%) Caucasian, 5 (8%) American Indian, 2 (3%) Asian American, 1 (2%)
African American, and 2 (3%) who did not report. Forty-six (77%) participants reported
being most comfortable speaking English, while 12 (20%) were most comfortable
speaking Spanish and 2 (3%) were most comfortable speaking another language. All
participants were able to read and respond to experimental questions in English. See
Table 1 for demographic details broken down by condition.
<table>
<thead>
<tr>
<th></th>
<th>Direct (n = 20)</th>
<th>Guided (n = 20)</th>
<th>Discovery (n = 20)</th>
<th>Total (n = 60)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>14.0 (70.0)</td>
<td>13.0 (65.0)</td>
<td>12.0 (60.0)</td>
<td>39.0 (65.0)</td>
</tr>
<tr>
<td>Female</td>
<td>6.0 (30.0)</td>
<td>7.0 (35.0)</td>
<td>8.0 (40.0)</td>
<td>21.0 (35.0)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8.0 (40.0)</td>
<td>9.0 (45.0)</td>
<td>10.0 (50.0)</td>
<td>27.0 (45.0)</td>
</tr>
<tr>
<td>10</td>
<td>12.0 (60.0)</td>
<td>11.0 (55.0)</td>
<td>10.0 (50.0)</td>
<td>33.0 (55.0)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>8.0 (40.0)</td>
<td>13.0 (65.0)</td>
<td>12.0 (60.0)</td>
<td>33.0 (55.0)</td>
</tr>
<tr>
<td>White</td>
<td>8.0 (40.0)</td>
<td>3.0 (15.0)</td>
<td>6.0 (30.0)</td>
<td>17.0 (28.3)</td>
</tr>
<tr>
<td>Native</td>
<td>2.0 (10.0)</td>
<td>2.0 (10.0)</td>
<td>1.0 (5.0)</td>
<td>5.0 (8.3)</td>
</tr>
<tr>
<td>American</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African</td>
<td>0.0 (0.0)</td>
<td>1.0 (5.0)</td>
<td>0.0 (0.0)</td>
<td>1.0 (1.7)</td>
</tr>
<tr>
<td>Native</td>
<td>1.0 (10.0)</td>
<td>0.0 (0.0)</td>
<td>1.0 (5.0)</td>
<td>2.0 (3.3)</td>
</tr>
<tr>
<td>Other</td>
<td>1.0 (10.0)</td>
<td>1.0 (5.0)</td>
<td>0.0 (0.0)</td>
<td>2.0 (3.3)</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>0.0 (0.0)</td>
<td>1.0 (5.0)</td>
<td>1.0 (5.0)</td>
<td>2.0 (3.3)</td>
</tr>
<tr>
<td>4th</td>
<td>12.0 (60.0)</td>
<td>15.0 (75.0)</td>
<td>11.0 (55.0)</td>
<td>38.0 (63.3)</td>
</tr>
<tr>
<td>5th</td>
<td>8.0 (40.0)</td>
<td>4.0 (20.0)</td>
<td>8.0 (40.0)</td>
<td>20.0 (33.3)</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>16.0 (80.0)</td>
<td>15.0 (75.0)</td>
<td>15.0 (75.0)</td>
<td>46.0 (76.7)</td>
</tr>
<tr>
<td>Spanish</td>
<td>2.0 (10.0)</td>
<td>5.0 (25.0)</td>
<td>5.0 (25.0)</td>
<td>12.0 (20.0)</td>
</tr>
<tr>
<td>Other</td>
<td>2.0 (10.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>2.0 (3.3)</td>
</tr>
</tbody>
</table>

Note: Number of participants in each category reported; percentage in parenthesis
**Design.** Participants were randomly assigned in randomized blocks of three, generated by an online randomizer, to one of the three following conditions: 1) direct instruction, where children received instruction through lecture and explanation of examples; 2) guided instruction, where children received instruction where they explained the examples with prompts; and 3) discovery instruction, where children received instruction through methods of self discovery. All participants completed the instruction phases and outcome activities as individuals.

**Instruments.** All instruments were modified from pre-existing measures to be more appropriate for the content or age group. Specifically, reading difficulties of the materials were modified so that they were at the 5.9 grade level according to the Flesch Kincaid Grade Level reading test. Item analyses were performed for each measure. For the present description Cronbach’s alphas are reported.

**Demographic questionnaire.** A five-item survey was used to collect demographic information. Participants were asked to self-report the year they were born, their gender, the grade they would be starting in August, their ethnicity/race, and the language they were most comfortable speaking. The questionnaire can be found in Appendix H.

**Prior knowledge pretest.** The pretest consisted of 10 multiple-choice items regarding concepts related to scientific reasoning. Test items were experimenter created with the guidance of items from the Standard Based Assessment in Science (SBAS) for grades three through seven (NM Science Standards, 2003). The items were modified to reduce linguistic complexity. The number of items answered correct were added together to produce a pretest score ($\alpha = .21$). Results of the item analyses revealed that the reliability of the scores was problematic for this measure, because several items
correlated negatively with other items. Specifically, items one and eight, if dropped, would raise the reliability the most. The content in both of these questions involve scientific observations, and in the case of item eight is the only question on this measure from the seventh grade SBAS. For questions see Appendix I and for complete results of the item analysis see Appendix J.

**Science self-efficacy questionnaire.** This measure contained 10 items addressing participant’s self-efficacy in science. This measure was modified using Bandura’s (2006) instructions on constructing self-efficacy items and was adapted from a math inventory for young students. The individual scores were averaged together to produce a single science self-efficacy score for each individual each time the inventory was taken. The inventory was taken before participation in ($\alpha = .90$) and after completion of the study ($\alpha = .92$). For questions see Appendix K, for complete results of the item analysis see Appendix L.

**Designed experiments.** The experimental design activity asked participants to design as many experiments as possible in a 15-minute period using CVS. The activity occurred directly after instruction. During the activity, participants designed an experiment and told the facilitator when it was completed. The facilitator acknowledged the completion, and recorded if the experiment was correct as per CVS standards while the participant worked to design another experiment. The total number of experiments designed correctly as well as the number of incorrectly designed experiments were recorded. Participants used the experimental ramp pictured in Figure 1 to design their experiments regarding factors affecting the distance the ball rolls. For the exact wording of the instructions see Appendix M.
Figure 1. Image of experimental ramp

Note: The ramps used in this study are similar to the experimental ramps used in Klahr and Nigam (2004).

**Cued recall of scientific vocabulary.** The multiple-choice recall test consisted of five items regarding types of variables in a scientific experiment and designing experiments using CVS. As with the pretest, the items were developed using the Standard Based Assessment in Science for grades three through five as a framework. Correct answers were given a point, and all items were added together for a final recall score ($\alpha = .57$). For questions see Appendix N, for complete results of the item analysis see Appendix O.

**Application of control of variable strategies.** The multiple-choice application test consisted of 10 items regarding designing experiments using CVS. For four questions, the instructor designed two experiments with the ramp and participants answered two questions regarding each experiment. Questions asked participants to identify variables and if the experiment was good using CVS. For the remaining six questions, experiments
were described within the question and participants answered regarding the variables and experimental design. Items answered correctly were totaled for a final application score ($\alpha = .62$). For questions see Appendix P, for complete results of the item analysis see Appendix Q.

**Evaluation of science fair posters.** The multiple-choice evaluation test consisted of four items evaluating a study presented in the form of a 5th grade science fair poster. Questions asked participants to evaluate the poster in terms of controlling for variables important to the experiment. Correct responses were added together for a final evaluation score ($\alpha = .21$). For questions see Appendix R, for complete results of the item analysis see Appendix S.

**Procedures.** Participation in this experiment happened in two sessions. Each session occurred in a classroom with 20-30 student desks. The first session included obtaining the assent of the children and completion of the pretest and science self-efficacy questionnaire. The first session was completed in groups of five to twelve students and lasted no longer than 20 minutes. The facilitator was a male friend of the researcher. Students completed the first session on the day parental consent was received. The second session occurred one to three days after the completion of the first session. Before the second session, participants were randomly assigned to one of three conditions with 20 random blocks of three. Randomizations were created by an online random number generator.

Participation in the second session included an introduction to the facilitator and experimental ramp, the instructional treatment, the design activity and completion of the outcome measures. To begin the second session, the facilitator introduced herself, asked
the participant his or her name, and to develop rapport asked the participant about his or her experiences at the camp. See Appendix T for the introduction to the second session.

Next, the facilitator introduced the experimental ramp by asking the participant to describe the ramp. Then the facilitator asked the participant to study the ramp and tell them about something that moves and then move the parts of the ramp. This series of questioning was done until the participant changed all of the variables that could be manipulated. The following variables on the ramp were manipulated: steepness of ramp (flat, shallow, steep), surface of ramp (wood or carpet), length of ramp (short, medium, long), and type of ball (rubber, ping pong, golf).

The treatments for this study involved the instructional method for delivering content concerning control of variable strategies (CVS). The instructional wording was adapted from the terminology and definitions used in Harcourt Science: Grade 4 to describe experiments using CVS (Frank, Jones, Krockover, Lang, McLeod, Valenta, & Van Deman, 2000). This instruction started with an explanation of scientists using experiments to study relationships in nature. Then the concept of ‘variable’ was introduced. The types of variables were introduced focusing on independent, dependent, and control variables and their relationships in using CVS to design unconfounded experiments. Finally, examples of unconfounded experiments and confounded experiments were given with the types of variables being highlighted and how CVS were being used. Instructions unique to each treatment condition are described in the following paragraphs.

**Direct instruction.** The instructions for the direct instruction condition were similar to Klahr & Nigam (2004) where students received a detailed description of the
variables and explanations for how CVS were being used or not used in the examples. The instruction was delivered without soliciting responses from the participant. When the examples were given, participants were asked yes or no questions regarding if the experiment was unconfounded. This question was imbedded into the treatment to cognitively engage the participant in the discussion similar to Klahr and Nigam (2004).

**Guided instruction.** The guided instruction instructions were similar to procedures used by Case (1974). After introduction to scientists’ use of experiments, participants received the same instruction on the types of variables as given in the direct instruction treatment. While the same experiment examples were also used, participants in the guided instruction treatment were asked to evaluate the experiment using CVS for being confounded or unconfounded. They were also asked to label the variables, and to elaborate on how they knew which variables were which. When wrong answers were given, the facilitator encouraged the participant to try again or look at the experiment closer.

**Discovery.** The discovery condition instructions were also similar to those used by Klahr & Nigam (2004). Participants were given an introduction into how scientists used experiments to study relationships in nature, the same as in the direct and guided instruction treatments. They were then asked to design their own experiments. The experiments they were asked to design were the same ones designed by the facilitator in the direct instruction treatment.

After receiving their assigned instructional treatments, all participants engaged in designing their own experiments. Participants then were given the recall, application, and
evaluation measures which were followed by the final science self-efficacy measure. Participants were allowed as much time as needed to complete the outcome measures.

**Analyses.** Analysis of mean differences for the primary study was done in three parts: 1) comparing the number and correctness of experiments designed during the second session; 2) investigating differences in scores on the learning outcomes; and 3) comparing changes in science self-efficacy scores between conditions. The pretest, time, and all outcome measures were analyzed using ANOVA and the changes in science self-efficacy were analyzed using repeated-measure ANOVA.

**Analysis of variance.** One-way ANOVAs were conducted with learning outcomes, the pretest, and time as dependent variable and instructional strategy as a three-level between-subjects factor. Significant omnibus tests were followed with Fisher LSD comparisons. Fisher’s LSD was chosen because when there are three groups this technique has greater statistical power, controls Type I error rates at the desired level, and will identify at least one significant pairwise comparison if the omnibus test is significant (Levin, Serlin & Seaman, 1994). Statistically significant pairwise comparisons were followed with the calculation of Cohen’s $d$ a measure of effect size. Familywise type I error rate was set at .05.

**Repeated measures analysis of variance.** A repeated-measures ANOVA was used to analyze condition-related changes in science self-efficacy. Significant omnibus tests were followed by tetrad comparisons with the primary effect of interest being the interaction between condition and time. Tetrad comparisons allowed for the identification of condition differences in change scores (i.e., change in means from time one to time two). Cohen’s $d$s were calculated for significant tetrad comparisons.
CHAPTER 4

RESULTS

Analysis Overview

In order to address the three hypotheses, the number of experiments designed and the multiple choice learning outcomes were analyzed using analysis of variance (ANOVA). For analysis of the change in self-efficacy scores, a repeated-measures ANOVA was conducted. Prior to each analysis, assumptions associated with ANOVA and repeated-measures ANOVAs were tested. Time spent in the experiment was also analyzed with an ANOVA to ascertain whether time was a plausible explanation for condition-related differences on learning measures.

Preliminary Analyses

Preliminary analysis included: 1) testing statistical assumptions associated with the primary analyses; 2) determining whether condition-related differences in the amount of time individuals spent in the session, during instruction, and completing the outcomes were present; and 3) calculating the zero-order correlations for the continuous variables.

Assumptions. Prior to each analysis, assumptions associated with ANOVA and repeated-measures ANOVA were tested. These tests include testing for normality and homogeneity of variance. All assumptions were met.

Pretest. Scores from the pretest were tested using a one-way ANOVA with condition as the between-subjects factor and the pretest scores as the dependent variable. The main effect of the condition was not statistically significant, \( F(2,57) = 2.91, p = .063 \). Due to the low reliability of the pretest scores and observed negative correlations
with most measures, the pretest was dropped as a covariate from the primary analyses.

The source table for the analysis of the pretest scores can be found in Appendix U.

**Analysis of Time.** Length of time of different phases (study time, instructional time, and outcome completion time) of the experiment and total time of the experiment session was analyzed using a one-way ANOVA with condition as the between-subject factor and type of time as the dependent variable. Descriptive statistics for time can be found in Table 2 and source tables for the analysis of time can be found in Appendix V.

Table 2

*Descriptive statistics for time in the pilot and main study*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time for session 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot Study</td>
<td>18.17</td>
<td>0.88</td>
</tr>
<tr>
<td>Main Study</td>
<td>15.34</td>
<td>1.04</td>
</tr>
<tr>
<td>Time for introduction through instruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot Study</td>
<td>12.09</td>
<td>0.77</td>
</tr>
<tr>
<td>Direct Instruction</td>
<td>14.96</td>
<td>0.17</td>
</tr>
<tr>
<td>Guided Instruction</td>
<td>14.99</td>
<td>0.28</td>
</tr>
<tr>
<td>Discovery Learning</td>
<td>14.87</td>
<td>0.24</td>
</tr>
<tr>
<td>Time for design activity through last outcome measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot Study</td>
<td>24.83</td>
<td>1.087</td>
</tr>
<tr>
<td>Direct Instruction</td>
<td>30.03</td>
<td>0.39</td>
</tr>
<tr>
<td>Guided Instruction</td>
<td>30.07</td>
<td>0.34</td>
</tr>
<tr>
<td>Discovery Learning</td>
<td>29.93</td>
<td>0.57</td>
</tr>
<tr>
<td>Total time for session 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot Study</td>
<td>36.84</td>
<td>0.48</td>
</tr>
<tr>
<td>Direct Instruction</td>
<td>44.99</td>
<td>0.49</td>
</tr>
<tr>
<td>Guided Instruction</td>
<td>45.06</td>
<td>0.51</td>
</tr>
<tr>
<td>Discovery Learning</td>
<td>44.80</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Note: Time is in minutes
Time for the total session. For the total time of experimental session the main effect of instructional condition was not statistically significant, $F(2, 57) = 1.26, p = .290$. Participants spent similar amounts of time in learning sessions regardless of condition.

Time for the introduction through instruction. For the time spent in the introduction and instructional phases the main effect of instructional condition was not statistically significant, $F(2, 57) = 1.45, p = .244$. Regardless of condition, participants spent statistically comparable amounts of time being introduced to the materials and receiving instruction.

Time completing outcomes. For the time spent completing all outcomes including designing experiments, cued recall, application and evaluation and science self-efficacy measure, the main effect of instructional condition was not statistically significant, $F(2, 57) = .51, p = .605$, meaning participants spent similar amounts of time completing the outcome measures.

Correlation of Continuous Variables. Pearson’s product moment correlations were calculated for all continuous variables, the results of which can be found in Table 3. The pre-and post science self-efficacy scores correlate highly while the other continuous variables had low correlations. These low correlations indicate that the different learning outcome variables are measuring theoretically different types of learning. However, higher correlations were expected due to the similar content matter in each outcome.
Table 3

Inter-Item Correlation Matrix For Outcome Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pretest</td>
<td>2.23</td>
<td>1.38</td>
<td>(.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Pre CSAI</td>
<td>5.15</td>
<td>1.81</td>
<td>-.31*</td>
<td>(.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Design Experiments</td>
<td>4.67</td>
<td>1.30</td>
<td>-.18</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Recall</td>
<td>1.92</td>
<td>1.12</td>
<td>.15</td>
<td>-.21</td>
<td>.24</td>
<td>(.57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Application</td>
<td>5.47</td>
<td>2.37</td>
<td>-.05</td>
<td>-.05</td>
<td>.04</td>
<td>.24</td>
<td>(.62)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Evaluation</td>
<td>1.75</td>
<td>0.99</td>
<td>-.21</td>
<td>.04</td>
<td>.16</td>
<td>-.01</td>
<td>.12</td>
<td>(.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Post CSAI</td>
<td>5.53</td>
<td>2.02</td>
<td>-.36*</td>
<td>.87*</td>
<td>.41*</td>
<td>-.01</td>
<td>.01</td>
<td>.08</td>
<td>(.92)</td>
<td></td>
</tr>
<tr>
<td>8. Age</td>
<td>9.55</td>
<td>0.50</td>
<td>.03</td>
<td>-.14</td>
<td>-.01</td>
<td>.23</td>
<td>.05</td>
<td>.11</td>
<td>-.14</td>
<td></td>
</tr>
</tbody>
</table>

Cells below the diagonal contain bivariate correlations. The diagonal contains reliability coefficients (Cronbach’s alpha). Design of experiments does not have a reliability coefficient because it was based on the observations of the experimenter.

* Correlation is significant at $p < .05$.

Primary Analyses

The primary analyses consisted of independently analyzing each dependent variable. This included testing for differences in the number of experiments designed, learning outcomes, and changes in the scores on the science self-efficacy from pretest to posttest. Table 4 contains descriptive statistics for all outcome measures broken out by condition.

Designed Experiments. Analyzing data from the number of experiments the participants designed occurred in two parts. First, an analysis was conducted with the total number of experiments designed during the timed session as a dependent variable. Second, an analysis was conducted with the percentage of correctly designed experiments as a dependent variable. Each analysis was performed using a one-way ANOVA with condition as the between-subjects factor. Statistically significant omnibus tests were followed up with pairwise comparisons using Fisher LSD comparisons and a type I error
rate of .05 (Levin, Serlin & Seaman, 1994). Source tables for the analysis of designed experiments can be found in Appendix W.

Table 4

*Means by condition on learning outcome measures (standard errors in parentheses)*

<table>
<thead>
<tr>
<th></th>
<th>Direct Instruction</th>
<th>Guided Instruction</th>
<th>Discovery Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest</strong></td>
<td>2.60 (0.28)</td>
<td>1.65 (0.21)</td>
<td>2.45 (0.39)</td>
</tr>
<tr>
<td><strong>Experiment Design</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage Correct</td>
<td>50.30&lt;sup&gt;a&lt;/sup&gt; (13.43)</td>
<td>66.70&lt;sup&gt;b&lt;/sup&gt; (21.98)</td>
<td>33.45&lt;sup&gt;c&lt;/sup&gt; (16.43)</td>
</tr>
<tr>
<td>Total Designed</td>
<td>4.80 (1.24)</td>
<td>4.45 (1.23)</td>
<td>4.75 (1.45)</td>
</tr>
<tr>
<td><strong>Learning Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cued Recall</td>
<td>2.45 (0.94)</td>
<td>2.30 (0.92)</td>
<td>1.00&lt;sup&gt;a&lt;/sup&gt; (0.91)</td>
</tr>
<tr>
<td>Application</td>
<td>6.10 (2.42)</td>
<td>5.90 (2.29)</td>
<td>4.40&lt;sup&gt;a&lt;/sup&gt; (2.11)</td>
</tr>
<tr>
<td>Evaluation</td>
<td>2.10 (0.91)</td>
<td>1.95 (0.88)</td>
<td>1.20&lt;sup&gt;a&lt;/sup&gt; (0.95)</td>
</tr>
<tr>
<td><strong>CSAI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>4.58 (1.69)</td>
<td>5.57 (1.95)</td>
<td>5.29 (1.74)</td>
</tr>
<tr>
<td>Post</td>
<td>4.63 (1.74)</td>
<td>6.90 (1.67)</td>
<td>5.07 (1.94)</td>
</tr>
<tr>
<td>Change</td>
<td>0.50 (0.50)</td>
<td>1.34&lt;sup&gt;a&lt;/sup&gt; (0.96)</td>
<td>-0.23 (0.28)</td>
</tr>
</tbody>
</table>

Notes: For each measure, means with different superscripts are statistically different with a type I error rate of .05.

**Total number of experiments designed.** For the total number of experiments designed the main effect of instructional condition was not statistically significant, $F(2,57) = .42, p = .661$. Regardless of condition, participants designed a statistically comparable number of experiments, though this finding could be a result of the experiment not having enough power to find the difference.

**Percentage of correctly designed experiments.** The main effect of instruction was statistically significant, $F(2,57) = 17.73, p < .001$. To further examine the effect of condition, Fisher LSD pairwise comparisons were performed. The mean of the guided instruction condition ($M = 66.13$) was statistically greater than the means of direct instruction ($M = 50.54$) and discovery learning ($M = 33.59$), $F(1,57) = 35.53$ and $9.13, p =
.005 and < .001, Cohen’s $d_s = 0.90$ and $1.71$ respectively. Furthermore, the difference between direct instruction and discovery learning was statistically significant, $F(1,57) = 8.649$, $p = .004$, $d = 1.12$. As hypothesized, participants who received guided instruction correctly designed more experiments correctly than participants who received direct instruction or discovery learning.

**Learning Outcomes.** Analyzing the multiple choice learning outcomes was completed with one-way ANOVAs with condition as the between-subjects factor and the learning outcome as the dependent variable. Statistically significant main effects were investigated with Fisher LSD pairwise contrasts with type I error protected at .05 for each analysis. Source tables for the analysis of learning outcomes can be found in Appendix X.

**Cued recall of scientific vocabulary.** The main effect of instruction was statistically significant, $F(2,57) = 14.75$, $p < .001$. To further examine the effect of condition, Fisher LSD pairwise comparisons were performed. The mean of the guided instruction condition ($M = 2.30$) was not statistically greater than the means of direct instruction ($M = 2.45$) but was statistically greater than discovery learning ($M = 1.00$), $F(1,57) = .26$ and $24.32$, $p = .611$ and < .001, Cohen’s $d$ for the latter contrast was $1.41$. Furthermore, the difference between direct instruction and discovery learning was also statistically significant, $F(1,57) = 24.32$, $p < .001$, $d = 1.56$. As hypothesized, participants who received guided instruction answered more cued recall questions correctly than participants who received discovery learning. Contrary to the hypothesis regarding guided instruction vs. direct instruction, participants in the guided instruction treatment did not answer more cued-recall questions than those in the direct instruction treatment.
Application of control variable strategies. The main effect of instructional condition was statistically significant, $F(2,57) = 3.32$, $p = .043$. Fisher LSD pairwise comparisons revealed the mean of the guided instruction condition ($M = 5.90$) was not statistically greater than the mean of the direct instruction condition ($M = 6.10$), $F(1,57) = .08$ and $p = .780$. The mean of the guided instruction condition was statistically greater than the discovery learning condition ($M = 4.40$), $F(1,57) = 4.33$, $p = .042$, $d = 0.68$. Furthermore, the difference between direct instruction and discovery learning was also statistically significant, $F(1,57) = 5.56$, $p = .022$, $d = 0.75$. As hypothesized, participants who received guided instruction answered more application questions correctly than participants who received discovery learning. Contrary to what was hypothesized, participants in the guided instruction treatment did not answer more application questions than those in the direct instruction treatment.

Evaluation of science fair posters. The main effect of instruction was statistically significant, $F(2,57) = 5.53$, $p = .006$. Fisher LSD pairwise comparisons were indicated the mean of the guided instruction condition ($M = 1.95$) was not statistically greater than the means of the direct instruction condition ($M = 2.10$) $F(1,57) = .27$, $p = .607$. However, guided instruction was statistically greater than the discovery learning condition ($M = 1.20$), $F(1,57) = 6.69$, $p = .012$, $d = 0.82$ respectively. Furthermore, the difference between direct instruction and discovery learning was also statistically significant, $F(1,57) = 9.63$, $p = .003$, $d = 0.97$. As hypothesized, participants who received guided instruction answered more evaluation questions correctly than participants who received discovery learning. Contrary to expectations, participants in the guided instruction
treatment did not answer more evaluation questions than those in the direct instruction treatment.

**Science Self-Efficacy.** A repeated-measures ANOVA was conducted with condition as the between-subjects factor and time as the within-subjects factor. The statistically significant omnibus test for the interaction was interpreted with bar graphs and tetrad comparisons. Source tables for the analysis of changes in science self-efficacy scores can be found in Appendix Y.

The main effect of time was significant, $F(2,57) = 4.40$, $p = .001$. At pretest the mean self-efficacy was 5.15 and at the posttest the mean self-efficacy was 5.53, $d = 0.26$. The interaction between time of science self-efficacy measurement and instruction was statistically significant, $F(2,57) = 33.56$, $p < .001$. To further investigate this interaction, cell means were plotted (see Figure 2) and tetrad comparisons were performed. The participants who received guided instruction had greater change in science self-efficacy scores ($M_1 = 5.57$ and $M_2 = 6.90$) than those who received direct instruction ($M_1 = 4.58$ and $M_2 = 4.63$) and those who received the discovery learning treatment ($M_1 = 5.29$ and $M_2 = 5.07$). Tetrad comparison of guided instruction vs. direct instruction and guided instruction vs. discovery learning means revealed a statistically significant difference in change scores, $F(1, 57) = 39.99$ and $58.85$, $ps < .001$, $ds = 0.27$ and 0.32, respectively. Furthermore, the tetrad comparison of the change scores for direct instruction vs. discovery learning was not statistically significant, $F(1,57) = 1.82$, $p = .183$. Thus, as hypothesized participants who received guided instruction had greater changes in self-efficacy than those who received direct instruction or discovery learning. Contrary to
expectations, participants who received direct instruction did not have greater changes in self-efficacy than those who received discovery learning.

Figure 2 Mean in science self-efficacy at pretest and posttest by condition.
CHAPTER 5
DISCUSSION

Study Overview

This study investigated whether the form of instruction affected student learning and science self-efficacy when nine- and ten-year-old children learn control of variable strategies (CVS). Sixty children were randomly assigned to three conditions: 1) guided instruction, where children received instruction with examples and explanations; 2) direct instruction, where children received instruction through lecture with examples; and 3) discovery learning, where children received instruction through methods of self-discovery. Before starting the experiment, participants completed a prior knowledge pretest and a measure of science self-efficacy. After receiving a common introduction to scientists’ use of experiments, children received the instructional treatment and completed the learning outcomes including: 1) designing of their own experiments; 2) a recall measure about CVS; 3) an application measure about applying CVS to scientific experiments; 4) an evaluation measure using CVS to evaluate scientific experiments and 5) a final science self-efficacy measure. The following sections will discuss the theoretical and educations implications, strengths and limitations of this study and future research.

Results from the current study have theoretical implications to add to our understandings of how instructional strategies impact learning and science self-efficacy. Generally, there are condition-related differences related to the design of experiments, learning outcomes and changes in science self-efficacy. For the purpose of this section, the results will be discussed in relationship to the three main hypotheses tested and
conclude with theoretical implications.

The first hypotheses concerned instructional strategy-related differences in the number of experiments designed and percentage of unconfounded experiments designed. It was expected that children who received guided instruction would design a higher number of experiments with a higher percentage of unconfounded experiments than those who received direct instruction and children who receive direct instruction would design a higher number of experiments with a higher percentage unconfounded experiments than those receiving discovery learning. This hypothesis is partially supported by these results. There were group differences, suggesting that the level of instructional support affects student ability to design unconfounded experiments in a single instructional session. Specifically, though all groups designed a similar number of experiments, students receiving guided instruction designed a significantly higher percentage of unconfounded experiments than those receiving direct instruction and those receiving discovery learning designed a significantly lower percentage of unconfounded experiments than direct instruction.

Those children who received direct instruction designed a higher percentage of correct experiments than those children receiving discovery methods of instruction. Klahr and Nigam (2004) and Klahr and Li (2005) found similar findings in regard to the stronger performance of the children in the direct instruction when compared to children learning through discovery when designing unconfounded experiments. Klahr and associates attribute the improved performance of the children receiving direct instruction to the efficiency in learning by being instructed in exactly what was needed. Kuhn and Dean (2005) also found evidence supporting this claim, as children in their study were
measured after the first session and the children instructed with direct instruction were able to design more unconfounded experiments.

In regards to the strong performance of those children who received guided instruction, the literature is conflicting. Klahr and Chen (1999) used similar material and found that children who received direct instruction designed more unconfounded experiments than those in discovery or guided instruction conditions. On this study guided instruction was in the form of giving correct explanations when a wrong answer was given. The difference between the current study and their study was the operationalization of guided instruction. Case (1974) had results similar to the current study. Case used a comparable operationalization of guided instruction in a study with children of a similar age. Case found the children in the guided instruction designed more unconfounded experiments than those receiving discovery learning methods. This study did not have a direct instruction condition.

The second hypotheses proposed that instructional strategy-related differences would exist on the multiple-choice learning outcomes. Specifically it was hypothesized, children who received guided instruction would have higher scores on measures of recall, application and evaluation of CVS than those who received direct instruction and those children who received direct instruction would have higher scores than those who received discovery learning. This hypothesis is partially supported by the current study. Children receiving guided instruction and direct instruction did not perform significantly different on any of the multiple-choice measures. However, both groups of students did significantly better on these measures than those students instructed through discovery.
Recall of terminology and application as used in this study were not part of other studies examining control of variable studies. Klahr and Nigam (2004) employed similar evaluation measures and found a significant difference with a higher number of students who received direct instruction being able to correctly evaluate science fair posters than the number of students who received discovery methods. The main difference between the outcome measures in Klahr and Nigam (2004) and the current study was that the current study employed a multiple-choice format while their study used an interview format. Klahr and Nigam (2004) created a rubric to evaluate student responses to questions and employed cut-off points to distinguish between those who could and could not evaluate the science fair posters.

The lack of a difference between guided instruction and direct instruction in the current study on measures of evaluation is surprising. This relationship has been observed in past studies (Beilin, 1965; Brainerd, 1972; May & Tisshaw, 1975; Wallach & Sprott, 1964). There are several possible reasons for why this finding was absent. First, it could be that there are no group differences. Second, the scores were found to have low reliability (Nunnally & Bernstein, 1994), which impacts statistical power. Third, direct and guided instruction could be equally salient in this context. Finally, the small sample size reduces the power of the study to find significant group differences.

The final hypothesis was that there would be instructional strategy-related differences in change in science self-efficacy. It was expected that children who received guided instruction would have greater changes in science self-efficacy than those who received direct instruction and discovery. Furthermore, children who received direct instruction were expected to exhibit greater changes in science self-efficacy than those
who received discovery learning. Results from the current study provided partial support for the hypotheses. Children who received guided instruction had greater changes in their science self-efficacy than children who received direct instruction or discovery. The finding that guidance enhanced science self-efficacy is consistent with current research supporting the relationship between success and self-efficacy in science (Britner, 2008; Britner & Pajares, 2006; Gungoren & Sungur, 2009; Karaaslan & Sungur, 2011).

**Theoretical Implications**

The current study has theoretical implications regarding how level of instructional guidance impacts student learning and science self-efficacy. First, these findings add to the literature concerning the nature of developing scientific reasoning skills. Piaget (1970) theorized that children learn through their independent interactions with the environment. Counter to Piaget’s theory, this study provides evidence supporting the claim that scientific reasoning skills can be taught to children (Case, 1974; Chen & Klahr, 1999; Dean & Kuhn, 2007; Klahr & Nigam, 2004; Kuhn & Dean, 2005). What needs further study is Piaget’s (1970) claim that learning induced by external forces produces temporary or little change in children’s ability to reason, as this study does not address issues of durability in the learning.

Shulman and Keisler (1966) theorized that the reason direct instruction was more effective than discovery learning was because learner attention was focused to salient aspects during instruction. They argued that discovery was ineffective because some learners never discover the important aspects of the skill. Klahr and Nigam (2004) argue this guidance in focusing on important aspects of instruction facilitates learning in shorter periods of time making direct instruction more efficient than discovery. Evidence from
this study supports the notion that providing guidance produces greater learning in a single session over discovery learning, suggesting it is more effective. Further study is needed to isolate the relationship between focusing student attention and being able to demonstrate the desired skill to examine why this difference exists between instructional strategies.

Finally, the current study provides evidence to support the positive relationship between learning science and science self-efficacy (Britner & Pajares, 2001; Pajares, Britner, & Valiante, 2000). These findings also provide evidence that an instructional strategy can impact student science self-efficacy in a single session. If prior success is a source of science self-efficacy (Britner, 2008) and there is a positive relationship between science self-efficacy and learning, then one possible explanation for the success of those students in the guided instruction conditions is that the interaction during instruction positively affected student beliefs regarding ability to design unconfounded experiments more so than students who received direct instruction or discovery learning. More investigation into the relationships between instructional strategies to teach science and science self-efficacy are needed.

Educational Implications

This study further supports incorporating instructional guidance when teaching young students CVS. The relatively low performance of discovery condition children in designing experiments and answering questions was consistent throughout the study. This evidence suggests that providing guidance during instruction helps students focus their attention on relevant aspects of the instruction, which has demonstrated an increase in learning in shorter periods of time (Klahr & Nigam, 2004). Classroom teachers must use
their time carefully to ensure adequate coverage of curriculum; and efficiency when time and resources are important. Caution must be taken when applying these results to a classroom situation. While students who received guided instruction and direct instruction performed better on learning measures after a single session the durability of their learning over time is unknown. Other studies (Dean and Kuhn; 2007; Kuhn and Dean; 2005) have demonstrated that learning may be maintained longer when discovery methods are employed during instruction that is longer than a single session. This suggests that mastery cannot happen in a short amount of time.

Results from the study also suggest it is important to consider the alignment between instruction and assessment. The experimental design activity and multiple-choice outcomes provide evidence supporting the effectiveness of different instructional strategies. The mixed findings suggest that when instructing on CVS the correct amount of instructional guidance might be dependent on the learning outcome. For example, students who received guided instruction performed statistically better than those receiving direct instruction when designing experiments but comparably when they answered multiple-choice questions. Therefore, educators should consider student-learning outcomes when selecting instructional strategies to use in their classrooms.

Educational implications of these findings also suggest that level of instructional support does impact student science self-efficacy. When teaching the use of CVS to design unconfounded experiments guided instruction may be a better way of instructing when considered with students’ science self-efficacy over using discovery or direct instructional methods in a single session setting. Self-efficacy, as a construct of motivation, is important to the learning environment because students must be motivated
to engage in learning science (Sanfeliz & Stalzer, 2003). If students do not have high self-efficacy they are less likely to engage in complex cognitive processes, regardless of the amount of instructional guidance present in the lesson.

**Strengths and Limitations**

This study has five main strengths involving causality and generalizability. First, this study employed random assignment of individuals to one of the three conditions. This strengthens causal claims that the results are due to treatment, and not differences that were preexisting. Second, the children in this study are of the age most likely to encounter this topic in their science courses as indicated in the state standards for fourth grade (NM, 2003). Third, this is likely the first study to examine science self-efficacy in this context. Previous work on the affects of instructional strategies has focused on learning outcomes and not impacts on motivation. Fourth, this study adds to our understanding of changes in science self-efficacy in elementary-age children, which is an age group that has not received a lot of examination in the literature. Finally, the materials incorporated in this study as well as the sequence of events (instruction-practice-assessment) within the sessions follow a common sequence of instruction that can be found in science classrooms.

This study has four limitations involving causality and generalizability of results. First, this sample was from a convenience sample with a small number of students participating in a summer fitness camp in the southwestern United States. It is probable they differ in a substantive way from other nine- and ten-year old populations. Caution should be used when generalizing results to other children. Future studies are needed to examine the effects of instructional guidance on learning and science self-efficacy for a
variety of student characteristics. These characteristics might include ethnicity, age, gender and prior knowledge.

Second, the score reliability and validity evidence from the multiple-choice learning measures are of concern. With the exception of the science self-efficacy measure the scores from the measures have low reliability. The correlations of the measures are not correlated (except the pre- and post- science self efficacy measure), this suggests the measures may be measuring different constructs. However, the low correlations and unexpected directions of the relationships are of concern. For example, the evaluation and recall measure were negatively correlated so that as a student did better on the recall they did worse on the evaluation. For these reasons, the scores produced by these measures might not be valid representations of student learning.

Third, the lack of an instrument that measured children’s knowledge before starting this experiment is a limitation. Currently, there is no way to tell how different students are after their session in understanding CVS, because the pretest measure was dropped from the analysis. Even though students were randomly assigned to conditions it is possible there were more advanced students in terms of science reasoning in the guided instruction conditions. The pretest could be improved by: 1) dropping items that correlate poorly with other items; 2) including more items; 3) using a different format besides multiple choice such as asking students to design experiments before they begin; and 4) align test content with Piaget’s (1970) stages of development instead of methods employed in the scientific process. For example, instead of asking students about how to make scientific observations the pretest could ask them to solve problems involving conservation or other reasoning tasks.
Finally, due to the nature of this research, there are limits to the interpretability of these results, as conditions for this study are not found in actual classrooms. This relatively quick, scripted, single session intervention one-on-one with a researcher is lacking in some of the contextualized features found in classrooms. In classroom settings, instruction interacts with other support beneficial to learning such as materials, caring instructors who answer questions and other students. Also, the motivation for learning might be different during an experimental session than a real classroom context since the stakes are low and not tied to course grades or other performance outcomes.

**Future Directions**

Studying instructional strategies that help students achieve mastery of using control of variable strategies to design unconfounded experiments has many potential areas of research. The phase model of research, as described in Marley and Levin (2011) and Levin and O’Donnell (1999), can be used to consider future directions. The phase model is a systematic way to scale research from the generation of ideas, through laboratory studies to whole classroom experiments in order to ensure prescriptive statements are based in credible research from multiple studies. The first phase involves collecting observations, completing case studies, or administering surveys. This allows researchers to generate hypotheses and identify important patterns. During the second phase, researchers would follow up on their findings from phase one by employing systematic research incorporating random assignment. The focus on the second phase should be to gather evidence to isolate specific information in regards to which students and in what contexts the strategies are effective. During the third phase, researchers should begin addressing issues of ecological validity in regards to the targeted
environment such as a classroom setting. The focus of the third phase should be to increase the generalizability of the research without sacrificing internal validity.

An example of the phase model of research as applied to the future directions implied by this study might include investigating the interaction between time, instructional strategy and mastery of CVS. The first phase of this research agenda would include operationalizing terminology increasing the reliability in the scores from outcome measures, and generating hypotheses. First, a consistent definition of guided instruction would be needed to begin programmatic study. This could be gained through reviews of literature as well as interviews and surveys with practitioners and researchers. Second, the reliability of scores produced by the measures incorporated in the current study must be addressed. This would include a review of literature detailing the instruments being employed to measure mastery of CVS to design unconfounded experiments to ascertain validity evidence. Then studies could be conducted to improve the reliability in scores produced by the measures. Finally, two studies similar to Kuhn and Phelps (1982) would be conducted. These researchers made careful observations and conducted detailed interviews concerning how old strategies are replaced with new strategies when instruction is done through discovery. Their reports contain rich descriptions of how success in one session did not guarantee the strategies incorporated in following sessions and how the processes of replacing ineffective strategies involve a merging of new and old before the more effective strategy prevails. Similar studies should be completed with students instructed with direct and guided instruction.

During the second phase, studies could be done on relationships identified in phase one employing tightly controlled laboratory-based experiments. This might include
investigating, age, experience, gender, race or other such individual characteristic on learning with a specific strategy. For this study, the relationship between time during treatment, time after treatment, and mastery of CVS could be investigated by systematically manipulating these variables over repeated experiments. First, a series of experiments would be done to isolate the role of time in mastery of CVS strategies. These studies would include varying the amount of time spent in an instructional session (e.g. 30, 60, 90 minutes), manipulating the number of sessions (e.g. 1, 2, 3 sessions), and investigating the effects of the amount of time between sessions (e.g. 1 day apart, 1 week apart, 1 month apart). Beyond session time, programmatic study should include investigating the durability of the learning by giving outcomes at the end of sessions, but also days, weeks, and/or months after the sessions have ended. The goal in the second phase would be to isolate the specific relationship between the time spent with instruction, the durability of the learning, and instructional strategies.

At the third phase, increasing the generalizability without weakening the internal validity is most important. The goal is to scale research from the laboratory environment to conditions similar in the targeted applied setting. For example, systematically increasing the number of students who receive instruction at a single time could occur until the conditions of the study are consistent with classroom contexts. Allowing students to work together in pairs and then larger groups would also increase the generalizability of findings to an actual classroom setting. Another study could be done to incorporate the isolated instructional session used in phase two into a larger cohesive unit. For example, incorporating several learning experiences of the curriculum within an entire course could be done. At this phase, and the previous phases, replication of
findings is important before prescriptive statements about the effectiveness of an instructional strategy in the classroom.

**Conclusion**

This study was designed to investigate whether level of instructional guidance impacts student learning and science self-efficacy. Guided instruction was a better instructional strategy than direct instruction and discovery when the learning outcome was designing unconfounded experiments. Guided instruction and direct instruction produced statistically comparable scores that were better than scores produced by students who received discovery learning on learning outcome measures in multiple choice formats. In addition, the results indicate the use of guided instruction produced greater changes in science self-efficacy than discovery and direct instruction. The results have implications both theoretically and educationally. However, there are limitations to the study that are important to keep in mind. Future research is needed in understanding the formation of strategy use in all three types of instruction and the role time plays in young children developing mastery of using control of variable strategies to design unconfounded experiments.
References


Brainerd, C. J. (1972). Reinforcement and reversibility in quantity conservation 

Brainerd, C. J. (1978). The stage question in cognitive-developmental theory. The 
Behavioral and Brain Sciences, 1, 173-213.

differences in life, physical and earth science classes. Research In Science 
Teaching, 45, 955-970.

Britner, S.L., & Pajares, F. (2001). Self-efficacy beliefs, motivation, race, and gender in 
middle school science. Journal of Women and Minorities in Science and 


advanced placement intent of high school students learning science. Science 
Education, 14, 1047-1065.

Bolte (Eds.), Scientific literacy, (pp. 37-68). Kiel, Germany: Institute for Science 
Education.


Chen, Z., & Klahr, D. (1999). All other things being equal: Children’s acquisition of the


grade and eighth-grade students in an international context (NCES 2009-001).

Retrieved from the National Center for Educational Statistics website

students’ motivational beliefs in science. *The International Journal of Learning,*
16, 496-506.

self-efficacy for African Americans and European Americans. *Journal of Career
Assessment,* 9, 261 – 281.

Hackett, G., & Betz, N. E. (1981). A self-efficacy approach to the career development of


Leadership,* 16(52), 13-16.

McNally.


Karaarslan, G. & Sungur, S. (2011). Elementary students’ self-efficacy beliefs in science:
Role of grade level, gender, and socio-economic status. *Science Education
International,* 22(1), 72-79.


MI: National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan.


Appendix A: Session One Script for Pilot and Primary Study

Session 1
Introduction:

“Thank you for being here today. My name is Glenn and I am a researcher here at the University of New Mexico. I am working on a project to understand how we learn science. If you agree, you will be asked to spend time with another researcher where you will learn about science experiments and answer some questions. First, though, please read through this assent form, as I read it aloud and if you agree sign it before we begin. Remember, even though you are signing it now you can withdraw at anytime during any of the session. (READ FORM) Do you have any questions about being in this study?”

(Students receive and fill out the assent form. The researcher will read it to them before they sign)

“Today I am going to ask you to answer some questions regarding science and how you feel about science. Please answer the questions to the best of your ability.”

(Students fill out the pre-test and the science self-efficacy scale)
Appendix B: Direct Instruction Script, Pilot and Primary Study

For the script, sections that are in *italics* are common statements used by the researcher in all conditions. Sections that are **bold and italics** are statements that were varied in each treatment. Sections without bolding and italics are representing where a researcher action or student response occurs.

“Scientists use an experiment to search for relationships in nature. In other words, they design an experiment so that changes to one thing can cause something else to vary in a predictable way.

These changing quantities are called variables. A variable is any factor, or trait that can exist in differing amounts or types. An experiment usually has three kinds of variables: independent, dependent, and controlled.

The independent variable is the one that is changed by the scientist. A good experiment usually has only one independent variable. As the scientist changes the independent variable, he or she watches what happens in the other variables.

The scientist focuses his or her observations on the dependent variable to see how it responds to changes made to the independent variable.

For example, if you wanted to see what affects the distance the ball rolls you might be curious about the type of surface the ball is rolling across. So you want to see how surface type (the independent variable) affects the distance the ball rolls (dependent variable)—you might observe that the ball rolls farther on smooth surfaces.

Experiments also have constant variables. Constant variables are quantities that a scientist wants to remain constant, and must watch them as carefully as the dependent variables.”

(Researcher designs a confounded experiment: ramp with carpet and ramp with wood set up and steep ramp vs shallow ramp)

“Can you tell if the starting position affects the distance the ball rolls?”

(Students answer: No)

“If you wanted to see what affects the distance the ball rolls you might be curious about the surface type under the ball, it is important to make sure that the steepness of the ramp (the constant variable) is held constant as well as all other variables. That’s because both the steepness of the ramp and the surface have an impact on how far the ball rolls. If we change both of them at the same time, we can’t be sure how much of the change is because of the steepness and how much is because of the surface. In this study the surface is the independent variable and the distance the ball rolls in the dependent variable.”
(Researcher designs an unconfounded experiment: one ball ping pong the other golf)

“Can you tell if the ball type affects the distance the ball rolls?”

(Students answer: yes)

“If you wanted to see what affects the distance the ball roll you might investigate how the type of ball impacts the distance, it is important to make sure that the steepness of the ramp (the constant variable) is held constant and all other variables are the same. That’s because both the steepness of the ramp and the surface have an impact on how far the ball rolls. If we change both of them at the same time, we can’t be sure how much of the change is because of the steepness and how much is because of the surface. In this study the type of ball is the independent variable and the distance the ball rolls in the dependent variable and all other variables are constant.

Before we start the next activities, do you have any questions regarding designing science experiments with independent, dependent and control variables?”
Appendix C: Guided Instruction Script, Pilot and Primary Study

For the script, sections that are in *italics* are common statements used by the researcher in all conditions. Sections that are **bold and italics** are statements that were varied in each treatment. Sections without bolding and italics are representing where a researcher action or student response occurs.

“Scientists use an experiment to search for relationships in nature. In other words, they design an experiment so that changes to one thing can cause something else to vary in a predictable way.

These changing quantities are called variables. A variable is any factor, or trait that can exist in differing amounts or types. An experiment usually has three kinds of variables: independent, dependent, and controlled.

The independent variable is the one that is changed by the scientist. A good experiment usually has only one independent variable. As the scientist changes the independent variable, he or she watches what happens in the other variables.

The scientist focuses his or her observations on the dependent variable to see how it responds to changes made to the independent variable.

For example, if you wanted to see what affects the distance the ball rolls you might be curious about the type of surface the ball is rolling across. So you want to see how surface type (the independent variable) affects the distance the ball rolls (dependent variable)—you might observe that the ball rolls farther on smooth surfaces.

(RESEARCH DESIGNS EXPERIMENT)

Experiments also have constant variables. Constant variables are quantities that a scientist wants to remain constant, and must watch them as carefully as the dependent variables.”

(Researcher designs a confounded experiment: ramp with carpet and ramp with wood set up and steep ramp vs shallow ramp)

“Can you tell if the starting position affects the distance the ball rolls? Why?”

(Students answer: No because there are 2 variables being changed: surface and length)

“How can we change this experiment so that we can make a clear conclusion regarding the surface area and the distance the ball rolls?”

(Students answer: you need to use the same length on each ramp)
“Which variable is the independent variable and which is the dependent variable? What are some other variables we have controlled for in this experiment?”

(Students answer: IV: surface, DV distance rolled, CV: steepness, type, length)

(Researcher designs an unconfounded experiment: one ball ping pong the other golf)

“Can you tell if the ball type affects the distance the ball rolls? Why?”

(Students answer: yes all variables are held equal)

“Which variable is the independent variable and which is the dependent variable? What are some other variables we have controlled for in this experiment?”

(Students answer: IV ball type, DV length of roll, CV all other variables)

“Before we start the next activities, do you have any questions regarding designing science experiments with independent, dependent and control variables?”
Appendix D: Discovery Script, Pilot and Primary Study

For the script, sections that are in *italics* are common statements used by the researcher in all conditions. Sections that are *bold and italics* are statements that were varied in each treatment. Sections without bolding and italics are representing where a researcher action or student response occurs.

“Scientists use an experiment to search for relationships in nature. In other words, they design an experiment so that changes to one thing can cause something else to vary in a predictable way.

“Can you design an experiment to see how surface type affects the distance the ball rolls?”

*(Students design experiments)*

“Can you design an experiment to see how starting length affects the distance the ball rolls?”

*(Students design experiments)*

“Can you design an experiment to see how the ball type affects the distance the ball rolls?”

*(Students design experiments)*
Appendix E: Introduction Session Two Script, Pilot

Introduction:
"Hello! My name is Cari and I want to thank you for being here today. Today we are going to learn about designing science experiments. I am going to tell you about how we design experiments and then allow you to design some. Then you will answer some questions on your own."

Training Session:
“Today we will be investigating what affects the distance the ball rolls with this ramp. You will notice there are some things about the ramp you can change such as the surface of the ramp, the type of ball used, the steepness of the ramp, and the distance of the ramp. Do you have any questions? Would you like to play with the ramps before we start?”
Appendix F: Instructions for Designing Experiments and Learning Outcomes, Pilot

**Design Activity:**

“Ok. So now I have a challenge for you. After this you will do some activities alone and then return to your camp group. I am going to give you 10 minutes, and I would like you to design as many experiments as possible that involve this ramp and the distance the ball rolls. So you are going to design an experiment (just like I did above) and let me know it is done. I will check and then you will design another one. I will count how many are designed, try and get as many as you possibly can in 10 minutes. Do you have any questions? Okay. Ready, Set, Go!”

(Students then designs an experiment with the two ramps)

“Excellent. You have 5 minutes left. Keep it up”

(Students then designs an experiment with the two ramps)

“Excellent. You have 2 minutes left. Keep it up”

(Students answer questions above)

**Testing**

“Well done, you designed _{(number of experiments)}_ experiments.

*Here is a test to see what you learned today. Please do the best you can on these questions. If you need help, please ask and I will do my best to answer your question.*”

(Student does recall, application, evaluation, and CSAI assessment)

“Thank you for participating today. I hope you enjoy the rest of your summer! Now lets go find your group.”
Appendix G: Study Evaluation of Science Experiments, Pilot

Algae Bloom

<table>
<thead>
<tr>
<th>Question: How does fertilizer effect algae?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis: The more fertilizer there is the more the algae will grow</td>
</tr>
<tr>
<td>Materials: 8 4 oz jars Fertilizer Water Measuring cup Algae</td>
</tr>
</tbody>
</table>

| Procedure: |
| 1. Fill each jar with water |
| 2. Add fertilizer to each jar. |
| In 2 jars don’t add any |
| In 2 jars add 1 tablespoon |
| In 2 jars add 2 tablespoons |
| In 2 jars add 3 tablespoons |
| 3. add the algae to each jar |
| 4. label each jar |
| 5. put each jar into sunny windows |

| Results: |
| The jar with 2 grew good but not the best. The jar with none kinda grew. The jar with 1 tablespoons grew the best. |

| Conclusion: |
| Fertilizer helps algae grow but not too much. |

1. What is the independent variable?
   a. The amount of algae that grew
   b. **The amount of fertilizer**
   c. The water in the jar
   d. The sunny window

2. What was the dependent variable?
   a. The measuring cups
   b. The different amounts of fertilizer
   c. **How much algae grew**
   d. Water used in the jars

3. Did they control for any other variables? If so, what were they?
   a. **Yes. They used the same size jar and put all jars in the same window**
   b. No. They did not need to keep variables controlled
   c. Yes. They used different amounts of fertilizer
   d. No. The conclusion does not mention any controlled variables

4. Were there other variables that should be controlled?
   a. Controlling variables in this experiment is not important
   b. They should have added 4 tablespoons of fertilizer to the jar
   c. They should not label each jar
   d. **They should make sure to add the same amount of water to each jar**
5. How would you make this experiment better?

### Can Dogs See Color?

<table>
<thead>
<tr>
<th>Question: Can dogs see different colors?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis: Dogs can see in color, they’ll probably like blue the best</td>
</tr>
<tr>
<td>Materials: 10 dogs, 5 sheets of paper of different colors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Procedure:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lay out 5 different colors of papers.</td>
</tr>
<tr>
<td>2. Cover them with a towel</td>
</tr>
<tr>
<td>3. Bring the dog out and uncover the papers</td>
</tr>
<tr>
<td>4. Count how many times a dog smells each color</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Results:</th>
</tr>
</thead>
</table>
| ![Bar graph showing sniffs by color with blue having the most sniffs]

| Conclusion: Each dog liked a different color, but the color with the most sniffs was black. Dogs can see color they just like black better than all the other colors |

6. What is the independent variable?
   a. The number of dogs
   b. **Color of papers**
   c. The number of sniffs on each paper
   d. The blue paper

7. What was the dependent variable?
   a. 5 different colors of paper
   b. Type of dog used
   c. **How many times each paper was sniffed**
   d. Towel used to cover the paper

8. Did they control for any other variables? If so, what were they?
   a. **Yes. They covered the paper before the dog arrived**
   b. No. They did not need to keep variables controlled
   c. Yes. Counted how many times the dog smelled each color
   d. No. The conclusion does not mention any controlled variables
9. Were there other variables that should be controlled?
   a. Controlling variables in this experiment is not important
   b. They should have used a pink paper
   c. They should have used different sizes of paper instead of just color
   d. They should make sure to test all dogs in the same location

10. How would you make this experiment better?
Appendix H: Demographic Questions, Pilot and Primary Study

Demographic Survey Questions:

Please answer the following questions to the best of your ability.

1. What year were you born? ________________
2. What is your gender (circle one)  Male   Female
3. What is your ethnicity _____________
4. What grade will you be starting in August? ______________
5. What language do you prefer to speak? ______________
Appendix I: Pretest, Pilot and Primary Study

Please circle the answer you feel is the best.

1. Which sentence describing a snake is an example of a scientific observation?
   a. It is fast
   c. **It does not have legs**
   d. It is an excellent hunter
   e. It eats lots of mice

2. Why is it important to write each step of your scientific experiment?
   a. So others can repeat the experiment and compare results
   b. So others can copy the steps and then do a different experiment
   c. So others can make a conclusion based on their own opinion
   d. So others can make a conclusion before they look at the data

3. Which of the following is the best way for students to record the number of leaves on a plant?
   a. Building a model
   b. Drawing pictures
   c. Writing a description
   d. **Recording the number in a table**

4. Identify something a person should do to make their observations more accurate.
   a. Make several observations and record them
   b. Change the location of the observations
   c. Just look carefully
   d. You can’t make an observation accurate

5. Which statement about fruit is a fact rather than an opinion?
   a. Oranges are too hard to peel
   b. Apples are the best tasting fruit
   c. Plums are the hardest fruit to eat
   d. **Bananas turn brown after a few days**

6. What is a hypothesis?
   a. A guess
   b. An explanation for something observed
   c. Scientific theory
   d. The question in a scientific experiment
7. How fast does a candle burn? What would be the dependent variable in this experiment?
   a. Size of the candle
   b. Match used to light the candle
   c. Type of candle
   d. **Time it takes candle to burn**

8. Two students think that wintertime thunderstorms are always followed by snow. Which of the following would be the best way to test this hypothesis?
   a. Measure the snowfall after the first winter thunderstorm
   b. Compare data collected for one week with data from another group
   c. Tape record a show on tornadoes and their destruction
   d. **Analyze winter weather records over several years**

9. Students are planning an experiment to determine if bean seeds can grow without light. The dependent variable for this experiment is
   a. Amount of soil in the pots
   b. Amount of light given to the plants
   c. Amount of water given to the plants
   d. **Amount of growth of the plant**

10. A variable in a scientific experiment is
    a. The poster designed after the experiment
    b. The question asked by the scientist
    c. The way you measure something
    d. **A factor that can exist in different amounts in the experiment**
Appendix J: Pretest Item Analysis

Sample items:

1. Which statement about fruit is a fact rather than an opinion?
   a. Oranges are too hard to peel
   b. Apples are the best tasting fruit
   c. Plums are the hardest fruit to eat
   d. Bananas turn brown after a few days

2. What is a hypothesis?
   a. A guess
   b. An explanation for something observed
   c. Scientific theory
   d. The question in a scientific experiment

<table>
<thead>
<tr>
<th>Pre1</th>
<th>Pre2</th>
<th>Pre3</th>
<th>Pre4</th>
<th>Pre5</th>
<th>Pre6</th>
<th>Pre7</th>
<th>Pre8</th>
<th>Pre9</th>
<th>Pre10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item-Total Statistics

<table>
<thead>
<tr>
<th>Scale Mean if Item Deleted</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
<th>Percent Participants Answering Correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre1</td>
<td>2.12</td>
<td>2.105</td>
<td>-.067</td>
<td>.206</td>
<td>.268</td>
</tr>
<tr>
<td>Pre2</td>
<td>2.10</td>
<td>1.854</td>
<td>.156</td>
<td>.171</td>
<td>.159</td>
</tr>
<tr>
<td>Pre3</td>
<td>2.05</td>
<td>1.947</td>
<td>.037</td>
<td>.319</td>
<td>.222</td>
</tr>
<tr>
<td>Pre4</td>
<td>2.03</td>
<td>1.965</td>
<td>.014</td>
<td>.155</td>
<td>.235</td>
</tr>
<tr>
<td>Pre5</td>
<td>2.08</td>
<td>1.671</td>
<td>.325</td>
<td>.344</td>
<td>.060</td>
</tr>
<tr>
<td>Pre6</td>
<td>2.08</td>
<td>1.773</td>
<td>.221</td>
<td>.170</td>
<td>.121</td>
</tr>
<tr>
<td>Pre7</td>
<td>2.17</td>
<td>2.073</td>
<td>-.006</td>
<td>.165</td>
<td>.235</td>
</tr>
<tr>
<td>Pre8</td>
<td>1.95</td>
<td>1.981</td>
<td>-.025</td>
<td>.109</td>
<td>.263</td>
</tr>
<tr>
<td>Pre9</td>
<td>2.03</td>
<td>1.660</td>
<td>.286</td>
<td>.154</td>
<td>.072</td>
</tr>
<tr>
<td>Pre10</td>
<td>1.93</td>
<td>2.097</td>
<td>-.112</td>
<td>.184</td>
<td>.313</td>
</tr>
</tbody>
</table>
Appendix K: Science Self-Efficacy Questions, Pilot and Primary Study

Please rate how certain you are that you can do each of the things described below by circling the number:

I can learn science

0 1 2 3 4 5 6 7 8 9 10
Cannot do it Pretty sure I can do it For sure I can do

I can finish my science homework by deadlines

0 1 2 3 4 5 6 7 8 9 10
Cannot do it Pretty sure I can do it For sure I can do

I can get myself to study science when there are other interesting things to do

0 1 2 3 4 5 6 7 8 9 10
Cannot do it Pretty sure I can do it For sure I can do

I can concentrate on science during class

0 1 2 3 4 5 6 7 8 9 10
Cannot do it Pretty sure I can do it For sure I can do

I can take good notes during science instruction

0 1 2 3 4 5 6 7 8 9 10
Cannot do it Pretty sure I can do it For sure I can do

I can design science experiments well

0 1 2 3 4 5 6 7 8 9 10
Cannot do it Pretty sure I can do it For sure I can do

I can ask good science questions

0 1 2 3 4 5 6 7 8 9 10
Cannot do it Pretty sure I can do it For sure I can do
I can remember well information presented in science class

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot do it</td>
<td>Pretty sure I can do it</td>
<td>For sure I can do</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I can make a good scientific hypothesis

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot do it</td>
<td>Pretty sure I can do it</td>
<td>For sure I can do</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I can get myself to do my science school work

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot do it</td>
<td>Pretty sure I can do it</td>
<td>For sure I can do</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix L: Science Self-Efficacy Item Analysis

Sample items:
1. I can design science experiments well
2. I can ask good science questions

<table>
<thead>
<tr>
<th></th>
<th>Cannot do it</th>
<th>Pretty sure I can do it</th>
<th>For sure I can do it</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Inter-Item Correlation Matrix: Time 1*

<table>
<thead>
<tr>
<th></th>
<th>SE1.1</th>
<th>SE1.2</th>
<th>SE1.3</th>
<th>SE1.4</th>
<th>SE1.5</th>
<th>SE1.6</th>
<th>SE1.7</th>
<th>SE1.8</th>
<th>SE1.9</th>
<th>SE1.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE1.1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE1.2</td>
<td>.436</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE1.3</td>
<td>.522</td>
<td>.330</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE1.4</td>
<td>.413</td>
<td>.293</td>
<td>.396</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE1.5</td>
<td>.518</td>
<td>.434</td>
<td>.465</td>
<td>.563</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE1.6</td>
<td>.466</td>
<td>.441</td>
<td>.512</td>
<td>.433</td>
<td>.462</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE1.7</td>
<td>.672</td>
<td>.487</td>
<td>.619</td>
<td>.492</td>
<td>.683</td>
<td>.609</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE1.8</td>
<td>.468</td>
<td>.463</td>
<td>.473</td>
<td>.519</td>
<td>.609</td>
<td>.523</td>
<td>.762</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE1.9</td>
<td>.273</td>
<td>.327</td>
<td>.300</td>
<td>.411</td>
<td>.369</td>
<td>.324</td>
<td>.408</td>
<td>.804</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>SE1.10</td>
<td>.275</td>
<td>.404</td>
<td>.350</td>
<td>.413</td>
<td>.535</td>
<td>.413</td>
<td>.522</td>
<td>.884</td>
<td>.892</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Item-Total Statistics: Time 1*

<table>
<thead>
<tr>
<th></th>
<th>Scale Mean</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
<th>Item Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE1.1</td>
<td>45.9500</td>
<td>270.387</td>
<td>.608</td>
<td>.539</td>
<td>.895</td>
<td>5.500</td>
</tr>
<tr>
<td>SE1.2</td>
<td>46.0000</td>
<td>283.085</td>
<td>.533</td>
<td>.324</td>
<td>.900</td>
<td>5.450</td>
</tr>
<tr>
<td>SE1.3</td>
<td>46.3833</td>
<td>268.173</td>
<td>.594</td>
<td>.438</td>
<td>.897</td>
<td>5.067</td>
</tr>
<tr>
<td>SE1.4</td>
<td>46.0500</td>
<td>270.455</td>
<td>.588</td>
<td>.432</td>
<td>.897</td>
<td>5.400</td>
</tr>
<tr>
<td>SE1.5</td>
<td>46.2667</td>
<td>265.114</td>
<td>.703</td>
<td>.625</td>
<td>.889</td>
<td>5.183</td>
</tr>
<tr>
<td>SE1.6</td>
<td>46.3667</td>
<td>276.507</td>
<td>.629</td>
<td>.450</td>
<td>.894</td>
<td>5.083</td>
</tr>
<tr>
<td>SE1.7</td>
<td>46.6167</td>
<td>254.952</td>
<td>.806</td>
<td>.831</td>
<td>.882</td>
<td>4.833</td>
</tr>
<tr>
<td>SE1.8</td>
<td>46.7833</td>
<td>266.444</td>
<td>.848</td>
<td>.916</td>
<td>.882</td>
<td>4.667</td>
</tr>
<tr>
<td>SE1.9</td>
<td>46.2000</td>
<td>268.095</td>
<td>.598</td>
<td>.845</td>
<td>.897</td>
<td>5.250</td>
</tr>
<tr>
<td>SE1.10</td>
<td>46.4333</td>
<td>272.046</td>
<td>.707</td>
<td>.907</td>
<td>.890</td>
<td>5.017</td>
</tr>
</tbody>
</table>
### Inter-Item Correlation Matrix: Time 2

<table>
<thead>
<tr>
<th></th>
<th>SE2.1</th>
<th>SE2.2</th>
<th>SE2.3</th>
<th>SE2.4</th>
<th>SE2.5</th>
<th>SE2.6</th>
<th>SE2.7</th>
<th>SE2.8</th>
<th>SE2.9</th>
<th>SE2.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE2.1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE2.2</td>
<td>.494</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE2.3</td>
<td>.586</td>
<td>.349</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE2.4</td>
<td>.495</td>
<td>.318</td>
<td>.462</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE2.5</td>
<td>.670</td>
<td>.481</td>
<td>.524</td>
<td>.602</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE2.6</td>
<td>.538</td>
<td>.434</td>
<td>.542</td>
<td>.594</td>
<td>.595</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE2.7</td>
<td>.685</td>
<td>.496</td>
<td>.644</td>
<td>.621</td>
<td>.684</td>
<td>.717</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE2.8</td>
<td>.485</td>
<td>.508</td>
<td>.519</td>
<td>.599</td>
<td>.636</td>
<td>.613</td>
<td>.758</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE2.9</td>
<td>.309</td>
<td>.329</td>
<td>.332</td>
<td>.474</td>
<td>.467</td>
<td>.419</td>
<td>.422</td>
<td>.792</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>SE2.10</td>
<td>.363</td>
<td>.430</td>
<td>.363</td>
<td>.530</td>
<td>.567</td>
<td>.482</td>
<td>.475</td>
<td>.852</td>
<td>.915</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Item-Total Statistics: Time 2

<table>
<thead>
<tr>
<th></th>
<th>Scale Mean if Item Deleted</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
<th>Item Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE2.1</td>
<td>49.4167</td>
<td>339.129</td>
<td>.669</td>
<td>.611</td>
<td>.914</td>
<td>5.917</td>
</tr>
<tr>
<td>SE2.2</td>
<td>49.5833</td>
<td>354.518</td>
<td>.544</td>
<td>.377</td>
<td>.920</td>
<td>5.750</td>
</tr>
<tr>
<td>SE2.3</td>
<td>49.9000</td>
<td>334.837</td>
<td>.620</td>
<td>.468</td>
<td>.918</td>
<td>5.433</td>
</tr>
<tr>
<td>SE2.4</td>
<td>49.5167</td>
<td>334.152</td>
<td>.680</td>
<td>.510</td>
<td>.914</td>
<td>5.817</td>
</tr>
<tr>
<td>SE2.5</td>
<td>49.7833</td>
<td>323.291</td>
<td>.763</td>
<td>.638</td>
<td>.909</td>
<td>5.550</td>
</tr>
<tr>
<td>SE2.6</td>
<td>49.9333</td>
<td>334.063</td>
<td>.718</td>
<td>.575</td>
<td>.912</td>
<td>5.400</td>
</tr>
<tr>
<td>SE2.7</td>
<td>50.1000</td>
<td>318.939</td>
<td>.807</td>
<td>.840</td>
<td>.906</td>
<td>5.233</td>
</tr>
<tr>
<td>SE2.8</td>
<td>50.2333</td>
<td>323.572</td>
<td>.853</td>
<td>.897</td>
<td>.904</td>
<td>5.100</td>
</tr>
<tr>
<td>SE2.9</td>
<td>49.6333</td>
<td>335.490</td>
<td>.638</td>
<td>.848</td>
<td>.916</td>
<td>5.700</td>
</tr>
<tr>
<td>SE2.10</td>
<td>49.9000</td>
<td>332.261</td>
<td>.723</td>
<td>.907</td>
<td>.911</td>
<td>5.433</td>
</tr>
</tbody>
</table>
Appendix M: Instructions for Designing Experiments and Learning Outcomes, Primary Study

**Design Activity:**

“Ok. So now I have a challenge for you. After this you will do some activities alone and then return to your camp group. I am going to give you 15 minutes, and I would like you to design as many experiments as possible that involve this ramp and the distance the ball rolls. So you are going to design an experiment (just like I did above) and let me know it is done. I will check and then you will design another one. I will count how many are designed, try and get as many as you possibly can in 10 minutes. Do you have any questions? Okay. Ready, Set, Go!”

(Students then designs an experiment with the two ramps)

“Excellent. You have 5 minutes left. Keep it up.”

(Students then designs an experiment with the two ramps)

“Excellent. You have 2 minutes left. Keep it up.”

(Students answer questions above)

**Testing**

“Well done, you designed ___(number of experiments)___ experiments.

Here is a test to see what you learned today. Please do the best you can on these questions. If you need help, please ask and I will do my best to answer your question.”

(Student does recall, application, evaluation, and CSAI assessment)

“Thank you for participating today. I hope you enjoy the rest of your summer! Now lets go find your group.”
Appendix N: Recall of CVS, Pilot and Primary Study

1. A variable is
   a. A factor in an experiment that can exist in different amounts
   b. A number within an experiment
   c. Anything that can be put on a graph
   d. A factor that needs to be taken out of an experiment

2. A variable that is changed by the scientist is called a
   a. Dependent variable
   b. Independent variable
   c. Controlled variable
   d. Semi-dependent variable

3. The variable measured as it changes because of other variables is the
   a. Dependent variable
   b. Independent variable
   c. Controlled variable
   d. Semi-dependent variable

4. The variable that is held constant is the
   a. Dependent variable
   b. Independent variable
   c. Controlled variable
   d. Semi-dependent variable

5. Why is it important to control some variables during an experiment?
   a. So you can tell how much change in the dependent variable is due to the independent variable
   b. So you can tell how much change in the independent variable is due to the dependent variable
   c. So you can tell how much change in the controlled variable is due to the independent variable
   d. So you can tell how much change in the controlled variable is due to the dependent variable
Appendix O: Recall Item Analysis

Sample Item:
1. The variable measured as it changes because of other variables is the
   a. Dependent variable
   b. Independent variable
   c. Controlled variable
   d. Semi-dependent variable

2. The variable that is held constant is the
   a. Dependent variable
   b. Independent variable
   c. Controlled variable
   d. Semi-dependent variable

Inter-Item Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Recall_1</th>
<th>Recall_2</th>
<th>Recall_3</th>
<th>Recall_4</th>
<th>Recall_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall_1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall_2</td>
<td>.276</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall_3</td>
<td>.110</td>
<td>.284</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall_4</td>
<td>.056</td>
<td>.168</td>
<td>.417</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Recall_5</td>
<td>.336</td>
<td>.168</td>
<td>-.067</td>
<td>.295</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Item-Total Statistics

<table>
<thead>
<tr>
<th></th>
<th>Scale Mean if Item Deleted</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
<th>Percent Participants Answering Correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall_1</td>
<td>1.62</td>
<td>1.596</td>
<td>.304</td>
<td>.181</td>
<td>.517</td>
<td>.380</td>
</tr>
<tr>
<td>Recall_2</td>
<td>1.60</td>
<td>1.532</td>
<td>.358</td>
<td>.153</td>
<td>.485</td>
<td>.420</td>
</tr>
<tr>
<td>Recall_3</td>
<td>1.58</td>
<td>1.603</td>
<td>.290</td>
<td>.285</td>
<td>.525</td>
<td>.400</td>
</tr>
<tr>
<td>Recall_4</td>
<td>1.63</td>
<td>1.524</td>
<td>.376</td>
<td>.291</td>
<td>.475</td>
<td>.300</td>
</tr>
<tr>
<td>Recall_5</td>
<td>1.63</td>
<td>1.626</td>
<td>.283</td>
<td>.255</td>
<td>.529</td>
<td>.370</td>
</tr>
</tbody>
</table>
Appendix E: Application of CVS, Pilot and Primary Study

1. In this experiment on the table, can you tell how the type of ball affects the distance the ball rolls?
   a. Yes. All other variables are the same and only the ball types are different between the two ramps
   b. No. More than one variable is different in the two ramps
   c. Yes. The ball is rolling down the ramp
   d. No. There are no numbers involved with this experiment

2. Which variable is the independent variable?
   a. Steepness of the ramps
   b. Surface type on the ramp
   c. Length of the ramp
   d. Type of ball

3. In this experiment on the table, can you tell how the surface type affects the distance the ball rolls?
   a. Yes. All other variables are the same and only the ball types are different between the two ramps
   b. No. More than one variable is different in the two ramps
   c. Yes. The ball is rolling down the ramp
   d. No. There are no numbers involved with this experiment

4. Which variable is the independent variable?
   a. Steepness of the ramps
   b. Surface type on the ramp
   c. Length of the ramp
   d. Type of ball

5. A student wants to learn if potassium is needed in soil for plant growth. Which variable will be the independent variable in this experiment?
   a. Amount of water
   b. Amount of potassium
   c. Amount of sunlight
   d. Amount of plant growth
6. A student wants to learn if potassium is needed in soil for plant growth. Which variable will be the dependent variable in this experiment?
   a. Amount of water
   b. Amount of potassium
   c. Amount of sunlight
   d. Amount of plant growth

7. A student wants to learn if potassium is needed in soil for plant growth. Which variable will be a variable that needs to be controlled so it stays the same in this experiment?
   a. Amount of water
   b. Amount of potassium
   c. The table where the student writes their observations
   d. Amount of plant growth

8. Students are planning an experiment to determine if bean seeds can grow without light. The independent variable for this experiment is
   a. Amount of soil in the pots
   b. Amount of light given to the plants
   c. Amount of water given to the plants
   d. Amount of growth in the plant

9. Students are planning an experiment to determine if bean seeds can grow without light. The dependent variable for this experiment is
   a. Amount of soil in the pots
   b. Amount of light given to the plants
   c. Amount of water given to the plants
   d. Amount of plant growth

10. Students are planning an experiment to determine if bean seeds can grow without light. An example of a variable that needs controlled so it stays the same for this experiment is
    a. The table the student records their observations
    b. Amount of light given to the plants
    c. Amount of water given to the plants
    d. Amount of plant growth
Appendix Q: Item Analysis-Application

Sample Items:
1. In this experiment on the table, can you tell how the surface type affects the distance the ball rolls?
   a. Yes. All other variables are the same and only the ball types are different between the two ramps
   b. No. More than one variable is different in the two ramps
   c. Yes. The ball is rolling down the ramp
   d. No. There are no numbers involved with this experiment

2. A student wants to learn if potassium is needed in soil for plant growth. Which variable will be the independent variable in this experiment?
   a. Amount of water
   b. Amount of potassium
   c. Amount of sunlight
   d. Amount of plant growth

Inter-Item Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>App1</th>
<th>App2</th>
<th>App3</th>
<th>App4</th>
<th>App5</th>
<th>App6</th>
<th>App7</th>
<th>App8</th>
<th>App9</th>
<th>App10</th>
</tr>
</thead>
<tbody>
<tr>
<td>App1</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App2</td>
<td>.408</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App3</td>
<td>.313</td>
<td>.299</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App4</td>
<td>-.022</td>
<td>-.098</td>
<td>.402</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App5</td>
<td>-.258</td>
<td>.045</td>
<td>.209</td>
<td>.197</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App6</td>
<td>.102</td>
<td>.205</td>
<td>.103</td>
<td>.299</td>
<td>.294</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App7</td>
<td>.206</td>
<td>.084</td>
<td>.269</td>
<td>.220</td>
<td>.053</td>
<td>.336</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App8</td>
<td>.015</td>
<td>.058</td>
<td>.029</td>
<td>.105</td>
<td>.125</td>
<td>.144</td>
<td>.416</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>App9</td>
<td>-.037</td>
<td>-.045</td>
<td>-.006</td>
<td>-.130</td>
<td>.303</td>
<td>.045</td>
<td>.230</td>
<td>.421</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>App10</td>
<td>.066</td>
<td>.161</td>
<td>-.070</td>
<td>.137</td>
<td>.017</td>
<td>.242</td>
<td>.039</td>
<td>.155</td>
<td>.323</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Item-Total Statistics

<table>
<thead>
<tr>
<th></th>
<th>Scale Mean if Item Deleted</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
<th>Percent Participants Answered Correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>App_1</td>
<td>4.77</td>
<td>5.029</td>
<td>.177</td>
<td>.343</td>
<td>.624</td>
<td>.700</td>
</tr>
<tr>
<td>App_2</td>
<td>4.93</td>
<td>4.809</td>
<td>.248</td>
<td>.350</td>
<td>.610</td>
<td>.500</td>
</tr>
<tr>
<td>App_3</td>
<td>4.95</td>
<td>4.591</td>
<td>.354</td>
<td>.432</td>
<td>.586</td>
<td>.500</td>
</tr>
<tr>
<td>App_4</td>
<td>4.98</td>
<td>4.796</td>
<td>.253</td>
<td>.419</td>
<td>.609</td>
<td>.500</td>
</tr>
<tr>
<td>App_5</td>
<td>5.05</td>
<td>4.862</td>
<td>.228</td>
<td>.373</td>
<td>.614</td>
<td>.420</td>
</tr>
<tr>
<td>App_6</td>
<td>5.00</td>
<td>4.475</td>
<td>.414</td>
<td>.337</td>
<td>.571</td>
<td>.500</td>
</tr>
<tr>
<td>App_7</td>
<td>4.82</td>
<td>4.491</td>
<td>.435</td>
<td>.360</td>
<td>.568</td>
<td>.700</td>
</tr>
<tr>
<td>App_8</td>
<td>4.90</td>
<td>4.634</td>
<td>.337</td>
<td>.314</td>
<td>.590</td>
<td>.570</td>
</tr>
<tr>
<td>App_9</td>
<td>4.88</td>
<td>4.817</td>
<td>.250</td>
<td>.443</td>
<td>.609</td>
<td>.600</td>
</tr>
<tr>
<td>App_10</td>
<td>4.92</td>
<td>4.823</td>
<td>.242</td>
<td>.292</td>
<td>.611</td>
<td>.550</td>
</tr>
</tbody>
</table>
Appendix R: Evaluation of Science Experiments, Primary Study

Algae Bloom

<table>
<thead>
<tr>
<th>Question: How does fertilizer effect algae?</th>
</tr>
</thead>
</table>

| Hypothesis: The more fertilizer there is the more the algae will grow |

| Materials: 8 4 oz jars Fertilizer Water Measuring cup Algae |

<table>
<thead>
<tr>
<th>Procedure:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fill each jar with water</td>
</tr>
<tr>
<td>2. Add fertilizer to each jar. In 2 jars don’t add any In 2 jars add 1 tablespoon In 2 jars add 2 tablespoons In 2 jars add 3 tablespoons</td>
</tr>
<tr>
<td>3. Add the algae to each jar</td>
</tr>
<tr>
<td>4. Label each jar</td>
</tr>
<tr>
<td>5. Put each jar into sunny windows</td>
</tr>
</tbody>
</table>

| Results: The jar with 2 grew good but not the best. The jar with none kinda grew. The jar with 1 tablespoons grew the best. |

| Conclusion: Fertilizer helps algae grow but not too much. |

1. What is the independent variable?  
   a. The amount of algae that grew  
   b. **The amount of fertilizer**  

The water in the jar  
The sunny window

2. What was the dependent variable?  
   a. The measuring cups  
   b. The different amounts of fertilizer  
   c. **How much algae grew**  
   d. Water used in the jars

3. Did they control for any other variables? If so, what were they?  
   **Yes. They used the same size jar and put all jars in the same window**  
   No. They did not need to keep variables controlled  
   Yes. They used different amounts of fertilizer  
   d. No. The conclusion does not mention any controlled variables

4. Were there other variables that should be controlled?  
   a. Controlling variables in this experiment is not important  
   b. They should have added 4 tablespoons of fertilizer to the jar  

They should make sure to add the same amount of water to each jar
Appendix S: Evaluation Item Analysis

Sample items:
1. What is the independent variable?
   a. The amount of algae that grew
   b. The amount of fertilizer
   c. The water in the jar
   d. The sunny window

2. What was the dependent variable?
   a. The measuring cups
   b. The different amounts of fertilizer
   c. How much algae grew
   d. Water used in the jars

Inter-Item Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Eval_1</th>
<th>Eval_2</th>
<th>Eval_3</th>
<th>Eval_4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eval_1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eval_2</td>
<td>-.068</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eval_3</td>
<td>-.365</td>
<td>.236</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Eval_4</td>
<td>-.233</td>
<td>.100</td>
<td>.735</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Item-Total Statistics

<table>
<thead>
<tr>
<th></th>
<th>Scale Mean if Item Deleted</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
<th>Percent Participants Answering Correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eval_1</td>
<td>1.48</td>
<td>1.305</td>
<td>-.294</td>
<td>.137</td>
<td>.625</td>
<td>.500</td>
</tr>
<tr>
<td>Eval_2</td>
<td>1.48</td>
<td>.830</td>
<td>.148</td>
<td>.068</td>
<td>.125</td>
<td>.450</td>
</tr>
<tr>
<td>Eval_3</td>
<td>1.53</td>
<td>.660</td>
<td>.376</td>
<td>.604</td>
<td>-.227^a</td>
<td>.400</td>
</tr>
<tr>
<td>Eval_4</td>
<td>1.50</td>
<td>.661</td>
<td>.372</td>
<td>.547</td>
<td>.430</td>
<td></td>
</tr>
</tbody>
</table>
Appendix T: Introduction Session Two Script, Primary Study

**Introduction:**
"Hello! My name is Cari and I want to thank you for being here today. What’s your name? Are you having fun today at camp? What is your favorite thing about camp? Today we are going to learn about designing science experiments. I am going to tell you about how we design experiments and then allow you to design some. Then you will answer some questions on your own."

**Training Session:**
“Today we will be investigating what affects the distance the ball rolls with this ramp. Can you describe the ramp to me? Can you tell me about a part that moves? How does it move?”

(Ask these questions until all variables are discussed: the type of ball used, the steepness of the ramp, and the distance of the ramp.)

“You have moved the ramp in all the different ways. Do you have any questions about the ramp?”
Appendix U: Analysis of Pretest

*Tests of Between-Subjects Effects*
Dependent Variable: Pretest

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>10.433</td>
<td>2</td>
<td>5.217</td>
<td>2.907</td>
<td>.063</td>
</tr>
<tr>
<td>Error</td>
<td>102.300</td>
<td>57</td>
<td>1.795</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>112.733</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix V: Analysis of Time

Tests of Between-Subjects Effects
Dependent Variable: Total session time

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>2557.433</td>
<td>2</td>
<td>1278.717</td>
<td>1.264</td>
<td>.290</td>
</tr>
<tr>
<td>Error</td>
<td>57685.500</td>
<td>57</td>
<td>1012.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>60242.933</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects
Dependent Variable: Time for introduction through instruction

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>569.100</td>
<td>2</td>
<td>284.550</td>
<td>1.445</td>
<td>.244</td>
</tr>
<tr>
<td>Error</td>
<td>11221.750</td>
<td>57</td>
<td>196.873</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11790.850</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects
Dependent Variable: Time completing outcomes

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>715.833</td>
<td>2</td>
<td>357.917</td>
<td>.506</td>
<td>.605</td>
</tr>
<tr>
<td>Error</td>
<td>40298.750</td>
<td>57</td>
<td>706.996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>41014.583</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix W: Analysis of Designed Experiments

*Tests of Between-Subjects Effects*
Dependent Variable: Total number experiments designed

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>1.433</td>
<td>2</td>
<td>.717</td>
<td>.417</td>
<td>.661</td>
</tr>
<tr>
<td>Error</td>
<td>97.900</td>
<td>57</td>
<td>1.718</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>99.333</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Tests of Between-Subjects Effects*
Dependent Variable: Percentage of correctly designed experiments

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>11056.300</td>
<td>2</td>
<td>5528.150</td>
<td>17.773</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Error</td>
<td>17729.350</td>
<td>57</td>
<td>311.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>28785.650</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix X: Analysis of Learning Outcomes

Tests of Between-Subjects Effects
Dependent Variable: Recall

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>25.433</td>
<td>2</td>
<td>12.717</td>
<td>14.748</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Error</td>
<td>49.150</td>
<td>57</td>
<td>.862</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>74.583</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects
Dependent Variable: Application

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>34.533</td>
<td>2</td>
<td>17.267</td>
<td>3.321</td>
<td>.043</td>
</tr>
<tr>
<td>Error</td>
<td>296.400</td>
<td>57</td>
<td>5.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>330.933</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects
Dependent Variable: Evaluation

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>9.300</td>
<td>2</td>
<td>4.650</td>
<td>5.528</td>
<td>.006</td>
</tr>
<tr>
<td>Error</td>
<td>47.950</td>
<td>57</td>
<td>.841</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>57.250</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix Y: Analysis of Changes in Science Self-Efficacy

**Tests of Between-Subjects Effects**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>3420.804</td>
<td>1</td>
<td>3420.804</td>
<td>550.598</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Error</td>
<td>54.706</td>
<td>2</td>
<td>27.353</td>
<td>4.403</td>
<td>.017</td>
</tr>
<tr>
<td>Total</td>
<td>354.135</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tests of Within-Subjects Effects**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSAI</td>
<td>4.524</td>
<td>1</td>
<td>4.524</td>
<td>21.747</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CSAI*Condition</td>
<td>13.963</td>
<td>2</td>
<td>6.982</td>
<td>33.560</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Error(CSAI)</td>
<td>11.858</td>
<td>57</td>
<td>.208</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>