5-2018

Formal Learning in an Informal Setting: The Cognitive and Affective Impacts of Visiting a Science Center during a School Field Trip

Charlie Belin
University of Arkansas, Fayetteville

Recommended Citation

Belin, Charlie, "Formal Learning in an Informal Setting: The Cognitive and Affective Impacts of Visiting a Science Center during a School Field Trip" (2018). *Theses and Dissertations*. 2701.
http://scholarworks.uark.edu/etd/2701

Formal Learning in an Informal Setting: The Cognitive and Affective Impacts of Visiting a Science Center during a School Field Trip

A dissertation in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Education Policy

by

Charlie M. Belin
University of Arkansas, Monticello
Bachelor of Science in Natural Science, 2004
University of Arkansas, Monticello
Bachelor of Arts in Communications, 2004
Texas Wesleyan University
Master of Education in Science Education, 2010

May 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

William F. McComas, Ph.D.
Dissertation Co-Director

Jay P. Greene, Ph.D.
Dissertation Co-Director

Gary W. Ritter, Ph.D.
Committee Member

Abstract

With the pressures of budget cuts, many schools--particularly schools serving low-income students--are eliminating school field trips to museums, zoos, and other cultural institutions despite their widely reported utility. First, through a systematic review of the literature, I examine the findings from 17 international studies on the benefits of visiting an informal science education institution during a K-12 school field trip. Almost all pre-post studies reported a positive change in both cognitive and affective outcomes after visiting an ISEI. However, studies that also included a control group to compare students who visited the ISEI with students who had yet to visit the ISEI reported mixed findings, and only three of those studies used randomization in placing students into the control and treatment groups. Second, I use a random assignment experimental design to study the impacts of visiting a science center during a school field trip on student interest in studying science, interest in visiting science centers, and knowledge obtained from attending an educational program as part of the science center experience. Survey data from 1,830 third through eighth graders showed small positive results suggesting that science museums encourage students to become connoisseurs[1] of science, and this effect was slightly greater for minority students, boys, and first time visitors. Also, short science center educational programs increased boys', first time visitors', and minority students' knowledge of science concepts found on state standards but did not benefit the average student visitor.

---

[1] Expression coined by McComas (2010) to describe encouraging student interest in science from an avocational perspective as one of the goals of science instruction.

Acknowledgements

This venture would not have been possible without the help of so many friends, colleagues, family members, and teachers. I would first like to thank Dr. William McComas for taking me under his wing and showing me how to be a stronger advocate for science education through policy. I would also like to thank Dr. Jay Greene who continuously provides a different and much needed perspective to education policy that I have been unable to find in our mainstream education system and Dr. Gary Ritter whose passion in the classroom has reignited my own passions for teaching. Crossing the finish line would not have been possible without the continuous and encouraging support of Dr. Dirk van Raemdonck. He has patiently and kindly helped me every step of the way. Next and equally as important, I would like to thank my other education policy professors. The University of Arkansas has such an amazing Education Reform Department and that program would not excel in such a way as it has over the last ten years without the passion and contributions of Drs. Robert Costrell, Patrick Wolf and Robert Maranto. Although Dr. Sandra Stotsky has returned to her home state of Massachusetts, she helped refine my overall education philosophy about the types of teachers needed to improve our current school system. I would also like to thank Dr. Brian Kisida, my colleague and partner in this study. Besides these great mentors, I want to thank my family, specifically my parents, Tom and Betty Belin, my husband, Jonathan Ruiz, and my sons, Jax, Nathan, and Ian for their sacrifices so I could finish. I also am deeply grateful for my children’s Nonnie, Maricarmen Ruiz, who has kept my toddler entertained and busy so I could have peace of mind. Lastly, I would like to thank two dear friends, Ione Stavron, for her encouragement and kindness, and to one of my oldest friends, Otto, for inspiring me to dream and always reminding me that dreams require action because we cannot idly sit by on wishful thinking.

Dedication

For Jax

Table of Contents

List of Tables and Figures

Chapter One: Why another Study about School Field Trips to Science Centers

School field trips are so much more than just a free day from school where students socialize with friends and play recess-type games. Out-of-school excursions provide enhanced learning and produce unique experiences not easily replicated inside the classroom, which can advance curriculum, offer real-world opportunities, give access to unfamiliar environments, teach responsible citizenship, improve critical thinking, and increase empathy and tolerance (Greene, Kisida, & Bowen, 2014a; Nabors, Edwards, & Murray, 2009). These benefits even carry into adulthood. Adults who experienced more hands-on activities during field trips often retain more information regarding the subject matter and found educational value in visiting museums, historical sites and zoos. Also, many adults report even revisiting former school field trips sites later in life (Pace & Tesi, 2004).

Yet with declining school budgets (Leachman, Masterson, and Figueroa, 2017), emphasis on standardized testing (Columbus Dispatch, 2012), and stricter field trip guidelines (Honeycutt Spears, 2014), museum and other off-campus expeditions are becoming limited are even non-existent for many students across the United States. After surveying superintendents across the country, the American Association of School Administrators reported that 37% of school districts would no longer provide field trips in the 2017-18 school year, of which 9% had actually eliminated field trips over five years ago (N. Ellerson, Associate Executive Director, personal communication, January 9, 2018). The New Jersey School Board of Association (2012) reported 41% of their respondents in 2012 saw a decline in field trips over the past three years, and districts were increasingly asking parents to pay the entire cost or some part of the trip. The Los Angeles Unified District reported their head count for field trips fell 56% between 2007-2008 and 2013-2014 even though the district enrollment fell just 6 percent, and in Long Beach

Unified their student head count for field trips fell 34% between 2002-2003 and 2013-2014 compared to 17% for their total student enrollment (Plummer, 2014). Field trip decline, particularly for low-income students, has inspired several legislators in Delaware to pass a field trip funding bill which would provide $25 a year for field trips for students in schools primarily serving low-income students (Ohlandt, 2017).

Unsurprisingly, the students who experience the greatest decrease in out-of-school excursions are those in heavily populated urban and low-income areas such as San Antonio, Texas; Miami-Dade County, Florida; and Los Angeles, California (Terrero, 2012). For many low-income and minority students (a population of interest in this study), school field trips likely provide the only opportunity for a holistic learning experience where they can see and handle real objects, access ideas and emotions not created in a school classroom, and uncover a passion for a future career or hobby. Many scientists have stated that the first draw towards a life-long career in science occurred because of a museum visit (Csikszenmihalyi & Hermanson, 1995). In an interview with 300 full-time science and engineering professors, a field trip experience was a top factor leading to their future career choice (Nazier, 1993).

With the troubling results reported from American students on international science assessments paired with the high demand for more science, technology, and engineering post-secondary workers; providing more visits to zoos, aquariums, natural history museums, and science centers should be a priority (McComas, 2006). Unfortunately for many policy-makers and school administrators, this need is not evident. Instead, they constantly weigh whether these excursions are worth the money and time away from important classroom instruction (Davidson, Passmore, & Anderson, 2010). Orion and Hofstein (1994) found that “the field trip is one of the most complex and expensive activities in the education system.” (p. 117), and Zoldosova and

Prokop (2006) questioned whether the educational benefits justify such a large investment. Not only do schools, districts, and students financially invest in museum learning, museums contribute over $2 billion a year to educational activities primarily aimed for K-12 students (American Alliance of Museums, 2014), even though little research exists that suggests a single field trip has a significant impact on student learning (Burchenal & Grohe, 2008).

Purpose of Study

The present study is designed to examine the differences in student attitudes (affective impact) and content knowledge (cognitive impact) between students who visited a science center as part of a school field trip and students who had not yet visited the science center as part of a school field trip. The study specifically focuses on (1) student interest in visiting science centers, (2) student interest in studying science both in their current school and as a possible future career, and (3) the amount of content learned after attending an educational program led by the science center staff during the school visit. Lastly, the study also evaluates the changes in attitude and content knowledge across several student populations of interest: white students, minority students, girls, boys, students who were a first time visitor to a science center, and students who had previously visited a science center.

Research questions. The study addresses three research questions:

1) After visiting a children's discovery science center during a school field-trip, how does <u>student's attitudes</u> regarding science centers and studying science differ from students who had not yet made such a visit?
2) After visiting a children's discovery science center during a school field-trip, how does <u>students' science knowledge</u> differ from students who had not yet made such a visit?

3) What overall impacts in attitudes or knowledge does the science center visit offer on certain subpopulations of students? Such populations include: white students, minority students, girls, boys, students who were a first time visitor to the science center, and students who had previously visited a science center before the study began.

Study site. This study was conducted using 3rd- 8th grade students who visited the Museum of Discovery in Little Rock, Arkansas during April, May, October, and November of 2012. The Museum of Discovery is a Smithsonian affiliated science center that houses five permanent galleries with over 90 hands-on exhibits. At the time of this study, the Museum of Discovery also provided 17 different educational programs for schools covering topics ranging from health to tinkering labs (Museum of Discovery, 2013). Three months prior to the study, the Museum of Discovery was renovated from a natural history museum to a science center, so throughout this discussion, I will refer to the *Museum* as a *science center* rather than a museum.

Significance of Study

This study advances the literature on informal school learning at science centers' by using a randomized-control experimental design with a large sample of survey respondents and as the first study to evaluate the cognitive and affective impacts on students who visit the Arkansas Museum of Discovery on a school field trip. Other studies have explored similar topics; particularly focusing on the affective and cognitive impacts of informal learning during school field trips, yet many of these researchers suggest that more rigorous research is needed. While evaluating changes in middle school student attitudes after students visited the Middle East Technical University's Science Center, Sentürk and Özdemir (2014) called for more studies that focus on both cognitive and affective variables. Also, several researchers have called for more rigorous and larger-scale evaluations (Andre, Durkson, & Volman, 2017; Dewitt & Storksdieck, 2008; Gutwill & Allen, 2012). Most of the research targeting the benefits of visiting science centers during school field trips is based on small sample sizes, usually a single class or classes from a single school. Students in these studies are rarely randomly sorted into a control and treatment group, and many studies do not even include a non-randomized control group. Most studies use either pre-/ post-assessment or interviews. To the best of my knowledge at the time of writing this publication, this is the largest randomized study examining the cognitive and affective benefits from vising a science center on a school field trip. A similar and larger study published by colleagues at the University of Arkansas evaluated the cognitive and affective benefits of visiting the Crystal Bridges Museum of American Art in Northwest Arkansas (Greene, Kisida, & Bowen, 2014b). Through a systematic review of the literature, I only found three studies from the ISEI field trip literature published since 2000 that incorporated a randomized control-treatment method that compared the difference between students who visited a science venue on a school field trip with students who have not yet visited. Holmes (2011) surveyed 228 6th grade student participants from a single school who visited a science center. Prokop, Tuncer, and Kvansičák (2007) surveyed 140 6th grade students from three schools who visited an outdoor park. Lastly, Itzek-Greulich, Flunger, Vollmer, Nagengast, Rehm, and Trautwein (2015) surveyed 770 9th grade students from 33 classes who visited an outreach science lab.

Brief Overview of Methods

This study examined if students experience an increase in knowledge after attending an educational program pre-selected by the teacher and taught by the science center staff and if students experience a change in attitudes about science while visiting the Museum of Discovery, a science center in Little Rock, Arkansas. I also explore how these experiences vary across different subpopulations of students, such as minority students and females.

To study the effects of the visit to the Museum of Discovery, I compared attitudes and knowledge derived from surveys (Appendix C) completed by students who had recently visited the science center (the treatment group) during a school field trip with attitudes and knowledge of students who had not recently visited the science center (the control group) during a school field trip. I developed the surveys with Brian Kisida, a research assistant at the University of Arkansas. The survey instrument is internally reliable (Cronbach's alpha = 0.85), was piloted at the science center before the study began, and reviewed by the educational director of the science center and three teachers. Unlike many other studies that focus on changes in student attitudes or content knowledge, I use randomized sampling techniques to place school groups of 3rd-8th grade students into either the treatment or control group. The treatment school groups were surveyed after visiting the center while the control student groups are those school groups who plan to visit but have not yet done so and completed the surveys before going. This increases the validity of the research design since the students are likely to be much the same in their overall nature and demographics. I administered surveys to 1,830 3rd-8th grade students who visited the center in April, May October, and November of 2012. Students in the control group completed surveys in their classrooms, administered by their teachers before having visited the science center on a school field trip. Students in the treatment group completed surveys also in their classrooms, administered by their teachers within two weeks after visiting the center on the school field trip.

Limitations, Delimitations, and Assumptions

Limitations. This study only evaluates the affective and cognitive impacts from a single science center. These findings, even with a large sample size, are recommended for generalization of the Museum of Discovery in Little Rock. This is particularly true of the cognitive domain because the survey was related to knowledge that might have been gained from programs and exhibits at this particular science center. Another limitation is about the type of school that chooses to visit a science center. Students surveyed primarily visited the Museum of Discovery from the central, more urban regions of the state or the northwest corner. Different outcomes may be measured for students from more rural areas and areas along the Mississippi River. Third, I only collected data over several months, a spring cohort and a fall cohort, so any findings are limited to the types of teachers and schools that self-select to visit this particular science center during these two times of the year. Many venues such as the Museum of Discovery experience an influx of students after state testing is over. Also in the final months of the school calendar, many students experience a decline in their learning motivation and academic achievement (Corpus, McClintic-Gilber, Hayenga, 2009). A fall visit to a science center may be more aligned with school curriculum and part of a larger instructional unit. Finally, although the treatment/control method should eliminate any potential bias from a specific group of students, there are several significant differences between the treatment and control groups such as grades and race.

Delimitations. Several choices I made regarding the design for this study are worth discussing. First, when deciding on the framework and methodology for both the Museum of Discovery study and the systematic review of previous studies on student learning during field trips, I originally planned to include Kindergarten through 2nd grade and 9th-12th grade. However, when piloting the survey at the MOD, several early elementary teachers said that they did not believe their students could complete the survey even if the survey was picture based and featured emoticons for the Likert-scale. One teacher said that for the students to understand she would likely have to help each student individually fill out the survey. Based on the teacher feedback, we decided not to include K-2nd grade students in the survey. I also did not include 9th-12th grade students in the study because only four high school groups made a reservation to visit the MOD during the study period. Other types of science centers may have larger participation of older students but for this venue, high school age students are a rare commodity.

Second, for the literature review in chapter two, I chose to conduct a systematic review instead of the more traditional review. In a traditional review, the author often selects literature that narrates the author's purpose for the research he or she is conducting. The empirical studies mentioned in a traditional review can be selected specifically to strengthen the author's perspective or to weaken the alternate perspective. A systematic review attempts to encompass all of the literature about a single topic that meets specific criteria that was pre-determined by the author. Not only does a systematic review help provide more overall conclusions from the literature but also provides specific details about the number of studies published about the topic, what countries the studies were conducted, the types of methodologies used, and the number of participants in each study. A systematic review follows a fairly specific process, and the final result is the collection of a group of articles that meet the inclusion criteria, and only those articles' results are tabulated in the overall findings of the review (Kowalczyk & Truluck, 2013). When I decided to research student benefits of visiting a science center on a field trip, there were already thousands of articles published on the topic. I wanted to know of those articles, how many were conducted in the last fifteen years, how many of the studies followed a randomized control/treatment design, how many students participated in each study, and what type of student participated. The best way to determine the answers to those questions was through a meticulous systematic review process. Ultimately, I also included a traditional literature review that was not based on a specific set of inclusion criteria f0or the first section of chapter two. These additional articles provide insight on field trips experiences by certain subpopulations. Most of the studies in the systematic review did not look at any subpopulations of interest such as minority students or girls. However, the final conclusions from chapter two are based entirely on the results of the studies from the systematic review.

Finally, for the study design, I chose to conduct a delayed-treatment RCT methodology rather than a baseline pre-post method. With the potential for such a large sample size, every student could have completed a pre-survey before the visit to the Museum and then a post-survey after the visit. I selected the RCT design instead of the pre-post method for two reasons. First, the majority of literature published after 2000 on field trip benefits utilizes a pre-post design. Very few informal science education institution (ISEI) field trip studies estimating the cognitive and affective impacts incorporate RCT methodology. An additional large-scale RCT study would help build the overall literature findings. Besides the limited number of RCT studies, I did not want to study just growth or change in a students' cognitive and affective domain after visiting a science center in comparison to students' attitudes and knowledge before the visit. Instead I wanted to determine the difference, if a difference even exists, between students who visited a science center with students who had not visited the science center. Over the same time period between the pre- and post-assessment, students who did not visit the science center could also show growth cognitively and potentially affectively just from being in the classroom. The RCT design eliminates this particular confounding variable. Also RCT designs are considered the gold-standard for program evaluations, and the methodological benefits of using an RCT will be discussed in greater detail in chapter 2.

Assumptions. First, estimates based on survey data require respondents to represent the population of interest, to answer the survey questions, and to answer the questions honestly (Murdoch et al., 2014). I will assume that students answered the survey items honestly and the cognitive questions to the best of their ability. Murdoch et al. (2014) found that people are more likely to answer self-reported questionnaires honestly if the questionnaire is anonymous rather than confidential. In a validation study about criminal behavior, Preisendörfer and Wolter (2014) concluded that female, better-educated, and older respondents were less likely to confess to committing a crime than their male, lesser-educated and younger respondents. Of course, in the study, the survey questions about attitudes about science learning are not as intimidating as when asking about someone's criminal history. Nonetheless, I am assuming the student participants are answering the questions honestly.

Second on the survey, I asked students if they had previously visited the Museum of Discovery. Seventy-percent of both the treatment and control students said they had visited the science center prior to the study. This study's primary premise relies on the difference between students who have visited the science center compared to students who had not yet visited the science center. Right before this study began, the Museum of Discovery just completed a major renovation from a collection-based natural history style museum to a more Exploratorium –style science center and had been closed for several months. So, although 70% of the comparison group had previously visited the science center, a second assumption is that the renovations create a completely new experience as if the students have never visited before.

A Preview of Chapter Two

Chapter two presents the literature background on informal science education institutions and school field trips. The literature review is divided into two parts. The first part follows a more traditional review and discusses the difficulties with measuring students who visit ISEIs during a school field trip; how schools, teachers and ISEIs partner together to align the goals of schools and teachers with the goals of ISEIs; current field trip policy and how that policy is playing out in today’s schools; how school field trips have the potential to increase diversity in STEM; and how informal learning impacts minority students, girls, boys, and first time visitors. The second part of chapter two discusses the steps taken and overall results of following a systematic approach to finding and then evaluating studies about student field trips to ISEIs. From this process, 17 studies were found that meet the inclusion criteria. These studies are divided into two categories based on if the researchers measured affective or cognitive outcomes, and then the findings from the studies are combined by category. The overall findings are discussed in detail.

Chapter Two: A Systematic Review on the Impacts of Visiting an ISEI by Students on a School Field Trip

The purpose of this chapter is to present the research on the impacts K-12th grade students receive after visiting an Informal Science Education (ISEI). The chapter is divided into two parts. The first part follows a more traditional literature review framework and is divided into three sections: (1) methodological difficulties when evaluating students who visit an informal learning environment during a school field trip, (2) interconnections between ISEIs, science curriculum, and school field trips, and (3) the impacts of school field trips on different types of students.

Besides the traditional review, I also include a systematic review as the second part of the literature review for several reasons. Both systematic reviews and traditional literature reviews are based on summarizing evidence; however, they vary significantly in approach. Systematic reviews use a systematic approach to critically appraise and synthesize research findings, while traditional literature reviews are more informal in nature and do not follow a standard scientific protocol (Kowalczyk & Truluck, 2013). The term *systematic review* originated in 1975 as a 'meta-analysis' by Gene Glass whose research focused on areas in public policy. Originally, systematic research methodologies expanded as a way to showcase 'evidence-based medicine.' Archie Cochrane's pioneering text, *Effectiveness and Efficiency'* published in 1972 pressed for a more rigorous way to compare health and medicine study results. This led to the formation of the Cochrane Collaboration in 1992, an international group of practitioners, researchers, and academics through Oxford University that review and combine health care research so that research findings are accessible to professionals and are quality assessed. Since the Cochrane Collaboration focuses on medical research, public policy researchers established the Campbell

Collaboration which is based on Cochrane methodology but applied to other policy agendas such as education, criminal justice, and social welfare (Oakley, Gough, Oliver, & Thomas, 2005).

There are not any specific guidelines for conducting a literature review, but the first step to a systematic review begins with problem formulation or statement of objective. After the first step, a systematic review follows a very structured set of procedures. First, the reviewers determine the quality standards and inclusion criteria for the review. Next, they begin collecting data from studies using an unbiased search of the literature. Once all the studies are collected, the reviewers evaluate the study design and determine if the methods meet the pre-determined quality standards. For the studies that do meet the quality standards, the study findings are combined, analyzed and interpreted either using qualitative or quantitative aggregation (Cooper, 1984).

Systematic reviews are often confused with meta-analysis. According to the 6th edition of Porta's *A Dictionary of Epidemiology* (2014), a systematic review is a "review of the scientific evidence which applies strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on the specific topic. Systematic reviews differ from traditional reviews, which tend to be mainly descriptive, do not involve a systematic search of the literature, and thus can suffer from selection bias" (p. 266) while a meta-analysis "is a statistical analysis of results…often performed on data located in a systematic review" (p. 184).

For this systematic review, I combine the cognitive and affective results from ISEI school field trip research that used either a pre-post survey or control trial design and met several other inclusion criteria. I compare studies across all types of ISEIs, but I do not include any meta-analytic averages of the effects of the studies found because the outcomes and methods are not similar enough, which I discuss in greater detail later in this chapter.

Part One: A Traditional Review of ISEI and School Field Trip Literature

Before discussing the benefits of visiting informal learning sites on various subpopulations of students, one major concern worth noting is the difficulties researchers experience when planning and conducting a study to an informal learning site. Also, science educators face several problems when trying to coordinate the learning goals of schools and teachers with the missions of informal learning sites. Both of those concerns are discussed in detail in the following two sections.

Difficulties with measuring informal learning impacts from a field trip. Researchers have published hundreds of studies investigating different elements of informal learning during school field trips to various science institutions around the world. The field trip literature catalogue includes studies that investigated long-term and short-term retention of information, change in student attitudes about science or a particular Informal Science Education Institution (ISEI), teacher education for planning field trips, ISEI program development, social interactions at the ISEI, gender or minority differences in student learning at an ISEI, student behavior at field-trip locations, curriculum modifications before, during, or after a field trip, virtual field-trips verses real field-trips, the use of mobile phones, cameras, worksheets, or probeware to enhance field trip learning, and differences in learning across subjects such as art compared to science. Even though the literature catalogue on field trip topics includes a variety of topics, the lion's share of the research focuses on benefits students receive when they visit an ISEI. These benefits are generally characterized as cognitive or affective (Dewitt & Storksdieck, 2008; Falk & Storksdieck, 2005; Griffin & Symington, 1997).

Although research on field trips is broad and abundant, little is known about actual learning, museum-school learning, and other learning outcomes students experience during a visit to an ISEI in the 21st century (Andre et al., 2017), in part because so many of these studies are based on a limited experimental design or weak statistic reporting (Zoldsova & Prokop, 2006). Leeming, Dwyer, Porter, and Cobern (1993) synthesized 34 studies that evaluated a change in content knowledge, attitude, or behavior after experiencing a classroom or informal learning treatment in environmental education. To draw any broad conclusions, they admitted to overlooking methodological imperfections within these studies.

Another reason for limited research on student learning at ISEIs is because the qualities and unique attributes of informal science learning venues make it difficult for researchers to evaluate field trip experiences in the same context as formal school learning. School assessments focus on cognitive impacts, particularly student demonstration of content attainment, while informal learning venues focus on increasing student interest and free-choice learning where students choose how they want to experience a venue (Sparks, 2011). Crane (1994) and Wellington (1990) believe it is almost impossible to accurately assess the cognitive and affective domains because of the complex nature of leaning inside a museum. Informal learning is so individualized and its impact cannot easily be evaluated using a letter grade and multiple-choice questions. Birney (1998) asks how one evaluates learning that is spontaneous, unguided, and inspires student discussion. Valuable informal learning experiences are so subtle that it often requires non-traditional methods of evaluation (Semper, 1990) because a single study cannot measure all the various elements of out-of-school learning, and different elements studied may require different approaches (Wellington, 1990). Most attempts to evaluate educational impact occur shortly after the visit and are conducted qualitatively through interviews and observations (Henry, 1992). Ramey-Gassert (1997) suggests rubric-based projects that combine classroom and museum learning is the best way to determine evidence of cognitive gains. Bamberger and Tai

(2008) call for a well-designed, large-scale study on field trip outcomes that could benefit museums and other informal institutions. Rennie and Johnston (2004) suggest:

> Visitors must be involved in the research process, not simply observed from a distance, because there is a sizable inferential gap between observing and interpreting. Seeing through the eyes of the visitor means that, at some state, data must be collected from the visitor and this requires self-report data, or recording what visitors both say and do (pg. S8).

Not only is research on informal learning difficult to measure and to align with both the goals of ISEIs and the goals of schools, researchers generally evaluate the impacts of ISEIs on students at a single venue, in a single grade, or in a single school or district (Kamarainen et al., 2013; Stavrova & Urhahne, 2010; Sweet, 2014). Since informal learning evaluations are often limited in scope, researchers often call for further studies similar to their own that branches out across different ISEIs and different grade levels (Holmes, 2011), use a different empirical strategy (Itzek-Greulich et al., 2015), evaluate different instructional techniques such as the use of museum educators, hand-held devices, or virtual fieldtrips to enhance the venue experience (Krombaß and Harms, 2008; Sweet, 2014), analyze long-term contributions on student attitudes, motivation, and learning (Stavrova & Urhahne, 2010), focus on other domains such as motivation or skills (Puhek, Perse, & Sorgo, 2012), or expand the research on cognitive and affective variables (Sentürk & Özdemir, 2014). Also, as ISEI directors incorporate new technology devices into their displays and exhibits, modify their curriculum based on new advances in science, and redesign the visitor’s experience based on a new mission or vision statement, researchers must continue evaluating the impacts ISEIs have on various student outcomes.

Cultivating field trips between schools, teachers, and ISEIs. The purpose and even definition of *school field trip* has evolved over the last century. Early literature defines school field trips as *school excursions*, *jaunts and journeys*, and *school journeys*, whereas the most recent literature has narrowed the definition to *educational trips*, *field observation visits*, *study tours*, or *educational tours* (Krepel & DuVall, 1981). Morag and Tal (2012) define a field trip as “arranged by schools, have educational purposes and take place in engaging and interactive settings” (p.746). Whitesell (2015) views field trips as a way “to strengthen students’ understanding of content, to expose them to broader educational settings, or to provide rewards” (p. 7). Behrendt and Franklin (2014) describe field trips as excursions that “take students to locations that are unique and cannot be duplicated in the classroom” (p. 236). With modern technology, Behrendt and Franklin’s definition would also include digital field trips. However, the National Research Council (2009) believes virtual field trips are a one-dimensional activity that does not allow students to use all of their senses and constricts the experience to only what the digital platform creator’s view important. Counter to virtual field trips, school field trips are “lived social events that become ways of knowing” (Scarce, 1997, p. 219).

In 2013, 151 science centers, science museums, and other related institutions in the United States reported serving 12.1 million school children, approximately 22% of the total private and public K-12 school population, through a school field trip (Association of Science-Technology Centers, 2014; US Department of Education, 2016). As of 2007, there were over 2500 ISEIs in the United States of various sizes and types. The majority partner with K-12 schools and offer a vast number of various programs for teachers, schools, and students (Dillon, 2007). The Association of Science Technology Centers (1996), the industry trade group for these organizations, defines an ISEI (also referred to as Informal Science Institutions, ISIs) as an institution whose primary purpose is to promote informal leaning in science. These institutions include science museums, botanical gardens, zoos, aquariums, arboretums, planetariums, natural history museums, nature centers, and science-rich children's museums.

Although science education venues are often described as places for informal learning, Eshach (2007) argues that for school field-trip purposes students are learning non-formally rather than informally or formally. In general, four characteristics (the process, location and setting, purpose, and content taught) distinguish whether learning is informal or formal, and both ISEIs and schools have a unique set of goals and purpose. Hodkinson, Colley, and Malcolm (2003) describe how these four characteristics differ across formal and informal learning environments. The process of learning in a formal environment is typically structured by a teacher and the outcome is assessed; whereas informal environments encourage spontaneity and self-exploration. Formal learning happens in schools while informal learning has no boundaries, no curriculum, no learning objectives, and no assessment. The purpose of formal learning focuses on the institution and the goals of others, while the purpose of informal learning is the actual learner and is initiated by the learner. Informal learning focuses on the development of something new via everyday practice while formal learning consists of increasing expert knowledge, practices, and understanding through acquisition of vertical knowledge. Schools today primarily focus on student growth or mastery of standards dictated by state policy and measured through standardized testing dictated by federal policy. ISEIs provide free roaming, non-structured environments where visitors are able to guide their own learning and social experience; however, the environment and experience is still controlled by the individual ISEI's goals and missions. This divergence in design is crucial to how both formal and informal settings contribute to educating the whole child. Eshach (2007) suggests that students only visit an ISEI during a school field trip occasionally and not necessarily by choice. When visiting an ISEI as part of a school field trip compared to a family outing, teachers often prepare and plan for the trip, and students tend to participate in structured activities similar to classroom activities such as attending a museum lab or completing a worksheet.

Other than the structured activities, motivation in ISEIs is often intrinsic where learning is self-directed, voluntary, and follows a ‘please touch’ approach (Gutwill & Allen, 2012). Informal sites usually include physical, social, and personal elements that not only focus on providing content but also stimulating an emotional and engaging hands-on experience. In a true informal experience, students not only have control over what they learn but their learning is not evaluated and graded (McComas, 2006). Informal education is often passed over as an area of learning (Ramey-Gassert, 1997). When informal learning does occur as part of the school curriculum, it usually means a 1-day trip to a science and technology center, natural history museum, zoo, aquarium, art museum, or botanical garden (Dierking, 1991). Even through a short field trip, informal learning “has many potential advantages: nurturing curiosity, improving motivation and attitudes, engaging the audience through participation and social interaction, and enrichment. By nurturing curiosity, the desire to learn can be enhanced,” (Ramey-Gassert, Walberg, & Walberg, 1994, p. 351).

Critiques, however, suggest that not all informal learning institutions are comparatively evaluated, and the limited research findings might misrepresent ‘learning’ for ‘fun.’ For example, researchers might observe students concentrating harder in science museums than in art museums simply because the students may have more fun in a science museum than an art museum, not because they are learning new scientific knowledge or even improving upon the scientific knowledge they already possess (Eshach, 2007). When comparing school students and students in family groups, school children behave, appear, and are even treated differently

(Griffin, 2004). Teachers place different constraints on school children for how they move through a museum and how they learn. The school group's needs prioritize over any individual student's needs. Kisiel (2005) identified eight motivational reasons teaches provided for why they embark on field trips and choose a particular venue: alignment with the curriculum, district or school expectations, learning opportunities, student motivation and interest, changing the learning environment, encouraging life-long learning, rewarding good behavior, and new experiences.

Informal education sites, particularly museums and science centers, offer a variety of different levels of experiences: highly structured programs such as classes and labs, moderately structured tours where students have the opportunity to interact with museum or center staff, or little structure where the teacher/school allows students to roam freely through the institution (Dewitt & Storksdieck, 2008). Typically, the highly structured classroom programs or science shows taught by a museum educator are optional for any field trip and these specialized programs transform the venue into a more formal setting which better aligns the experience with school or state curriculum. More structured programs or activities during the visit have been shown to maximize student learning and increase affective impacts by enhancing deeper engagement, encouraging adult-student interactions, and improving content learning (Dewitt & Storksdieck, 2008).

To advance structured informal learning and bolster the STEM education agenda, the National Science Board (NSB) (2007) recommends a partnership between schools, teachers, and informal learning environments. The National Research Council (2009) outlines six strands of what students should do and learn when visiting an ISEI. Students should:

> Strand 1: Experience excitement, interest, and motivation to learn about phenomena in the natural and physical world.

> Strand 2: Come to generate, understand, remember, and use concepts, explanations, arguments, models, and facts related to science.
>
> Strand 3: Manipulate, test, explore, predict, question, observe, and make sense of the natural and physical world.
>
> Strand 4: Reflect on science as a way of knowing; on processes, concepts, and institutions of science; and on their own process of learning about phenomena.
>
> Strand 5: Participate in scientific activities and learning practices with others, using scientific language and tools.
>
> Strand 6: Think about themselves as science learners and develop an identity as someone who knows about, uses, and sometimes contributes to science (p. 45).

Through these six learning goals, schools and ISEIs with very different purposes and missions can develop structured collaborations integrating science curriculum, offering mutually beneficial educational encounters, and eliminating school barriers such as cost and accountability that make it hard for teachers to completely exploit the infinite resources offered by ISEIs (Weinstein, Whitesell, and Schwartz, 2014). According to a survey of 475 United States ISEIs by the Centre for Informal Learning and Schools, 73% of ISEIs reported support for schools through programs, curriculum and materials, and workshops for students, teachers, and schools beyond offering a one-day field trip, and 53% reported that their programs were under capacity and could handle more participants. ISEIs in the study included natural history museums, science centers, aquaria, botanical gardens, arboreta, zoos, and planetaria (Phillips, Finkelstein, & Wever-Frerichs, 2007). Some ISEI venues report that teachers and students do not utilize many of the resources offered, and partnerships between schools and ISEIs are sporadic and depend highly on teacher motivation (Kisiel, 2010).

With or without partnerships between schools and ISEIs, people learn science in informal environments. The National Research Council (2009) cites an enormous collection of evidence that regardless of the type of environment (designed spaces, everyday experiences, or out-of-school programs); all age groups can learn science at these venues. Designed ISEIs such as zoos and science centers offer valuable, real-world encounters where anyone can broaden their scientific interests and participate in scientific inquiry. Unplanned everyday experiences offer individuals the chance to interact with nature and self-discover scientific processes embodied within the natural world; while structured, out-of-school science programs kindle science interests, impact student choices for future careers in science, and improve science achievement.

Out-of-school experiences that occur during school such as field trips or other science related extracurricular activities (science fairs, science clubs, and science competitions) can increase student enjoyment for learning science, confidence in performing science-related activities, and achievement on standardized assessments (Organization for Economic Co-operation and Development, 2012). Along with K-12 and higher education, informal education is one of the three integral pieces required to guarantee “U.S. economic competitiveness, particularly the future ability of the nation’s education institutions to produce citizens literate in STEM concepts and to produce future scientists, engineers, mathematicians, and technologists” (U.S. Department of Education, 2007, p. 5). Finally, informal environments are a crucial channel for increasing awareness, appreciation, and interest for science, technology, engineering, and mathematics (STEM) and must partner with formal environments to improve teacher development and bolster science curriculum (National Science Board, 2007).

I**ncreasing diversity in STEM through school field trips.** There is a growing concern that the United States does not have enough Information Technology (IT) or Science, Technology, Engineering, and Mathematics (STEM) workers. The U.S. Bureau of Labor Statistics (2014) projects STEM occupations will grow by more than 9 million jobs between

2012 and 2022 (about 13%) with a median income of $76,000 in 2013 (compared to $35,080 for all workers). Persistence towards a STEM career depends on many factors that include family and peer attitudes, career awareness, out-of-school experiences to informal learning venues, academic course rigor, and a student’s personal interest (Dorsen, Carlson, & Goodyear, 2006; Madill, Ciccocioppo, Stewin, Armour, & Montgomerie, 2004; Cleaves 2005). Some scientists have suggested that a visit to a zoo or natural history museum at an early age had a strong influence on their decision to pursue a career in the sciences (Csikszenmihalyi & Hermanson, 1995; Nazier, 1993). Sixty-five percent of scientists interviewed by Maltese and Tai (2010) reported their interest in science began before middle school. Also, students in middle school who conveyed an interest in science were three times more likely to earn a post-secondary degree in the sciences than middle school students that did not convey an interest (Tai, Liu, Maltese, & Fan, 2006).

The NSB’s (2016) Science and Engineering Indicators reported that high school female student achievement in math and science is similar to that of male students, and that females were just as likely to enroll in advanced math and science courses as their male counterparts, except for engineering and computer science. Black, Hispanic, and American Indian/Alaskan Native students or those students from lower income families experience the largest gap on standardized mathematics and science tests. Also students from lower income families or with less educated parents were less likely to enroll in level-1 science courses, while sex, race, and ethnicity did not impact student enrollment at this level. Females were slightly more likely to enroll in advanced science courses than males. Only 7% of males and 4% of females enrolled in computer science courses and only 3% of males and 1% of females enrolled in engineering. Although a disparity exists between males and females, this number is low across the board.

Although at the secondary level, males and females are comparable in achievement and course selection with some disparity between income and race, a gender disparity emerges at the post-secondary level, particularly for minority women (NSB, 2016). Women receive slightly more than half of degrees in the biological sciences, 17.9% of computer science degrees, 19.3% of engineering degrees, and 39% physical science degrees. Minority women only received 11.2% of science and engineering bachelor's degrees, 8.2% of science and engineering masters' degrees, and 4.1% of science and engineering doctoral degrees, yet they represent 18% of the entire US population (Kirby, 2012). Some evidence suggests that visiting an ISEI during a school field trip has a positive impact on minority, low-income, and female students. The next section explores the ISEI school field trip literature for these groups but also for students who have never visited a particular ISEI before.

Informal learning and minority students. Although research has shown that those of all ages can learn science in informal environments, survey and polling evidence suggests that not all groups take advantage of the experiences offered by ISEIs. At all levels of the ISEI network, efforts are ongoing to increase the presence of diverse groups, but these efforts have made little progress (National Research Council, 2009). For many English Language Learners, the zoo or museum field trip may be their first or only visit to these types of venues, and for the special education population, field trips offer a variety of unique and authentic experiences not found at home and in the classroom (Melber, 2008). Based on data collected by the Department of Education's National Center for Education Statistics (NCES) for the Early Childhood Longitudinal Study, Swan (2014) reported that many kindergarten children living in households of low social-economic-status (SES) or of Hispanic or Latino ethnicity were less likely to visit informal learning institutions than the higher SES children or non-Hispanic children. After controlling for income, race, and parental education, Swan concludes that children who do visit an informal institution during kindergarten had higher scores in math, science, and reading in third grade than those children who did not.

Besides primary and early elementary Hispanic students, junior and high school African-American students have also shown increased content knowledge, critical thinking, general interest in science, and an interest to study agricultural science in college after attending an annual science field trip from seventh grade to twelfth grade. The annual field trips include visits to a dairy farm (7th grade), animal science complex (8th grade), school of natural resources (9th grade), arboretum (10th grade), agricultural engineering center (11th grade), and Laboratory Science Center (12th grade) (Jones, 1997). In New York City, middle schools had the option to participate in the Urban Advantage (UA) program where schools partnered with eight ISEIs (New York Aquarium, Wildlife Conservation Society's Bronx Zoo, New York Botanical Garden, American Museum of Natural History, New York Hall of Science, Queens Botanical Garden, Brooklyn Botanic Garden, and the Staten Island Zoological Society) for free teacher professional development, lab kits, and funding for field trips. The UA program had the greatest impact on African-American students who slightly outperformed other African-American students who did not attend a UA school on the New York 8th grade science ILS exam. For African-American students, this achievement trend continued at least through the 9th grade year, even though the students were no longer attending a UA middle school. African-American UA students slightly outperformed the non-UA African-American students on the 9th grade NY Regent exam. The treatment effect was marginally higher on the 9th grade exam (0.086 standard deviations) than the 8th grade exam (0.080 standard deviations). The UA program did not have any significant impacts for Hispanic students, and poor students only experienced a very small impact on the 9th grade Regents exam (Weinstein, Whitesell, and Schwartz, 2014).

Similar to Swan's (2014) study, Martinello and Kromer (1990) found that low-income Hispanic students who received a six-week instructional unit on ecology including a field trip to an ecology exhibit had more and/or better inferences than students who did not experience the program or a shorter version of the program. Teachers in the six-week group reported strong student engagement, more student self-direction towards their learning, an increased level of student questioning, an ability to connect the lessons to their home environment by bringing artifacts from home, and that students would discuss the material for multiple days after the lessons were completed. However, Martinello and Kromer also compared low-income Hispanic students abilities for using descriptors, metaphors and supporting evidence and did not find any significant differences between students who receive the six-week or two-week program compared to students who did not receive any program at all.

Informal learning and gender. Fifty-two percent of female physicists and chemists reported that educational experiences such as camps, field trips, science competitions, and teacher demonstrations sparked their initial interest in science. Males, on the other hand, contributed their early interest to self-Initiated activities such as curiosity and playing with legos (Maltese & Tal, 2010). When visiting the Children's Museum in Boston, boys and girls race through the museum, are equally active, and share similar interests in model kitchens and model cars until they reach the first grade and then the girls start to favor the model kitchens and the boys prefer the model cars (Shapiro, 1990). Kremer and Mullins (1992) observed how different exhibits at the Center of Science and Industry in Ohio attracted different ratios of girls and boys. Boys were more interactive and spent more time at the water jets, bubbles, and build-a-house exhibits, while girls spent more time and were more interactive at the face painting and animal lab.

When a visit to an informal science center is paired with classroom instruction, the differences between girls and boys are minimal or non-existent. Martinello and Kromer (1990) evaluated whether a six-week or two-week ecology program produced any differences between 4th grade boy and girl low-income Hispanic students. They found that both girls and boys benefited equally from the program and that there were not any gender differences on students' abilities to use metaphors, supporting evidence, descriptors, or making inferences. A slight gender difference did emerge for 8th grade New York City students who attended an Urban Advantage (UA) school. Both UA males and females performed slightly higher (0.037 and 0.052 SDs, respectively) on the New York ILS exam than non-UA males and females. However, only males had any long-term marginal effects. Ninth grade males who were part of a UA middle school scored 0.072 SDs higher than non-UA 9th grade males on the 9th grade NY Regents exam (Weinstein et al., 2014).

Informal learning and 1st time visitors. At least since the 1970s, researchers have wondered about the benefits for students who are visiting an informal science venue for the first time. First-time visitors are overly excited about new opportunities to explore and are unable to focus on the educational material which could reduce any benefits (Dewitt & Storksdieck, 2008). Over several publications, Balling, Falk, and Martin have found that students did not retain any conceptual knowledge from an outdoor field-trip until the second experience with that particular setting (Balling & Falk, 1980; Falk, Martin, & Balling, 1978; Martin, Falk, and Balling, 1981). When the novelty effect was reduced on sixth grade students visiting the Pacific Science Center's Playground, students displayed more on-task exploration, had higher exploratory behavior scores, and increased cognitive learning than a similar group that did not receive the reduced novelty program. Researchers reduced novelty by showing students the day before the field trip a 15 minute slide presentation on what an actual visit to the Center would look like (Kubota & Olstad, 1991). Anderson and Lucas (1997) used a similar approach but expanded the 15 minute presentation to 40 minutes and had the presentation three days before the field trip instead of the day before. They also controlled for students who had been to the ISEI before and found that the greatest impact on post-test scores was by students who received the orientation and had been to the ISEI before. The novelty orientation also diminished any differences in cognitive learning between males and females.

Conclusion. The literature on informal learning, particularly to science education institutions, is vast. However, most of the previous research relies on weak methodologies, and many researchers argue that evaluating students' experiences at informal learning venues is difficult because of the unique characteristics of such institutions (Crane, 1994; Wellington, 1990). Also, as the definition of *school field trip* has evolved, many research findings from the 1980s and 1990s may be irrelevant to today's students, especially considering students can visit almost anywhere through a virtual experience. Field trips to educational institutions provide students an opportunity to enhance learning and offer experiences not easily replicated in the classroom. These experiences are not as formal as learning in a classroom, but they are also not completely informal, or spontaneous. Eshach (2007) describers school field trip experiences to ISEIs as non-formal. Teachers choose a particular field trip site for various reasons. Regardless of venue, teachers often plan and prepare for the trip, and the experience is often structured with some type of learning activity. Even with this type of structure, Eshach (2007) warns that observational researchers might construe learning for fun. Many ISEIs partner with teachers and schools to ensure their missions and exhibits are aligned with the goals of schools so that students can learn as well as have fun.

One of the purposes of visiting ISEIs is to encourage more students to pursue a career in science, technology, engineering, and mathematics. STEM jobs are in great demand and females and minorities are underrepresented. Some scientists have contributed a visit to an ISEI as a young child as one of the reasons they ultimately chose a career in science (Csikszenmihalyi & Hermanson, 1995; Nazier, 1993). Several studies reported that minority students who visit an informal education institution have higher scores in science, math, and reading (Swan, 2014; Weinstein et al., 2014), increased content knowledge, general interest, and critical thinking (Jones, 1997), better inferences, higher levels of student questioning, and can connect the experiences to their home environments than students who did not visit the ISEI (Martinello & Kramer, 1990). However, visiting an ISEI as part of a school field trip may not help increase the percentage of females in STEM career fields. Either both boys and girls benefit equally from the field trip experience (Martinello & Kramer, 1990) or boys have greater benefits (Weinstein et al, 2014). Besides females and minorities, ISEI researchers have also wondered about the impacts on first-time visitors. In some cases, the new experience is so overwhelming and exciting that students are unable to learn (Balling & Falk, 1980; Falk, Martin, & Balling, 1978; Martin, Falk, & Balling, 1981). To diminish the novelty effect, teachers are encouraged to introduce the site to students before the visit (Anderson & Lucas, 1997).

Part Two: A Systematic Review of ISEI and School Field Trip Literature

As stated earlier, the primary purposes of conducting a systematic review rather than a traditional literature review is to eliminate researcher bias in the selection of studies and to create a scientific protocol that other researchers could follow. In this section, I will briefly describe the systematic review methodology and discuss the overall findings of the studies included. Appendix D provides greater details on how the systematic review was conducted and the characteristics of the individual studies included in the review.

Inclusion criteria. I systematically compiled a list of current studies published in English that evaluated whether or not students benefit, cognitively or affectively, from visiting an ISEI as part of a school field-trip. In order for a study to be included, it must meet several criteria (see Table 2.1): the publication must have occurred after the year 2000, the study must have a comparison group either a control that did not visit the ISEI or a pre-test/pre-survey, the study participants must be in kindergarten through twelfth grade, the school trip must be a single-day field trip, the results must be quantitative with statistical significance reported, the study must have a sample size of at least 30 per treatment/control group, and the study must focus on outcome measures related to cognitive impacts and/or affective/attitudinal impacts.

Table 2.1

Inclusion Criteria Used to Select Studies for the Systematic Review

Review question	How do students benefit, affectively or cognitively, from visiting a science-orientated educational site during a school field trip?
Population(s)	Children in grades kindergarten-12th grade
Intervention(s)	Students visit an informal science education institution (ISEI) during a single school day as part of a school field trip and findings published between Jan 1st 2000 to May 9th 2016.
Comparator	Students who did not visit the ISEI site or students who took a pre-assessment/survey that can be compared with a post-assessment/survey
Outcomes	Change in students' understanding of a science concept or change in attitudes towards science or the ISEI.
Setting	Natural history museum, zoos, aquariums, science centers, or outdoor venue
Study design	A quantitative RCT, non-randomized control-treatment, or pre/post-test. Must have a minimum sample size of 30 students for each study/treatment group/control group
Analysis	Analysis must include information about whether the effect size was statistically significant at the 90% confidence level or higher.

The inclusion criteria were chosen based on the following premises: I am including only studies published after 2000 because education has changed tremendously over the last two decades. Students have Google, access to countless number of educational apps for their personal devices, virtual labs, and high stakes testing driven by a common standards movement. The dynamics of how students learned in the 1980s and 1990s may no longer apply to today's generation of classroom learners. Also, multiple literature reviews of the impact of science centers were published in the 1980s, 90s and early 2000s (Bitgood, 1989; Bitgood, Serrell, Thompson 1994; Blosser & Helgeson, 1986; Dierking, Burtnyk, Buchner, & Falk, 2002; Garnett, 2001; Griffin, 2004; Koran, Koran, & Ellis,1989; McComas, Cox-Peterson, Olson, & Narguizian, n.d.; Prather, 1989; Price & Hein, 1991; Ramey-Gassett,1997; Rennie & McClafferty, 1995; Rudmann , 1994), yet only a few were recently published, (Genaux-Hauser, 2010.; Behrendt & Franklin, 2014; Dewitt & Storksdieck, 2008; Eshach, 2007; Osborne, Simon, & Collins, 2003), and only one which used a systematic approach (Andre et al., 2017). Lastly, and perhaps most importantly, the earlier research failed to follow the methodological standards we impose today. Even Falk, one of the most recognized informal science learning experts, questions the validity and reliability of his own research (Falk, 2013).

Pre-post study design. In previous studies about field trips and informal learning, researchers use a pre- to post- test design. Students are first assessed prior to the field trip (i.e., the baseline measure) and then re-assessed using the same instrument after the field trip. The results are compared typically using t-test analysis (Ballouard, Provost, Barre, & Boneet., 2012; Jarvis & Pell, 2002; Futer, 2005). The primary problem with this design is that everyone received the intervention and went on the field trip and almost all studies have shown a positive change between the pre-assessment to post-assessment. Researchers would like to assume this positive gain is due to the field trip intervention but Trochim (2006) describes several potential alternative explanations that pose a serious threat to internal validity.

The first alternative explanation is the *history threat*. Perhaps the students like the television show Myth-Busters, and many students in the experiment watched an episode of Myth-Busters about science. The episode discussed similar concepts that were on display at the ISEI students recently visited. In this case, any positive effect shown could be due to watching Myth-Busters and not necessarily visiting the ISEI. The next possible explanation is the *maturity threat*. Students may have had the same exact outcome without ever going on the field trip. Over time, students mature and learn and this could have contributed to a small positive change in scores, especially if the classroom lessons are aligned with the field trip experience. The third major threat to the internal validity of a single group, pre- to post-test design is the actual *testing threat*. Students after taking a pre-test may have searched for answers to questions they did not know, or they realized that the pre-test and post-test was measuring their change in attitude. On the post-test, they were more likely to agree with statements about positive science attitudes because they recognized this was the assessment's purpose and not because their attitudes actually changed over time. The last major threat to internal validity is the *regression threat*, or the 'you can only go up from here' threat. Trochim (2006) suggests this threat particularly occurs when the pre-test scores are low, especially when compared to the larger population. Even without any treatment, the scores can only increase. If students are taking a four-answer multiple choice assessment, there is a 25% chance they answer the questions correctly simply due to guessing. So if a student scores lower than a 25% on the pre-assessment, they would have better odds of increasing their score on the post-assessment. For all of these reasons, the results of pre-post studies should be viewed with skepticism.

RCT study design. One common way to eliminate these specific threats is by using a two–group study design, where one group is the control population and the other group is the intervention group. The only difference between the intervention group and the control population should be the actual intervention (Trochim, 2006; Moorehead, n.d.). In this design, the control population would encounter all the same maturation and history threats, have the same issues with instrumentation and testing, and experience similar rates of regression to the mean. However, with a two-group design, two new threats to internal validity arise: *selection threats* and *social threats* (Trochim, 2006).

To contribute an outcome to a specific intervention, the control and treatment populations must be comparable before the study, and the intervention being the only difference. If the groups are not comparable before the study, this creates a *selection threat*. The only way to eliminate selection bias is to randomly assign people into the two groups. This design is called a "true" experiment, and an RCT design does not require a pre-assessment of both groups because the design assumes that randomization eliminates any differences between groups. Randomized control trials have several key features: (1) randomization into intervention groups, (2) blindness or unawareness of the specific treatment given until after the study is completed, (3) test subjects typically remain in their allocated group regardless of if they experienced the intended intervention (intent to treat), and (4) analysis centers around estimating the magnitude of the difference in outcome measures between the treatment and control population (Meldrum, 2000; Sibbald & Roland, 1998).

The main concern with other types of study designs, including non-randomized controlled trials, is that even if a correlation between an intervention and an outcome exists, a third factor associated with both the intervention and the outcome could be driving that correlation. By randomizing into intervention groups, any systematic differences or similarities that might influence the outcome should be eliminated (Sibbald & Rioland, 1998), for example, in an educational control study where schools that experience a tutorial program intervention are matched with schools that did not experience such intervention. After data is collected and analyzed, any outcome differences such as improved test scores are attributed to the tutorial intervention. Yet matched schools may be different in other areas such as teacher quality or student demographics that could impact the results (Torgerson & Torgerson; 2001).

Critics of RCT methodology argue these types of studies are often more time-consuming and costly than non-RCT studies (Sibbald & Rioland, 1998). However, RCT methodology produces “unquestionably precise” results (Meldrum, 2000, p.746), and these results can be easily translated beyond the experimental setting (Porter, 1995). A second critique is that the practice of RCTs is unethical because some participants do not receive the intervention. In many cases, RCT studies are trying to determine if a particular drug or method works because at the time of the study, the answer is unclear. Treating one group and not the other is important to understanding if the method or drug works; otherwise we are giving everyone a treatment that could in fact be harmful. Also, an RCT could be considered ethical but not feasible, particularly if program stake-holders want all participants to receive the treatment intervention (Torgerson & Torgerson; 2001). When researchers are unable to use randomization techniques to create the two groups, the design is considered quasi-experimental, and pre-assessment of both groups becomes necessary to determine if a difference between the groups exists. This is still different from a pre-post analysis, because a quasi-experiment includes a control group that does not receive the treatment. The control group is just not randomly selected (Harris et al., 2006).

Regardless of design, both experimental and quasi-experimental studies are subject to several social threats, which are by-products of social pressure from the actual study. The first threat, *diffusion or imitation*, occurs when the comparison group realizes another group is receiving the treatment and then tries to imitate the treatment population, which would minimize the program effect if there is one. The second threat, *compensatory rivalry*, happens when the comparison group realizes they are not getting the treatment and then creates a competitive attitude, which again could potentially minimize the program effect if one exists. The third threat, *resentful demoralization*, is like compensatory rivalry except instead of trying to compete with the treatment population the students give up, making the program effect size larger than it should be. Lastly, the *compensatory equalization of treatment threat* happens when managers of the groups, such as teachers or parents, place pressure on researchers and administrators to be in the other group. If parents or teachers discovered that half the students in the school is going on a field trip but the other half is not, they may place pressure for their students to be able to go on the field trip (Trochim, 2006).

Although studies using an RCT experimental design are often considered the highest quality studies even with some of the methodological threats, only a handful of RCT studies on the impacts students receive when visiting an ISEI on a school field trip exist. Since pre/post assessment study designs are more often used by informal education researchers, I expanded the inclusion criteria to also include pre/post assessments as well as control-treatment without randomization (quasi-experimental). However, I separate those results from the RCT results.

Database searches. To find studies that could potentially meet the inclusion criteria; I searched JSTOR, EBSCO, Web of Science, Pro Quest, Science Direct, and the literature section of research studies that were coded for the review. I searched through the titles and abstracts of 1,925 articles and studies. If it was apparent from the title (and abstract if easily provided) that the study did not meet one of the criteria, I noted the inclusion reason and eliminated the study (see Table D.3 in Appendix D). I reviewed the methodology sections of 101 studies. From those, I coded 38 studies to determine if the studies meet all the inclusion criteria. Only 17 studies meet the inclusion criteria. Table 2.2 provides the combined totals of the three searches.

Table 2.2

Merged Search Results from JSTOR, EBSCO, Web of Science, ProQuest, Science Direct, and Other Literature Reviews

	Total Studies	Duplicates	Total Unique Studies	To Review	To Code	Studies Included
1st Search "science" "field trip" "not graduate" not undergraduate" "schools" students"	252	26	226	40	15	8
2nd Search "field trip" "science" "education	1487	269	1218	35	8	2
Bibliographies	186		186	26	15	7
TOTALS	1925	295	1630	101	38	**17**

Cognitive and affective studies. Table 2.3 provides the basic publication information for each of the 17 articles included in the systematic review. This information includes the authors, year published, title of article, country of study, ISEI, sample size, student grade levels, researcher's purpose for study, and if the study measures affective or cognitive outcomes. Three studies used a randomized control treatment design, although one of those three studies did not estimate the treatment effect. Six studies used a quasi-experimental design by including a control group, although not randomized. The final eight studies were non-experimental and did not include a control group.

Table 2.3

General Information about Each Study Included in the Systematic Review

Authors and Year Published	Study Design	Title	Country and ISEI	Sample	Grade(s)	Purpose of Study (Determine if…)	Aff.	Cog.
Holmes (2011)	Experimental: Random Control Treatment	Informal learning: Student achievement and motivation in science through museum-based learning	United States Louisiana Tech IDEA Place	228	6th	an exhibit, a museum lab program, or both changes students' attitude and knowledge	0*	0*
Itzek-Greulich, Flunger, Vollmer, Nagengast, Rehm, & Trautwein (2015)		Effects of a science center outreach lab on school students' achievement: Are student lab visits needed when they teach what students can learn at school?	Germany Experimenta Research Center	770	9th	a classroom lab or a lab in a science center has greater positive effect on student understanding		0*
Prokop, Tuncer, & Kvansičák (2007)		Short-term effects of field programme on students' knowledge and attitude toward biology: A Slovak experience	Slovakia Outdoor Nature Areas	143	6th	outdoor educational program changes attitudes or increases knowledge	+	+
Ballouard, Provost, Barre, & Bonnet (2012)	Quasi-Experimental: Control, Not-Random	Influence of a field trip on the attitude of schoolchildren toward unpopular organisms: An experience with snakes	France Outdoor Nature Area	520	1st-5th	attitude changes after attending program on snakes	+	
Puhek, Perse, & Sorgo (2012)		Comparison between a real field trip and a virtual field trip in a nature preserve: Knowledge gained in biology and ecology	Slovenia Nature Reserve at Maribor Island	211	8th	knowledge increases after experiencing a real field trip or a similar virtual field trip		M
Sentürk & Özdemir (2014)		The effect of science centres on students' attitudes towards science	Turkey METU's Science Centre	251	6th-8th	attending a science show with lab-based demos changes student attitudes	+*	
Sturm & Bogner (2010)		Learning at workstations in two different environments: A museum and a classroom	Germany Natural History Museum	163	6th	experiencing either museum workstations or classroom workstations changes attitudes	0*	+*
Sweet (2014)		*The effectiveness of virtual and on-site dairy farm field trips to increase student knowledge in science, social studies, and health and wellness standards* (Master thesis)	United States Kelsay Dairy Farm	46	3rd	knowledge increases after experiencing a real field trip or a similar virtual field trip		+
Wilde & Urhahne (2008)		Museum learning: A study of motivation and learning achievement	Germany Natural History Museum of Berlin	207	5th	Different field trip structure (closed-ended tasks vs. open-ended tasks) increases knowledge		M
Basten, Meyer-Ahrens, Fries, & Wilde (2016)	Non Experimental: No Control	The effects of autonomy-supportive vs. controlling guidance on learners' motivational and cognitive achievement in a structured field trip	Germany Zoo	206	5th	different field trip structure (autonomy supported vs. controlling) increases student knowledge		+
Freedman (2010)		A "healthy pizza kitchen" nutrition education program at a children's health museum	United States Hall of Health	151	5th	students have better understanding about nutrition after experiencing program on pizza		+
Futer (2005)		*Evaluating the effectiveness of environmental education essential elements in school field trip programming* (Master Thesis)	Canada Montreal Biodome	338	4th-6th	visiting one of three different ecosystem programs changes students' attitude and increases knowledge	M	+
Jarvis & Pell (2002)		Effect of the Challenger experience on elementary children's attitudes to science	United Kingdom Challenger Space Center	655	5th	visiting the exhibit changes attitude and increases student knowledge of space science	M	
Jarvis & Pell (2005)		Factors influencing elementary school children's attitudes toward science before, during, and after a visit to the UK National Space Center	United Kingdom Challenger Space Center	293	5th	visiting the exhibit increases student knowledge of space science	M	M
Kamarainen, Metcalf, Grotzer, Browne, Mazzuca, Tutwiler, & Dede (2013)		EcoMOBILE: Integrating augmented reality and probeware with education field trips	United States Outdoor Nature Area	71	6th	working with probeware changes student attitude and increases student chemistry knowledge	0	+
KrombaB & Harms (2008)		Acquiring knowledge about biodiversity in a museum: Are worksheets effective?	Austria Inatura Natural History Museum	148	6th-9th	using worksheets on a field trip increases student knowledge on animal biodiversity		+
Stavrova & Urhahne (2010)		Modification of a school programme in the Deutsches Museum to enhance students' attitudes and understanding	Germany Deutsches Museum	96	8th-9th	different field trip structure (highly vs. less) increases student knowledge		+

Studies grouped together by study design. Aff=Affective Outcome Studies, Cog.=Cognitive Outcome Studies, + = positive, significant findings ($p<0.5$), M=mixed findings that are either positive or did not find a difference between groups at a 95% confidence level, 0=no difference between groups at 95% confidence level. All studies used pre/post analysis. *= effect size was calculated between treatment group and control group.

Types of informal science educations intuitions represented. Combined, the 17 studies represent five different types of ISEIs: those occurring (1) in the natural environment or at (2) natural history museums, (3) science centers, (4) outreach labs, or (5) zoos, aquariums, and animal farms. Five studies occurred at science centers, while four studies each occurred in the natural environment or at a natural history museum. Zoos, aquariums, and farms consisted of three studies. Lastly, one study occurred at a science outreach lab. Table 2.4 provides a breakdown of those studies by ISEI type.

Table 2.4

Types of ISEIs Represented in the 17 Studies

Natural Environment	Ballouard et al., 2012 Kamarainen et al., 2013 Prokop et al., 2007 Puhek et al., 2012
Natural History Museum	KrombaB and Harms, 2008 Stavrova and Urhahne, 2010 Sturm and Bogner, 2010 Wile and Urhahne, 2008
Science Center	Freedman, 2010 Holmes, 2011 Jarvis and Pell, 2002 Jarvis and Pell, 2005 Sentürk and Özdemir, 2014
Zoos, Aquariums, and Farms	Basten et al., 2016 Futer, 2005 Sweet, 2014
Outreach Lab	Itzek-Greulich et al., 2015

Overall affective findings from the systematic review. Of the nine studies that looked at affective outcomes, six found at least one statistically positive impact from visiting an ISEI during a school field trip, regardless of whether the study relied solely on a pre-post analysis or included a control group. Table 2.5 provides the different outcomes by study. Together, the 9 studies reported results for 51 different affective outcomes. However, two studies (Ballouard et al., 2012; Jarvis & Pell, 2002) reported differences for individual Likert-style items. Also, Ballouard and colleagues did not report results for items that were not significantly different at the 95% confidence level. Their survey had 61 questions although some questions asked about student demographics. They only reported the results for four individual items. If 50 of the 61 items on their survey were affective questions, then the total number of individual Likert-items analyzed would be approximately 76 for Ballouard et al. (2012) and Jarvis and Pell (2002). In this case, only 16% (12 of 76) of the individual item outcomes were positive and significant at the 95% confidence level.

Four studies reported findings for 17 outcomes which were all multiple-item analysis of a specific category such as *anxiety* or *learning science in school*. Only 59% (10 of 17) were positive and significant at the 95% confidence level between the pre- and post-assessment. Researchers reported greater attitudes towards biology as a school subject, greater attitudes towards the natural environment, greater attitudes towards future biology work (RCT study) (Prokop et al., 2007), better attitudes about learning science in school, greater self-concepts in school science, greater interest in pursuing science outside of school, greater interest in future science work, and stronger beliefs in the importance of science (control study) (Sentürk, & Özdemir, 2014), and decreased anxiety and greater interest in space science (non-control) (Jarvis & Pell, 2005).

Lastly, four studies, reported results for the entire survey. Of the six different student groups surveyed (one of the four studies had three treatment groups), three groups (50%) had a greater overall attitude about science at the 95% confidence level, while three groups did not. The only RCT study from these four studies (Holmes, 2011) did not report a significant difference ($p<0.05$) between her three combined treatment groups (which she analyzed together) and the control group, and one of the non-control studies (Futer, 2005) only found significant positive results for one group, those who visited a tropical forest. Students who visited a Laurentian Forest or the St. Lawrence Forest did not answer significantly different on the post-assessment than the pre-assessment.

Randomized control treatment studies**.** Of the two RCT studies, one study (Holmes, 2011) did not find a significant difference at the 95% confidence level while Prokop et al. (2007) reported positive findings when estimating the mean difference between the pre-and post-tests for the treatment group and control group separately. The control group was not statistically different at the 95% confidence level on the pre-assessment and post-assessment, while the treatment group was significantly different ($p<0.05$) on all three instruments. Prokop et al., (2007) concludes that since the control group did not change between pre and post-assessment while the treatment group did that the intervention of visiting the venue worked. However, since Prokop et al. (2007) did not actually compute the effect size of the treatment group compared to the control group, it is unclear if the two groups are significantly different at the 95% confidence level, and how large of an effect the field trip had on students' attitudes compared to students who did not take the field trip.

Besides using an RCT design, Holmes (2011) calculated an F-statistic comparing students who visited a science center with students who had not yet visited. She randomly divided the intervention students into a guided exhibit tour group, a 30-minute lab activity group, and a guided exhibit tour/30-minute lab activity group and then compared the groups to a control group. She did not find any significant difference at a 95% confidence level between the combined intervention groups with the control group on the pre-Children's Academic Intrinsic Motivation Inventory (CAIMI) and the post-CAIMI. However, students who attended only the

30-minute lab activity at the IDEA Place scored significantly higher (effect size=0.222 standard deviations, $p<0.050$) on the post-assessment of the CAIMI than their pre-assessment. This higher score could be a reflection of how awesome the program was or could be the excitement the students felt as they knew they were about to actually visit the science center. When the lab group arrived at the science center they immediately went to the work station room but knew that once they completed the activity and the post-surveys they would be free to roam around and explore the science center.

Subpopulation analysis. Two researchers studied how visiting a field trip impacted certain subpopulations of students. Sentürk & Özdemir (2014) compared different groups of students who visited the Middle East Technical University's Science Centre and did not find any significant differences ($p<0.05$) between the post-attitudes of boys and girls, by different grade levels, or by different science achievement levels. The ages compared were between 11 and 14 and the science achievement level groups were fail, passable, average, good, and excellent. Jarvis and Pell (2005) divided the classes who visited the Challenger Space Centre by race and community type and by how enthusiastic students were about science on the pre-affective survey. Although insignificant at the 95% confidence level, they found that suburban-mixed race classes and primarily Caucasian classes in an urban school had a slight increase in science enthusiasm after visiting the Challenger Space Centre. They also found that suburban-Asian and inner city-mixed race classes had a slight decrease in science enthusiasm after visiting the Challenger Space Centre. For the suburban-Asian classes, Jarvis and Pell (2005) note how high their pre-assessment scores were and that there is not much room for an increase in science enthusiasm, which they also cite for the slight decrease in pre-to post-scores for students who have a high science enthusiasm before visiting the Centre.

For students who are not very enthusiastic about science prior to the field trip, they reported a significantly large change in their enthusiasm after the trip, while students who are extremely unenthusiastic about science had a significantly negative change in attitude after visiting the Centre. Jarvis and Pell believed the more enthusiastic students reflected the enthusiasm of their teachers. Teachers who dislike visiting science centers had large numbers of students who also dislike visiting science centers, while teachers who are extremely excited about these kinds of field trips are better prepared and have large numbers of students who are also very excited about science. Jarvis and Pell did not include a control group and these results may vary for a different set of students.

Treatment effect. Lastly, three studies estimated the treatment effect for visiting a science center compared to not visiting the science center, and only one of the three studies found a significant yet small effect by students who visited the ISEI then students who did not. The only RCT study (Holmes, 2011) did not find a difference between students who visited the IDEA Place and students who had yet to visit. Her control group completed the post-assessment minutes before visiting the science center because they were attending the science center with the treatment groups. One non-random control/treatment study (Sentürk, & Özdemir, 2014) reported a moderate positive change in attitude towards science by students who visited the Middle East Technical University's Science Centre compared to students who did not visit. The students who did not visit stayed in the classroom and continued with regular instruction and visited the center after participating in the study. The other non-random control/treatment study (Strum & Bogner, 2010) did not find a significant difference at the 95% confidence level by students who visited a natural history museum compared to students who stayed in the classroom. The control group received the same lessons and classroom activities as the treatment group.

Table 2.5

Affective Impacts from Visiting an ISEI during a School Field Trip

Researcher/ Year Published	Type of Experience	Ages/ Grade	Pop. Size	Study Group	Survey	Effect pre	post	Sign
Experimental Studies: Randomized Control-Treatment								
Holmes, 2011	Louisiana Tech *IDEA Place*- Science Center-Physical Science. Researcher created activities.	6th	56C	Control Group vs Treatment Group***	Children's Academic Intrinsic Motivation Inventory (122 items)	**$F_{(3,224)}$=2.05**		
						91.1	94.1	+
			53 T	Guided Tour Group		90.2	92.6	+
			61 T	30 Minute Lab Activity		94.3	97.7	**+**
						ES=0.222		
			58 T	Guided Tour & Activity		88.5	89.9	+
Prokop et al., 2007	Local Area-Environmental- Experimental students had 4 activities/tasks to complete about ecosystems-Researcher created activities.	6th	69 C 74 T	Control Group	Attitudes towards Biology as a School Subject (4 items)	14.5	14.3	-
				Experimental Group		14.5	15.5	**+**
				Control Group	Attitudes toward the Natural Environment (3 items)	10.5	10.2	-
				Experimental Group		10.6	11.1	**+**
				Control Group	Attitudes towards Future Biology Work (5 items)	13.2	13.1	-
				Experimental Group		13.6	14.9	**+**
Quasi- Experimental Studies: Non-Randomized Control-Treatment								
Sturm & Bogner, 2010	Natural History Museum-bird biology-8 workstations-Researcher created program.	6th	117 C 46 T	Museum Group Vs Classroom Group (same lesson)***	Interest/Enjoyment	X^2 = 0.026		+
					Perceived Choice	X^2 = 0.291		+
					Value/Usefulness	X^2 = 02.897		+
					Perceived Competence	X^2=5.160		+
Sentürk & Özdemir, 2014	*Middle East Technical University's Science Centre*- Science Show about 12 exhibits by Centre Staff- Free Roaming with Guide Assistance. Center created demonstration.	Ages 11- 14	117 C 46 T	Control Group vs Intervention Group***	Learning Science in School	η^2=0.62		**+**
					Self-concept in school science	η^2=0.05		**+**
					Practical work in school science	η^2=0.03		+
					Science outside of the school	η^2=0.06		**+**
					Future Science Work	η^2=0.04		**+**
					Science Importance	η^2=0.16		**+**
					Attitudes Toward Science Scale (33 items)	η^2=0.35		**+**
			46 T	*Gender*		*η^2=0.01*		+
				Age		*η^2=0.01*		+
				Prior Science Know.		*H^2=0.02*		+
Ballouard et al., 2012	Outdoor nature park-Guided Activities about Snakes-Researcher created program.	Ages 6 – 11	48C	Control Group	4 item analysis-see below	No difference any item		
			472T	Intervention Group	Afraid of Snakes (1 item)	33%	11%	**+**
					Like Snakes (1 item)	42%	53%	**+**
					Important to Protect Snakes (1-tem)	77%	94%	**+**
					Priority to Protect Snakes (1-item)	31%	73%	**+**

Table 2.5(cont.)

Affective Impacts from Visiting an ISEI during a School Field Trip

Researcher/ Year Published	Type of Experience	Ages/ Grade	Pop. Size	Study Group	Survey	Effect		
						pre	post	Sign
Non-Experimental Studies: No Control								
Kamarainen et al., 2013	Local Pond-Environmental- Probeware-Activities- Researcher created program	6th	71	All Students	7 Question Self-Efficacy Survey	ES=0.48		+
Jarvis & Pell, 2002	*Challenger Space Center*- small group, staff led activities plus teacher facilitated play and focused science talk during free-roaming time. Center created programs	Ages 10 to 11	655	All Students	*Science Enthusiasm (8 items)*			
					I often do science experiments at home	1.76	1.85	+
					I should like to be a scientist	1.97	2.05	+
					School science clubs are a good idea	Not significant		
					I should like to be given a science kit as a present	Not significant		
					I like science more than any other school work	Not significant		
					I like to watch TV science programs	Not significant		
					I am always reading science stories	Not significant		
					One day I should like to go to the moon	Not significant		
					Social Context (7 items)			
					Science is good for everybody	Not significant		
					Our food is safer thanks to science	Not significant		
					Science can make chemicals we need from rock	2.32	2.54	+
					Science makes me think	Not significant		
					It is easy to find out new things in science lessons	Not significant		
					TV, telephones, and radio have all needed science	Not significant		
					Science has made us better and safer medicines	Not significant		
					Lots more money should be spent on science	Not significant		
					Space (10 items)			
					I would be scared to go on a trip to space	1.80	1.67	+
					Computers control all space research	2.51	2.62	+
					We need telescopes at school	1.87	2.03	+
					One day humans will settle on other planets	2.00	2.08	+
					I should like to know how a rocket works	2.83	2.81	-
					Space satellites have improved our lives	2.49	2.50	**0**
					We can learn a lot by exploring space	2.86	2.85	**0**
					I would like to make a model rocket	2.74	2.77	+
					Space scientists will keep our planet safe	2.59	2.66	+
					I would hate to try experiments in space	1.70	1.72	-

,

Table 2.5(cont.)

Affective Impacts from Visiting an ISEI during a School Field Trip

Researcher/ Year Published	Type of Experience	Ages/ Grade	Pop. Size	Study Group	Survey	Effect		
						pre	post	Sign
Futer, 2005	Montreal *Biodome*-Environmental Program-Biodome created tour.	4th-6th	151	Tropical Forest Group	Environmental Attitudes and Personal Responsibility (10 items)	t=2.11		**+**
			172	Laurentian Forest Group		t=-0.29		-
			68	St. Lawrence Group		t=-1.02		-
Jarvis & Pell, 2005	*Challenger Space Center*- small group, staff led activities plus teacher facilitated play and focused science talk during free-roaming time. Center created activities.	Ages 10-11	293	All Students	Anxiety	15.3	14.5	**+**
					Space Interest	19.6	20.0	**+**
					Social Context	20.0	20.3	+
					Science Enthusiasm	*15.4*	*15.6*	+
			58	*Inner City Mixed Race*		*14.8*	*14.5*	**-**
			66	*Suburban Mixed Race*		*14.8*	*15.0*	+
			109	*Urban Mainly Caucasian*		*15.0*	*15.4*	+
			58	*Suburban Asian*		*17.6*	*17.0*	-
			59	*Pre-High Science Enthusiasm*		*17.6*	*17.0*	-
			85	*Pre-Low Science Enthusiasm*		*15.0*	*16.4*	**+**
			55	*Pre-Super Low Science Enthusiasm*		*14.5*	*13.9*	**-**

*** Studies that determined the treatment effect between the control and treatment groups. C=control, T=Intervention/Treatment, ES=effect size, FT=Field Trip Activities. Pre and Post score are mean averages unless designated otherwise. Pre and post scores listed unless researchers did not specifically state in the article. Gray shaded, bold sign means significant, $p<0.05$. Jarvis & Pell (2002) did not give pre/post scores for all items. * Studies that calculated a difference between the control and treatment group. *Italics*=subpopulation effects.

Overall cognitive findings from the systematic review. Of the 14 studies investigating knowledge gains, 43 cognitive outcomes are reported and 31 (72%) are positive and significant at the 95% confidence level. Table 2.6 provides the different outcomes by study. Only two studies reported single item-analysis (Freedman, 2010; Puhek et al., 2012), of which four of seven are positive and significant ($p<0.05$). Five studies reported 16 outcomes for sub-categories from the overall assessment such as tropical forest questions and conceptual knowledge. Of the 16 outcomes, 11 (70%) were significant and positive ($p<0.05$). Students who visited these ISEIs had greater knowledge on water, soil, meadow, and wood ecosystems (RCT study) (Prokop et al., 2007), factual knowledge, conceptual knowledge (pre-post only) (Basten et al., 2011) knowledge about tropical forests, Laurentian forests, and St. Lawrence Forests (pre-post only) (Futer, 2005), and knowledge about personal meaning mapping and specific information found in science center exhibits (Stavrova & Urhahne, 2010).

Randomized control-treatment studies. Only three studies used a randomized control treatment study design, and of those studies only one (Prokop et. al, 2007) found a positive impact of visiting an ISEI compared to students who had yet to visit ($p<0.05$). Only two of the three studies determined if a significant difference between the control and treatment groups exists, and both studies did not find a difference (Holmes, 2011; Itzek-Greulich et al., 2015.) In the study by Holmes (2011), the control group students completed all testing shortly after arriving at the field trip destination, the University of Louisiana’s IDEA Place, but in another building. Holmes did not provide any specifics about the control students pre-and post-scores, which would show if the change between the two scores was positive or negative. Holmes assigned sixth grade students into three intervention groups where students experienced (1) a guided tour of the exhibits, (2) a 30-minute workstation lab activity, or (3) both the tour and the activity. Only the students who experienced just the guided tour showed a significant positive growth in knowledge. The students who experienced both the tour and activity showed growth between the pre-and post-assessment, while the students who experienced only the activity scored on average a lower score on the post-assessment than the pre-assessment, although these findings were not significant ($p<0.05$). Also her conclusions are not surprising considering the cognitive test administered only asked questions from the guided tour and not the workstation activities. Students who experienced both the tour and the activity had less time for the guided tour than students who only experienced the guided tour, which was reflected in the difference between the pre-and post-assessments.

Itzek-Greulich et al. (2015) also compared the differences between a classroom group with a group that visited the Experimenta Research Center. Both groups of 9th grade students in middle-track schools experienced the same lab-based activities, pre-lesson, and post-activities. The only difference was the location of the labs. Students who conducted the labs at school had higher gains on the carbohydrate knowledge assessment, chemical analysis assessment, chemical terms assessment, and declarative knowledge assessment, while students who conducted the labs at the outreach lab had higher gains only on experimental knowledge. None of the findings were significant. Itzek-Greulich et al. did control for the pre-assessment scores; however, the school group scored much lower on all five pre-assessments than the outreach lab group. This study also did not discuss or address a potential novelty effect, which may have led to the greater outcomes by students at school.

Quasi-experimental studies. Five studies included a control group, but the researchers did not actually determine if the two group's results were significantly different (Prokop et al., 2007; Puhek et al., 2012; Sturm & Bogner, 2010; Sweet, 2014). Prokop et al. (2007) was the only study to randomly assign students into either the control or treatment groups. The other four studies did not use randomization. For these studies, the control groups served as a proxy to show that any growth between the pre-and post-assessment for the field trip students occurred solely from the field trip, as long as any gains in knowledge by the control groups was not significant ($p<0.05$). Although the control groups' knowledge gains are insignificant while the intervention groups' gains are significant, this type of comparison does not determine if the actual change in knowledge between the two groups is significantly different.

Sweet (2014) and Puhek et al. (2012) had students who did not visit the ISEI complete a virtual experience based on the actual field trips. Both studies calculated the effect sizes for the change in knowledge between the pre- and post-assessments, and both studies reported slightly larger effect sizes for the group of students who actually visited the real venue. Sweet (2014) reported moderate size effects on a health assessment by students using the virtual experience ($d=0.50$) but a large effect on the health assessment after the real field trip ($d=0.91$). For the science assessment, the effect was moderate for both groups (virtual, $d=0.65$; real, $d=0.74$). Puhek et al. (2012) reported mixed findings between the two groups depending on the material taught. For example, when studying tree rings, 8th grade students who visited the nature reserve scored 23% of standard deviation higher on the post assessment ($p<0.05$) while students who visited the reserve through a virtual experience scored only 8% of a standard deviation higher on the post-assessment. On the pH-scale assessment, students who visited the nature reserve scored 12% of a standard deviation higher on the post-assessment while students who visited the reserve through a virtual experience scored 4% higher on the post-assessment than the pre-assessment.

Sturm and Bogner (2010) created two control groups of students. The first control group was not taught any of the material covered in the field trip to a natural history museum where students participated in an 8-workstation lab program on bird biology. The other control group completed similar workstation activities but in the classroom. As expected the control group that was not taught the material had a similar score on the pre-assessment and the post-assessment. The other control group where the students were taught the same material but in a classroom did show significant improvement between the pre-and post-assessment (5.32 pre-mean average to 7.53 post-mean average) but not at the level of those students who actually visited the museum (5.15 pre-mean average to 10.3 post-mean average). Prokop et al., (2007) reported similar findings between a classroom group and a field trip group. Students in both groups showed academic gains on four assessments about the water ecosystem, soil ecosystem, meadow ecosystem, and wood ecosystem, even though the classroom group did not have a lesson or complete any activities on these ecosystems. However, the 6th grade students who actually visited and studied these ecosystems at a local nature park showed significant changes. For example, students who stayed in the classroom scored a mean average of 3.33 on the pre- water ecosystem test and a 3.41 on the post-test, while students on the field trip scored 3.31 on the pre-water ecosystem test and a 5.81 on the post-test.

Field trip structure. Finally, several studies focused on whether a more controlling structured field trip experience had a larger impact on student's knowledge gains than a more free-choice, less structured experience. All three studies included a classroom control group that did not visit the ISEI. Of the two groups from each study that did visit the ISEI, the high structured group (controlling) and the less structured group (autonomy supported), students showed similar gains (Basten et al. *2*011; Stavrova & Urhahne, 2010; Wilde & Urhahne, 2008). Wilde and Urhahne (2008) found that regardless of the structure of the field trip, students who visited the ISEIs on average scored 156% of a standard deviation higher on the post-assessment than the pre-assessment.

Since so many of these studies included in the review employ a very specific field trip routine without much time for students to freely explore the venue, this may factor into some of the mixed findings from the control-treatment studies. However, all three studies (Basten et al., 2011; Stavrova & Urhahne, 2010; Wilde & Urhahne, 2008) about field trip structure found that students in the more structured environment reported having a greater interest/enjoyment ($p<0.05$; $p<0.01$, respectively). However, students attending a more advanced, college preparatory school who experienced the field trip in a less-controlling environment reported having a higher interest/enjoyment. Basten et al.'s (2011) study differed from the other two studies by having the museum guides constantly hamper the highly structured group about running out of time and paying close attention to pass the test. The less-structured group of students was told not to worry about time and that there was not going to be a test. For Stavrova and Urhahne (2010) and Wilde & Urhahne (2008), the highly structured group of students did not get to choose their partners, completed more worksheets, answered more open-ended questions, and completed more specialized tasks.

Table 2.6

Cognitive Impacts from Visiting an ISEI during a School Field Trip

Researcher/ Publication Year	Type of Experience	Ages/ Grade	Pop. Size	Study Group	Survey	Effects		
						pre	post	sign
Experimental Studies: Randomized Control-Treatment								
Holmes, 2011	Louisiana Tech *IDEA Place*-Science Center-Physical Science-researcher created program.	6th	56 C	Control Group vs /Treatment Group***	30 question exhibit test covering electricity, light and optics, mechanics, sound and waves, and weather	$F_{(3,224)}$ =1.0		0
						t=0.932		+
			53 T	Guided Tour Group		t=2.371 ES=0.436		+
			61 T	30 Minute Lab Activity		t=-0.339		-
			58 T	Guided Tour and Activity		t=1.859		+
Itzek-Greulich et al., 2015	*Experimenta* Research Center-Half day lab experience-*Pre-tests scores for the School/Museum group were very low compared to the school group. Center created program.	9th middle tracked school	376 C 394T	School vs School/Museum***	Carbohydrate Knowledge (12 items)	ES=0.17		**-**
					Chemical Analysis-16 item multiple choice	ES=0.18		**-**
					Chemical Terms-9 terms covered	ES=0.04		**-**
					Experimental Knowledge-8 item multiple choice	ES=-0.01		+
					Carbohydrate Knowledge (12 items)	ES=0.17		**-**
Prokop et al., 2007	Local Area-Environmental- Experimental students had 4 activities/tasks to complete about ecosystems-Researcher created program.	6th	69 C 74 T	Control Group	Water Ecosystem	3.33	3.41	⩲
				Intervention Group		3.31	5.81	**+**
				Control Group	Soil Ecosystem	3.11	3.16	⩲
				Intervention Group		3.12	5.61	**+**
				Control Group	Meadow Ecosystem	3.36	3.4	⩲
				Intervention Group		3.39	6.01	**+**
				Control Group	Wood Ecosystem	3.1	3.12	⩲
				Intervention Group		3.03	5.92	**+**
Quasi- Experimental Studies: Non-Randomized Control-Treatment								
Wilde & Urhahne, 2008	*Natural History Museum of Berlin*-structured, only visited 3 of 5 sections of museum. 25 minutes per section then went and did a task. Small groups. Students attended a top tier, high academic school. Research created program	5th	33 C 207 T	Control Group	26 open and closed questions	9.21	8.09	**-**
				All Students		ES=1.56		**+**
				open ended tasks vs. closed ended tasks, very specific		$F_{(2,170)}$=1.74		0
Sturm & Bogner, 2010	Natural history museum-bird biology-8 workstations. Researcher created program.	6th	27 C1	Control Group (no lesson)	17 question multiple-choice test and 1 semi-open question over the pre-lesson and workstation activities.	Z=-1.090		⩲
			46 T	Museum Group		5..15	10.3	+
			117 C2	Control Group 2-Classroom Group (taught same lesson as Museum Group)		5.32	7.53	⩲

Table 2.6 (cont.)

Cognitive Impacts from Visiting an ISEI during a School Field Trip

Researcher/ Publication Year	Type of Experience	Ages/ Grade	Pop. Size	Study Group	Survey	Effects		
						pre	post	sign
Puhek et al., 2012*	Nature Reserve at Maribor Island. 8 lesson activity about trees-Researcher created	8th	133 C 78 T	Intervention-Real FT	Tree Rings (1 question)		ES=0.23	**+**
				Control –Virtual FT			ES=0.08	-
				Intervention-Real FT	The pH-Scale (1 question)		ES=0.12	-
				Control –Virtual FT			ES=0.04	+
				Intervention-Real FT	Illumination in the Forest (1 question)		ES=0.04	-
				Control –Virtual FT			ES=0.09	+
				Intervention-Real FT	Identifying Leaves (1 question)		ES=0.25	**+**
				Control –Virtual FT			ES=0.11	+
				Intervention-Real FT	Identifying Keys (1 question)		ES=0.21	**+**
				Control –Virtual FT			ES=0.28	**+**
				Intervention-Real FT	Biodiversity (1 question)		ES=0.06	+
				Control –Virtual FT			ES=020	**+**
				Intervention-Real FT	Composite Score on the six open question assessment		ES=0.23	**+**
				Control-Virtual FT			ES=0.18	**+**
				Boys-Real FT			ES=0.36	**+**
				Boys-Virtual FT			ES=0.30	**+**
				Girls-Real FT			ES=0.14	+
				Girls-Virtual FT			ES=0.10	+
				Satisfactory BAL-R			ES=0.36	+
				Satisfactory BAL-V			ES=0.06	-
				Good BAL-R			ES=0.28	+
				Good BAL-V			ES=0.13	+
				Very Good BAL-R			ES=0.17	**+**
				Very Good BAL-V			ES=0.23	+
				Excellent BAL-R			ES=0.20	+
				Excellent BAL-V			ES=0.22	**+**
Sweet, 2014	*Kelsay Dairy Farm*-educational guided tour based on state standards. Farm created program.	3rd	125 C 72 T	Control-Virtual FT	8-item science test designed by the Dairy Farm Educational Coordinator		**ES=0.65**	**+**
				Intervention-Real FT			**ES=0.74**	**+**
				Control-Virtual FT	7-item health and wellness test designed by Dairy Farm Educational Coordinator		**ES=0.50**	**+**
				Intervention-Real FT			**ES=0.91**	**+**

Table 2.6 (cont.)

Cognitive Impacts from Visiting an ISEI during a School Field Trip

Researcher/ Publication Year	Type of Experience	Ages/ Grade	Type of Study	Study Group	Survey	Effects		
						pre	post	Sign
Non-Experimental Studies: No Control								
Basten et al., 2011	Zoo- Students visited workstations in a zoological garden. (40 min). Completed Worksheets. Two Intervention Groups: (1) Controlled students constantly reminded of time constraints, lots of directives. (2)Autonomy-Supported told to take their time and do what they could do. Researcher created program.	Ages 10-11	100	Middle Track-Controlling	Factual Knowledge- 21 Questions	36	38	+
				Middle Track-Autonomy Supportive		37	38	+
				High Track-Controlling		39	44	**+**
				High Track-Autonomy Supportive		40	43	**+**
			106	Middle Track-Controlling	Conceptual Knowledge- 7 questions	3.1	6.3	**+**
				Middle Track-Autonomy Supportive		3.8	5.2	**+**
				High Track-Controlling		4.2	6.3	**+**
				High Track-Autonomy Supportive		3.8	5.5	**+**
Freedman, 2010	*Hall of Health-H*ealth museum.-(90 min trip, 30 min presentation) Hands-on activity, Museum created program	Ages 10 to 11	151	All Students	MyPryamid Test (19 Questions)	72%	93%	**+**
					Correct number of Food Groups	14%	54%	**+**
Futer, 2005	Montreal *Biodome*-Guided Environmental Tour- Biodome created program.	4th -6th	151	Tropical Forest Group	Tropical Forest Content-9 Questions	23% change		**+**
			172	Laurentian Forest Group	Laurentian Forest Content-9 Questions	12% change		**+**
			68	St. Lawrence Group	St. Lawrence Content-9 questions	18% change		**+**
Jarvis & Pell, 2002	*Challenger Space Center*- small group, staff led activities plus teacher facilitated play and focused science talk during free-roaming time. Science center created program.	Ages 10 to 11	655	All Students	10 question cognitive test about comets based on the 'Rendezvous with a Comet' experience.	4.40	5.24	**+**
				Boys		4.51	5.89	**+**
				Girls		4.28	5.67	**+**
Kamarainen et al., 2013	Local Pond-Environmental- Probeware, Lab-Based Activities-Researcher created program.	6th	71	All Students	14 Question Environmental Test	ES=1.0		**+**
Krombaß & Harms, 2008	*Inatura* Natural History Museum-(1 hour) Students completed 14 question worksheet based on completing specific tasks. Students attended a top tier, high academic school-Researcher created program.	6th -9th	148	All Students-	14 multiple-choice questions	ES=1.03		**+**
Stavrova & Urhahne, 2010	*Deutsches Museum.* Science center. Worksheets. Educational program/lab with guided tour of exhibits. Science center created program.	8th-9th	96	Autonomy Supported	Personal Meaning Mapping	2.73	2.84	**+**
				Controlling		2.54	3.17	**+**
				Autonomy Supported	11 question multiple choice based on exhibits/ guided tour	4.99	6.90	**+**
				Controlling		5.66	6.65	**+**
				Autonomy Supported	Summative Score of the two measures	7.71	9.74	**+**
				Controlling		8.19	9.81	**+**

The sample size is listed for each individual group assessed. C=control, T=Treatment, ES=effect size, BAL=Biology Achievement Level, V=Virtual Field Trip, and R=Real Field Trip. ***Studies that estimated the treatment effect between the control and treatment groups. Effect sizes listed when provided. Unless stated otherwise, pre- and post-scores are the average mean score. If the pre- and post-scores are not provided in the study, an alternate analysis is listed. Outcome effect is are positive (+), zero (0) or negative (-). Light gray shading, bold sign means effect was statistically significant (p<0.05).

Conclusion. The systematic review of the literature on the impact of field trips to an Informal Science Education Institution (ISEI) suggest not enough research, specifically RCT studies, exists to determine the cognitive and affective benefits 21st century students receive when visiting an ISEI during a school field trip. Table 2.7 below provides a summary of the overall cognitive and affective outcomes for all students. Subpopulation outcomes are not included. If a study calculated the control and treatment group separately and the difference between the pre- and post-scores were positive and significant at the 95% confidence level for both groups, I only considered the outcome positive if the effect size was larger for the treatment group (see Puhek et al. (2012) identifying keys).

Based only on the pre-post studies, students appear to have at least short term learning gains and a positive change in science interest following a visit to an ISEI. However, the limited RCT findings suggest that students do not have any short-term learning gains or changes in attitudes after visiting an ISEI. Prokop et al. (2007) was the only RCT study to report positive and significant findings, but the researchers did not estimate the differences between the two groups. All 75% of the positive outcomes from the RCT studies examining affective outcomes were from the Prokop et al. study.   What has not yet been demonstrated is whether the learning in the ISEI or the change in attitude after visiting an ISEI is significantly greater than not experiencing the field trip to an ISEI at all.

Table 2.7

Percent of Positive Outcomes by Study Type and Topic

Study Type	N of studies	N in US	Mean Sample Size	Total Number of outcomes studied	Percent of positive outcomes ($p<0.5$)
RCTs examining Cognitive Outcomes	3	1	380.3	10	50%
RCTs examining Affective Outcomes	2	1	186	4	75%
Pre/Posts examining Cognitive Outcomes (w/control)	4	1	208.5	11	55%
Pre/Posts examining Affective Outcomes (w/control)	3	0	258	15	66%
Pre/Posts examining Cognitive Outcomes	7	2	245	13	96%
Pre/Posts examining Affective Outcomes	4	1	353	32	41%

Preview of Chapter Three

Chapter three has five sections and discusses in detail survey development, administration, and study design. The first section describes the site selected for the study. The second section explains how the surveys were developed to measure affective and cognitive changes from visiting an ISEI during a school field trip, and how the final surveys were administered at each school. The third section discusses the validity and reliability of the instrument. The fourth section discussions demographics, participation rates, and passing rate for the cognitive assessments. The final section describes the estimation models used for analysis of affective and cognitive outcomes.

Chapter Three: Instrument Development, Administration of Surveys, and Methods

Hundreds of empirical and theoretical studies exist about the benefits students receive during field trips to science-related sites such as museums, science centers and outdoor environments. Here these sites are called Informal Science Education Institutions (ISEIs). Yet, researchers and educators continue to debate precisely what field trips offer beyond the regular classroom learning experience, and many informal learning experts have demanded more rigorous and larger-scale evaluations (Andre et al., 2017; Dewitt & Storksdieck, 2008; Gutwill & Allen, 2012). For this study, I use a randomized control trial (RCT) method to answer the following questions:

1) After visiting a children's discovery science center during a school field-trip, how do students' attitudes regarding science centers and studying science differ from students who had not yet made such a visit?
2) After visiting a children's discovery science center during a school field-trip, how does students' science knowledge differ from students who had not yet made such a visit?
3) What overall impacts in attitudes or knowledge does the science center visit offer on certain subpopulations of students? Such populations include: white students, minority students, girls, boys, students who were a first time visitor to a science center, and students who had previously visited a science center before the study began?

With the RCT design, I used an online random generator to place school groups who were planning a school field trip visit to the hands-on children's Museum of Discovery (MOD) in Little Rock, Arkansas, into either a treatment or control group. Randomization occurred at the school group level, and groups ranged from single classrooms to multiple grades. The treatment students completed a survey following their visit while those in the control group completed the survey before their visit. It is important to note that before this study commenced, the MOD had just reopened after a nine-month renovation that transformed the space from a children's museum to a science, technology, and math center (Tidwell, 2012). If students had visited the MOD on a previous school field trip before participating in the study, they would have visited a very different site; and therefore, it seemed reasonable to assume that this small group of students would not impact the results. Figure 3.1 illustrates the study design and administration of the surveys.

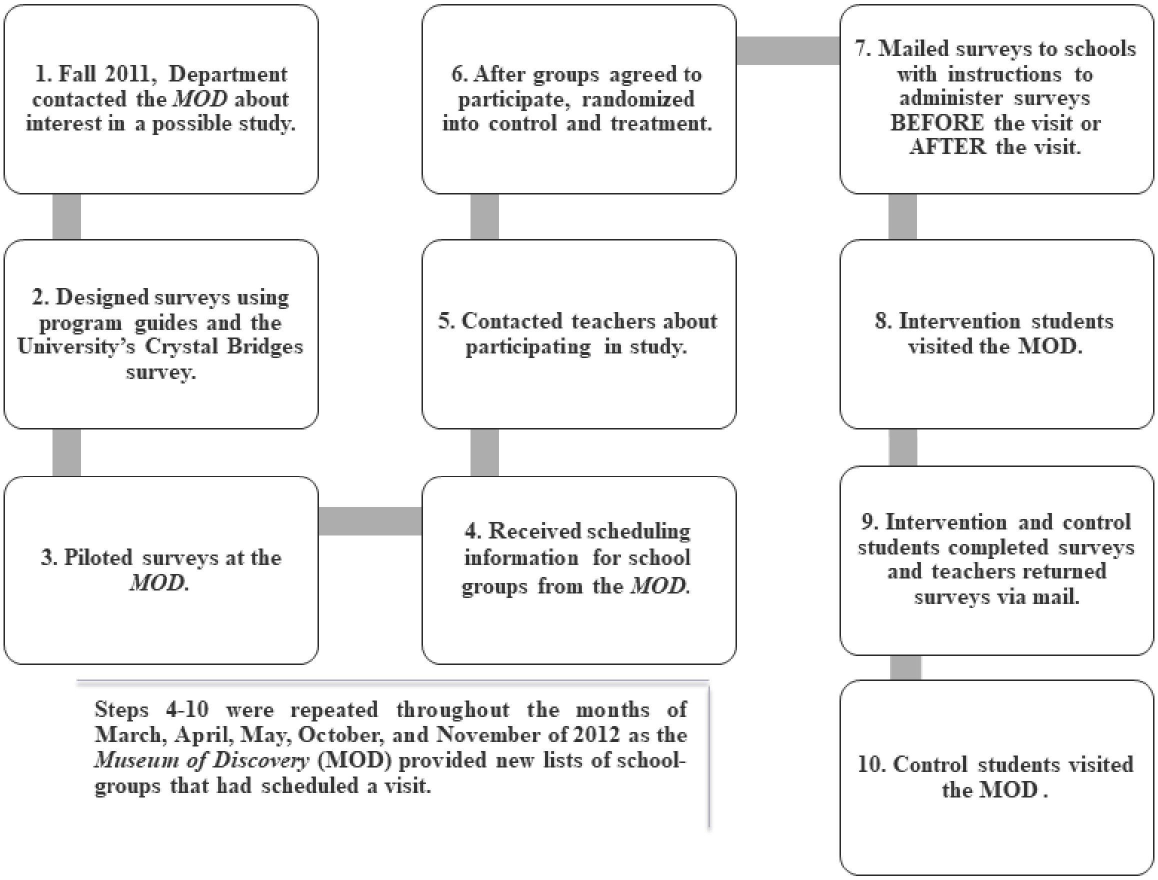


Figure 3.1. Steps Involved in the Development and Administration of the Affective Surveys and Cognitive Assessments

Site Selection

The Museum of Discovery (MOD) is a Smithsonian affiliate institution located in the downtown area of Little Rock, Arkansas. The museum, formally known as the Museum of Natural History and Antiquities, is the oldest museum in Little Rock and was established in 1927. The museum received its first accreditation by the American Association of Museums in 1993 and then received a re-accreditation in 2001. Before the first accreditation, the museum had collected around 14,000 historical and cultural artifacts and a vast number of species of animals and insects. In 1988, the museum moved to its current location and changed its name to the Museum of Discovery. The new museum provided more interactive programs and hands-on exhibits. In 2003, the MOD merged with the Children's Museum of Arkansas and began creating programs and exhibits for preschool aged children.

In 2011, just prior to the start of this study, the MOD underwent a complete renovation funded by a grant from the Donald W. Reynolds Foundation to transform the site from its role as a natural history museum to a science center (Museum of Discovery, n.d.). Science museums generally focus on collections of objects and educational displays, and the visitor experience is typically passive and based on observation, reading, discussing, and reflecting. Science centers emphasize technology to educate and primarily focus on the experiences created for visitors. Visitors participate in the learning experience by experimenting, exploring, and tinkering with objects. In reality, an ISEI can fall anywhere on the science center to science museum spectrum, and even though the ISEI may include 'museum' in its name, it may have more characteristics of a science center than a museum (IDEA, 2011). The Museum of Discovery has the word 'museum' in its name, and many of the newest renovations were geared for attracting an older audience. However, the MOD is still a children's discovery science center, and I will refer to it as a science center throughout this report.

After renovations, the Museum of Discovery featured five permanent galleries with over 90 hands-on exhibits and offered 17 educational programs led by museum educators including titles such as: Arkansas Animals, Arkansas Indians, Awesome Science, Body Basics, Boy Talk, Brian Dissection, Circuit Circus, Cool Canvas, Crime Solvers, Dinosaurs, Earthquakes, Girl Talk, Go Green, Heart to Heart, Invertebrates, Meet the Elements, Science of Toys, Senses, Sound is Groovy, and Sports Science. When school groups schedule a visit to the MOD, they can select an educational program for an additional cost. Many of the MOD's school groups that added an educational program selected either the *Arkansas Animal* or *Awesome Science* program. Therefore, for this study, I evaluated the impact of only those two programs on students' knowledge. Furthermore, both support Arkansas's K-12 science curriculum. Arkansas Animals focuses primarily on third to sixth grade standards while the Awesome Science program can be adapted for any age or grade.

Survey Development and Administration

Development of surveys. Along with Brian Kisida, a colleague and research assistant at the University of Arkansas, I created an instrument to measure general interest in museums and student attitudes about science. This instrument development included producing a basic test of science content knowledge that visitors would likely encounter from seeing either the Arkansas Animals program or the Awesome Science program. During the Arkansas Animals program, a MOD educator handles a live Texas Brown Tarantula, tree frog, toad, alligator, screech owl, doves, mice and a rabbit. The educator discusses the differences, physiologies, and life cycles of invertebrates, amphibians, reptiles, birds, and mammals. This program takes place in a MOD classroom which can hold about 30 people. During the Awesome Science program which occurs inside the science center theater, a MOD-educator demonstrates Bernoulli’s principle, the properties of solids, liquids, and gases, density, catalysts, and electrolysis through 12 different demonstrations. See Table 3.1 for a list of the Awesome Science and Arkansas Animals content questions.

Table 3.1

Questions Used to Assess Students Who Attended the MOD *Educational Programs : Arkansas Animals and Awesome Science*

Awesome Science Questions
1. What is the principle that explains how dry ice causes “metal to scream” called?
2. What is it called when a solid turns directly into a gas?
3. Why does a balloon filled with helium float?
4. Which gas escapes as dry ice heats up?
5. What is a compound that speeds up a chemical reaction?
6. What is a property of hydrogen?
Arkansas Animals Questions
1. Which of these is an invertebrate?
2. Which of these is a reptile?
3. Which of these is an amphibian?
4. Which of these is not cold-blooded?
5. All invertebrates do not have…?
6. Which of these is a mammal?

Besides questions about the two programs, the instrument (Appendix C) also included a section for demographic information and whether students have visited the museum previously. For the students in the control group who had not yet attended either program because they had not yet experienced the field trip, I included survey questions from the program they would eventually see at a later time. For control groups that did not select a program or did not select either the Arkansas Animals or Awesome Science programs, I either assigned those groups questions from the Awesome Science program or they were not asked any cognitive questions.

Not all schools attended a program, and so for those schools in the treatment group, I did not include science program content questions on their surveys.

We were fortunate to have an attitude instrument developed for use in a survey performed at Crystal Bridges Museum of American Art, and we used those items as a starting point for the affective items applied in this study. We developed the science knowledge questions using the program guides provided by the museum and by consulting the Arkansas K-12 state standards. For creating content questions, the program guides had several underlying concerns. According to the MOD program director, “Not all of [the program guides] are entirely complete but it should give you an idea of some of the content we deliver. We constantly adjust and alter our programs to match the audience depending on age, group size, capability etc.” (T. Lipman, personal communication, January 17, 2012). We then created questions based on the lower two-levels of Bloom’s Taxonomy (knowledge and comprehension) (Bloom, Englehart, Hill, & Krathwohl, 1956). For example, the Awesome Science part of the instrument asked, “Why does a balloon filled with helium float?” The four answer choices were (A) Helium is lighter than air, (B) The balloon expands, (C) The air heats up, and (D) Evaporation. We included only the two bottom levels of Bloom’s taxonomy because these programs were approximately 30 minutes and focused on introductory topics, thus it was unlikely students would have a more in-depth and higher level understanding of the material covered.

Before surveying new students in the fall, we made several changes to the survey instrument. First, we added two questions about the nature of science and a question about the value of science. In the spring analysis, students in the treatment group were scoring higher on the content questions than students in the control group and we wondered if the program might also impact students’ general understanding of science. Second, we also modified the questions about previous visits to museums. Lastly, we added two additional choices in response to the question about what is their favorite subject: art/music and PE. We added these two answer options to give students a non-core choice and 48% of students in the fall cohort selected either PE or music/art for their favorite subject (See Appendix C for final copies of the spring cohort survey and the fall cohort survey). These changes did not affect the outcomes used for this study.

Measuring affective outcomes. To determine the effects on student attitudes, I first combined student responses from the 11 affective items. Next, I divided the overall affective items into two smaller scales: *Interest in Science Centers/Museums* and *Interest in Studying Science*. Table 3.2 shows the individual items with a brief description of how the student could respond to each item.

Table 3.2

Items Included on the Overall Affective Instrument

Interest in Science Centers/Museums Scale	
How interested are you in visiting a science museum?	4-point Likert
Would you like more science museums in your town?	Yes or No
Trips to science museums are interesting.	4-point Likert
I plan to visit science museums when I am an adult.	4-point Likert
Trips to Science Museums are Fun	4-point Likert
Interest in Studying Science Scale	
How interested are you in learning about science?	4-point Likert
I would be interested in joining a science club if my school offered one.	4-point Likert
I would like to study science in college	4-point Likert
My favorite subject in school is?	Math, Science, History, PE, Reading/Writing, Art/Music[2]
When I grow up I want to be a(n)?	Open Item
Science is an important part of my life	4-point Likert

I created the Overall Affective Instrument, which includes all student responses about attitude towards science museums and towards studying science in school, by summing student responses for the 11 affective questions. Students received 4-points per item if the student

---

[2] Art/Music and PE were added as options to the fall survey but were not on the spring survey.

answered *yes, strongly agree, very interested*, reported their favorite subject was *science*, or wrote that when they grow up they want to work in a *science profession*. They received 3 points per item if they answered *somewhat agree* or *interested*. They received 2 points per item if they answered *somewhat disagree* or *a little interested*. They received 1 point per item if they answered *disagree, not interested, no*, listed another subject besides science as their favorite subject, or they listed a non-science profession for their future career choice. I calculated the average mean for all responses answered. Finally, I standardized the score. I will use standardized scores to estimate the impact of visiting a science center on student's affective domain compared to students who have not yet visited the MOD.

Table 3.3 shows the number of observations, mean, Cronbach's alpha, standard deviation, and minimum and maximum score for each inventory. On the Overall Affective Instrument, the average mean score was 2.87 out of 4 with a standard deviation of 0.63. On the Interest in Visiting Science Centers/Museums Scale, the average mean score was 3.35 out of 4 with a standard deviation of 0.67. On the Interest in Studying Science Scale, the average mean score was 2.46 out of 4 with a standard deviation of 0.73. Regardless of if students were in the treatment group or control group, students, on average, had more positive attitudes about visiting science centers than actually studying science.

Table 3.3

Descriptive Statistics for the Affective Domain Instruments

	α	Mean	SD	Min	Max
Overall Affective Items	0.85	2.87	0.63	1	4
Interest in Visiting Science Museums/Centers Scale	0.80	3.35	0.67	1	4
Interest in Studying Science Scale	0.77	2.46	0.73	1	4

Measuring cognitive outcomes. To gauge any cognitive impacts from participation in either the Arkansas Animals Program or the Awesome Science Program, the survey questionnaire included six questions for those enrolled in the Arkansas Animals Program (Arkansas Animal assessment) or 6 questions for those enrolled in the Awesome Science program (Awesome Science assessment). Some school groups did not choose to attend an educational program. I tallied a raw average based on the number of questions students answered, and then I converted the raw mean into a standardized score. Table 3.4 provides information about the average mean, standard deviation, Cronbach's alpha, the minimum score and the maximum score. I also merged the data from the two programs to create a Combined Knowledge Assessment. For the Arkansas Animals assessment, the average mean score was 65.4% with a standard deviation of 23%. For the Awesome Science assessment, the average mean was 39.5% with a standard deviation of 21%. For the combined assessments, the average mean score was 47.9% with a standard deviation of 25%.

Table 3.4

Descriptive Statistics for the Cognitive Domain Assessments

	# of Surveys	α	Mean	SD	Min	Max
Arkansas Animals Assessment	435	0.43	65.4%	23%	0%	100%
Awesome Science Assessment	915	0.33	39.5%	21%	0%	100%
Combined Assessments	1350	0.47	47.9%	25%	0%	100%

Administration of surveys. We did not want to take time away from the field trip experience and so decided to mail the instrument to the schools rather than to administer them on site. The final survey draft covered grades 3rd-8th rather than different grade bands because some groups had students attending across multiple grades. I did not include K-2nd grades in the study because during the visit to the museum, several K-2 grade teachers said that many of their students may have difficulty reading the questions. If the surveys went to the school, the teachers would have to help students complete the surveys and this could impact how students might answer the questions. I did not include 9th-12th grade students in the study because this age group was not making reservations at the MOD during the study. Only four groups at the high school age level scheduled a visit during the four months of this study.

In March of 2012, the museum provided a list of all the schools with contact information for all visits scheduled in April. The museum sent updated lists one month in advance; for example, I received the list of schools visiting the museum in May at the beginning of April. I contacted the schools by both email and phone. I emailed teachers a standard letter asking them to participate in our study and to provide us with the number of surveys and grade levels of students who will visit the museum. If I did not hear back from the teacher via email, I then called the teacher. Once a school agreed to participate, I randomly assigned school groups to either the treatment or control group. I mailed the survey version to teachers based on the educational program selected. I also included a parent letter providing an opportunity for a student to opt out of the study, an instruction letter that either told the teacher to administer the survey AFTER visiting the museum (treatment) or to administer the survey BEFORE their scheduled visit to the museum (control) (see Appendix B for a copy of the parent letter), and a teacher survey. The teacher surveys were dated so we could assess if the surveys were completed before or after the fieldtrip as outlined in the instruction letter.

I administered surveys to groups who had scheduled a visit to the Museum of Discovery between April 19th and May 17th (the spring cohort) and between October 14th and November 30th (the fall cohort.) At the end of survey collection in May, the sample size was 1250 students.

I continued surveying new students in the fall to increase the sample size. The final sample size was 1830 students. Since teachers in the fall needed time to plan and make reservations for a field trip, I waited until October to start surveying new school groups.

For this study, I used three different surveys (Appendix C). All three surveys had 8 demographic questions and 11 affective domain items about interest in science and science centers. Survey A also included six questions about the Arkansas Animals program and was administered to two treatment school-groups who had attended the Arkansas Animals program, two control school-groups that selected to attend the Arkansas Animals program when they visited the MOD at some later date, and two control school-groups of students that selected a different program to attend at some later date. Along with the 11 affective items and 8 demographic questions, Survey B also included six questions about the Awesome Science program and was administered to eight treatment school-groups who had attended the Awesome Science program, six control school-groups that requested the Awesome Science program when they visited the MOD at some later date, and three control school-groups that had either selected a different program or did not select a program at all. Survey C only had the 11 affective items and 8 demographic questions and was administered to eight treatment school-groups and two control school-groups that either did not select any educational program or selected a different educational program than the two programs evaluated. Also, all of the school groups that participated in this study selected at most one program for each student to attend. Table 3.5 lists the three surveys with the number of school groups and students by treatment, control, and the entire sample.

Table 3.5

Number of Students and School Groups Asked to Complete Each Survey Version

Survey	Control		Treatment		Total	
	Number of Students	Number of Groups	Number of Students	Number of Groups	Number of Students	Number of Groups
Survey A: *Arkansas Animals*	285	4	152	2	437	6
Survey B: *Awesome Science*	536	9	389	8	925	19
Survey C: No Program	140	2	328	6	468	8

Reliability and Validity

To determine survey reliability, I used Cronbach's alpha to measure internal consistency between survey statements. Overall, Cronbach's alpha is a 0.85. Rossi, Lipsey, and Freeman (2004) recommend reliability coefficients of 0.90 or higher, but also state that for program evaluation outcome measures this is a "relatively high standard" (p. 220). Table 3.3 and 3.4 above provide Cronbach's alpha for all of the instruments used.

Factor analysis. To determine construct validity of the multiple-choice questions for the educational programs, the lead museum educator and three teachers reviewed the questions. All four reviewers agreed that the questions measured the corresponding program. Two of the teachers recommended not surveying students in kindergarten to 2nd grade because they believed some students would have difficulties reading and answering the questions. To determine construct validity for the affective domain items, I conducted an exploratory factor analysis, which has several uses. First, exploratory factor analysis can determine the number of dimensions represented by a set of variables and the strength of correlations. Second, exploratory factor analysis can show evidence of construct validity from self-reporting instruments (Williams, Onsman, & Brown, 2012). The main section of the student survey was designed to measure a construct characterized as *interest in science.* Therefore, the eleven statements about interest in science should result in one factor that describes most of the variances between these eleven statements. When determining the number of factors, a common practice is to only include factors that have an eigenvalue greater than one (i.e., variances greater than 1). Eigenvalues are created by consolidating the variance into a correlation matrix. Factors with eigenvalues less than one are often omitted from solutions because these eigenvalues account for only a small fraction of the total variance. In factor analysis, the first factor accounts for the most variance, the second factor will account for the next level of variance, and so on. This variance is reported in eigenvalues (Gebotys, 2011). Table 3.6 contains the eigenvalues of the eleven factors. Factor 1 has an eigenvalue of 4.28 representing 87% of the total variance. All other factors have eigenvalues less than 0.66.

Table 3.6

Exploratory Factor Analysis for Face Validity to Determine if Survey Items Measure the Same Construct: Student Attitudes towards Science

Factor	Eigenvalue	Proportion
Factor 1	4.28	0.87
Factor 2	0.66	0.14
Factor 3	0.19	0.04
Factor 4	0.15	0.03
Factor 5	0.07	0.02
Factor 6	-0.007	-0.001
Factor 7	-0.03	-0.006
Factor 8	-0.05	-0.011
Factor 9	-0.06	-0.012
Factor 10	-0.09	-0.019
Factor 11	-0.17	-0.034

Rotated factor loadings. The factor analysis suggests that the student survey instrument does indeed measure a single construct, *attitude towards science*. Since the affective domain questions are about a student's attitude towards science with a general focus on either interest in science or interest in visiting science centers/science museums, I rotated the factor loadings to see how the questions sorted into the first two factors. An item loads on a specific factor based on the correlation between the items and the factor and how the items are weighted for each factor. Rotated factor loadings shifts the axes to better encompass the actual data points overall. This is particularly useful if the majority of the variance falls into one factor, which these items do. Logically similar items should load on the same factor (Rahn, 2017). Table 3.7 shows the rotated factor loadings for the 11 affective survey questions, which loaded onto the two sub-constructs exactly as expected. The items that specifically asked about interest in science museums loaded together and the items that asked about studying science in general loaded together.

Table 3.7

Rotated Factor Loadings that Show Correlation of Items on the Affective Subscales

Survey Item	Interest in Science Museums	Interest in Studying Science
Interest in Visiting Science Museums	**61%**	40%
Interest in Learning About Science	41%	**63%**
Like More Science Museums in Hometown	**47%**	23%
Science is an Important Part of My Life	33%	**52%**
Trips to Science Museums are Interesting	**81%**	24%
Favorite Subject in School is Science	12%	**51%**
Trips to Science Museums are Fun	**71%**	19%
Plan to Visit Science Museums as an Adult	**58%**	36%
Interested in Joining a School Science Club	38%	**63%**
Interested in Studying Science in College	28%	**69%**
Wants future Career to be in a Science Profession	1%	**34%**
Science is an Important Part of My Life	33%	**53%**

Bold/gray is for the greater percentage to show which construct each item loaded.

Internal validity threats. Trochim (2006) discusses several potential internal validity threats to studies with two group control/treatment methodologies. The first major threat is selection bias, which is generally eliminated through randomization of subjects into either control or treatment group; this is the technique used here. In the next section, I compare the demographic differences between the two groups and will discuss selection bias in greater detail. The other internal validity threats are based on social interactions between either the treatment and control groups or between managers of the groups and administrators (Trochim, 2006). Since students in the control group did not attend the same schools with students in the treatment group, except in one case where the students were in different grades and visited the museum several weeks apart, students did not know if they were in the treatment group or the control group. With a delayed RCT study design, diffusion limitation, compensatory rivalry, and resentful demoralization should not be an issue. Also, since students in the control groups eventually visited the museum, teachers or parents did not need to pressure administrators to be included in the treatment group and the threat of compensatory equalization of treatment also becomes non-existent.

Population and Sample

Participation rates. Between March and April, I contacted 59 groups scheduled to visit the museum, 28 groups agreed to participate, but only 21 groups returned the surveys. Between October and November, I contacted 21 groups scheduled to visit the museum, 11 groups agreed to participate, but only 10 groups returned the surveys. Table 3.8 provides information about the response rate. Overall, our response rate was 31 school groups out of 80 groups contacted, representing 39%. I refer to groups as *school-groups* rather than just *schools* because in two cases, two groups came from the same school but at different times, with different students in different grades, so I treated them as separate groups.

Table 3.8

Participation Rates by School Groups

	Teachers Contacted	Agreed to Participate	Returned Materials
Spring	59	28 47%	21 36%
Fall	21	11 52%	10 48%
Total	80	40 50%	31 39%

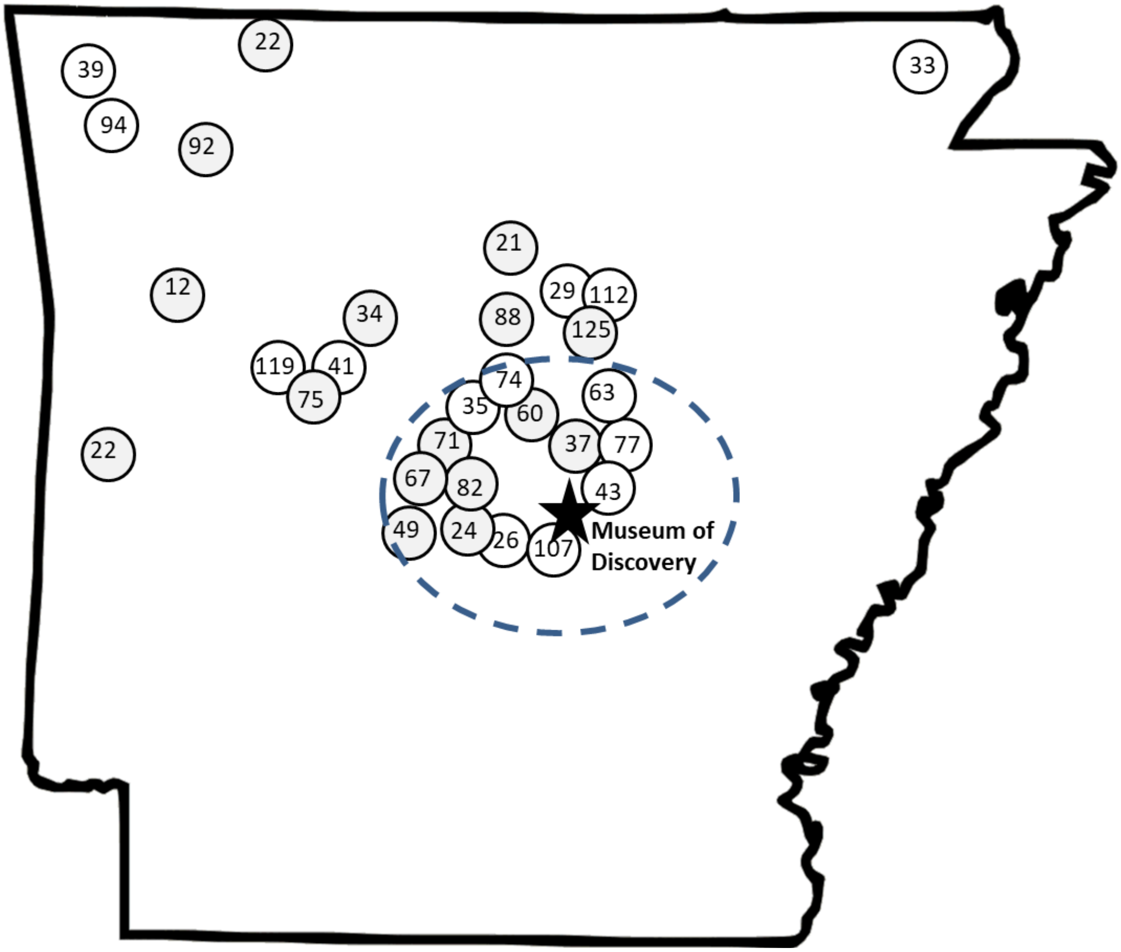
After each group agreed to participate, I randomly assigned them into the control or treatment group using an online random generator. Table 3.9 provides information about the rate returned by those assigned into the either control or treatment. In the spring of 2012, 15 groups were assigned treatment but only 12 treatment groups returned the materials, and 14 groups were assigned control but only 10 of those returned the materials. In the fall of 2012, six groups were assigned treatment but only four returned the materials. All seven groups assigned as control returned the materials. Seventy-six percent of those groups assigned to treatment returned the materials while 81% of those groups assigned to control returned the materials.

Table 3.9

Participation Rates by Control and Treatment School Groups

	Agreed to Participate	Returned Surveys	Percent Returned
Spring Treatment	15	12	80%
Spring Control	13	9	69%
Fall Treatment	6	4	67%
Fall Control	6	6	100%
All Treatment	21	16	76%
All Control	19	15	79%

Description of student participants. As observed from Figure 3.2, most of the school groups that visited the Museum of Discovery are from the central or northwest corner of the state of Arkansas. Sixteen school groups traveled more than 60 miles from their home campus to the MOD. The circle on the figure represents a 60 mile circumference around the Museum of Discovery. Darker shaded circles are treatment school groups and the lighter circles are control school groups. The numbers inside the circles are the population sizes for each school group.



Map not to scale and placement of school groups are approximate. Circles represent each school group included in the study and the number represents how many students completed the survey from that school group.

Figure 3.2 Approximate Location of Schools Compared to the MOD

Table 3.10 provides descriptive information about the control and treatment groups and the overall sample. In total, 1830 students submitted a survey. The control group consisted of 961 students (53% of the total number of subjects who completed surveys) and the treatment group consisted of 869 students (47% of all students). Although there was variance in school group sizes for both the control and treatment groups, the average school group size was 64 students for the control group and 54 students for the treatment group. The smallest group was 12 students and the largest group was 125 students. Based on student's reporting of race and ethnicity, all non-white students were combined to create a general minority category. In general, black students constitute 17% of the overall population, Hispanics constitute 6% of the overall population, and Asians, Native Americans, and students who selected *other* constitute about 10% of the overall population.

The randomization of school groups into treatment (the school groups who had visited the museum) or control (the school groups who have yet to visit the museum) should eliminate any significant baseline differences between the two groups. To determine any significant differences between the control group and the treatment group, I ran a two-sample t-test for each variable of interest. Any significant difference between the treatment and control groups, particularly the percentage of minority students or the percentage of students attending a majority FRL school, create a potential for selection bias. The t-test results showed several significant differences ($p<0.05$) between the control and treatment groups by student race, across individual grade levels, and percent of students attending a school with more than half of the student body receiving free and reduced lunch. I control for these differences using multiple linear regression by including the differences as covariates in the estimation.

Although a few differences exist between the treatment and control groups, the groups are very similar. Both groups have about the same percentage of boys and girls. The treatment and control groups each had approximately 15% of third graders, 2% of 7th graders, and 1% of 8th graders. The mean average grade level for those in the control group was 4.6 and for the treatment group was 4.7; this difference was statistically insignificant at the 95% confidence level. Just over 30% of students in the control group reported they had not visited the Museum of Discovery before the field trip, while 29% of the treatment group reported they had never visited the Museum of Discovery before the field trip.

Table 3.10

Descriptive Statistics for Control and Treatment Groups

	Sample Size	
	C	T
Population Size	961	869
White***	702 73%	522 61%
Minority***	255 (27%)	347 (40%)
Female	514 54%	450 52%
Male	446 46%	419 48%
Novelty	295 31%	254 29%
FRL50***	448 51%	550 63%
Grade 3	141 15%	128 15%
Grade 4***	398 41%	236 27%
Grade5***	163 17%	347 40%
Grade6***	230 24%	132 15%
Grade 7	16 2%	14 2%
Grade 8	13 1%	12 1%
Grade (Average)	4.60	4.65

All information obtained from survey responses except for the school percent for Free or Reduced Lunch (FRL) which was provided by the Arkansas Department of Education. C=Control population, T=treatment population, FRL50 = the number/percent of students who attended a school with more than 50% of the population received free or reduced lunch, Novelty = students who had never visited the MOD before. Two tailed t-test analysis completed. *** for each category, $p<0.01$

Passing rates. Although students may attend a science center educational program selected by their teachers, they may actually already know the material covered in the program. According to the passing rates for each question, students in the control knew some of the material and in some cases knew more than students who attended the program. These questions were based on the program guide provided by the MOD but different educators present the material different and may not cover all the material in the program. Table 3.11 provides the questions and passing rate for the control and treatment group. For both programs, the control group scored significantly higher on one question than the treatment group.

Table 3.11

Passing Rates for Each Cognitive Item by the Control and Treatment Students

Awesome Science Questions	Control Group	Treatment Group
1. What is the principal that explains how dry ice causes "metal to scream" called?**	17%	12%
2. What is it called when a solid turns directly into a gas?	29%	29%
3. Why does a balloon filled with helium float?**	69%	73%
4. Which gas escapes as dry ice heats up?	50%	53%
5. What is a compound that speeds up a chemical reaction?	22%	20%
6. What is a property of hydrogen?**	48%	56%
Average Overall Score	**39%**	**41%**
Arkansas Animals Questions	Control Group	Treatment Group
1. Which of these is an invertebrate?**	35%	55%
2. Which of these is a reptile?	70%	77%
3. Which of these is an amphibian?***	64%	72%
4. Which of these is not cold-blooded?	48%	52%
5. All invertebrates do not have…?	89%	90%
6. Which of these is a mammal?	70%	68%
Average Overall Score	**64%**	**67%**

Two-tailed t-test performed for each question. ***=p<0.01, **=p<0.05

Missing data. In any study missing data can be problematic. Missing data could come from an incomplete survey or a student not participating. A survey could be incomplete for multiple reasons: the student did not understand the question, believed the question to be too personal, or ran out of time. The question may not have the student's answer as an option or the student simply lost interest. If the data that are missing are not random, this could lead to misleading results by over representing or underreporting some data fields. For those students who did not complete a particular item on the survey, they become underrepresented (Graham, 2009). For example, if students did not know how to answer the question about how many times they had visited the MOD because they did not remember the name of the science centers they had previously visited, they may leave the question blank. If these are students who have never visited the MOD before then they would not be included in that particular subpopulation because the question has a blank response. If enough first-time visitors were not included in the analysis because they did not know how to answer the question about previous visits to the MOD, this could lead to inaccurate results about first-time visitors.

Only one student, who attended one of the control school-groups, opted out of the study, yet several students did not answer all questions on the survey. Two students, one each in the control and treatment groups did not answer any of the non-demographic survey questions and were excluded from the study. For individual item analysis, students who had a missing response for any of the variables in the estimation model would be excluded from the analysis. For both the cognitive and affective instruments, the final score was created by combining multiple items. Any blank response would not be included in the average raw mean score. The raw mean score was calculated based on the total number of items answered and not the total number of all items on that particular survey instrument.

Description of Data Analysis

Question 1: Affective impacts. To determine the impacts on student attitudes after students visit a science center compared to students who had yet to visit, I estimated using the following model for student *i*:

$$Survey_i = \alpha + \beta_1 T_i + \beta_2 Grade_i + \beta_3 Minority_i + \beta_4 Totalvisits_i + \beta_5 FRL_i + \varepsilon_i$$

where *Survey* is the student's percentage score for the Overall Affective Items or the student's percentage score on the survey section *Interest in Science Centers/Museums* or *Interest in Studying Science*. *T* is a binary variable equal to 1 if the student is in the treatment group and visited the Museum of Discovery during a school field trip and 0 if otherwise; $ß_1$ is the average effect from visiting an ISEI on a school field trip. Since randomization occurred at the group level, there is a possibility of some individual differences between the treatment and control group, so the model also includes grade, race, free or reduced lunch status, and total previous visits to science centers.

Grade is a vector of dummy variables that indicates student *i*'s grade (3^{rd} – 8^{th} grade) and 3^{rd} grade was omitted and serves as the comparison grade for the other grade levels. Since the variable for each grade has only 0s and 1s, the variable is considered a dummy variable. *Minority* is a dummy variable that indicates a student *i*'s race as Hispanic, black, Asian, American Indian, or other as 1, and white as 0.

Totalvisits is number of visits to science centers by student *i* prior to the study. *FRL* is the percent of students on Free or Reduced Lunch at student *i's* school. Lastly, ε is a stochastic error clustered by the school group that student *i* came with to visit the science center. A stochastic error is all of the variation in the dependent variable (the survey score) that cannot be explained by all of the independent variables (grade, race, treatment, total science center visits, free or reduced lunch status). Also, clustering errors helps control for within-cluster correlations which can lead to distorted small standard errors, narrow confidence intervals, low p-values, and large t-statistics (Cameron & Miller, 2015).

Question 2: Cognitive impacts. To determine the difference in knowledge between students who visited the science center and attended the Awesome Science program and students who have not yet visited the MOD , I estimated using the following model for student *i*:

(Model 1)

$$Awesome_i = \alpha + \beta_1 T_i + \beta_2 Grade_i + \beta_3 Minority_i + \beta_4 Totalvisits_i + \beta_5 FRL_i + \varepsilon_i$$

where *Awesome* is the student's standardized percentage score on the *Awesome Science* assessment, which consisted of six questions. *T* is a binary variable equal to 1 if the student is in the treatment group and visited the Museum of Discovery during a school field trip and attened the Awesome Science program, 0 otherwise. *Grade* indicates student *i*'s grade (3^{rd}-8^{th}) and is a vector of dummy variables. *Minority* is a dummy variable of 0 if student *i*'s race is white or 1 if student *i*'s race/ethnicity is Hispanic, black, Asian, American Indian, or other. *Totalvisits* is the total number of visits to science centers by student *i* prior to the study. *FRL* is the percent of students on Free or Reduced Lunch at student *i's* school. Lastly, ε, is a stochastic error clustered by the school group that student *i* came with to visit the science center.

To determine the difference in knowledge between students who visited the science center and attended the Arkansas Animals program and students who have not yet visited the MOD , I estimated using the following model for student *i*:

(Model 2)

$$Animals_i = \alpha + \beta_1 T_i + \beta_2 Minority_i + \beta_3 Totalvisits_i + \beta_4 FRL_i + \varepsilon_i$$

where *Animals* is the student's standardized percentage score on the Arkansas Animals -assessment, which consisted of six questions. *T* is a binary variable equal to 1 if the student is in the treatment group and visited the Museum of Discovery during a school field trip and attended the Arkansas Animals program, 0 otherwise. *Minority* is a dummy variable of 0 if student *i*'s race is white or 1 if student *i*'s race/ethnicity is Hispanic, black, Asian, American Indian, or other. *Totalvisits* is the total number of visits to science centers by student *i* prior to the study. *FRL* is the percent of students on Free or Reduced Lunch at student *i's* school. Lastly, ε, is a stochastic error clustered by the school group that student *i* came with to visit the science center. Student grade level was not included in this model because only students in grades 4th-6th attended the Arkansas Animals program.

To determine the overall impacts on students' science knowledge regardless of program attended during a science center visit compared to the knowledge of students who had not yet visited the science center, I estimated using the following model for student *i*:

(Model 3)

$$Cognitive_i = \alpha + \beta_1 T_i + \beta_2 Grade_i + \beta_3 Minority_i + \beta_4 Totalvisits_i + \beta_5 FRL_i + \beta_6 Program_i + \varepsilon_i$$

where *Cognitive* is the student's standardized percentage score either on the Awesome Science assessment or Arkansas Animals assessment. Both surveys consisted of six questions. *T* is a binary variable equal to 1 if the student is in the treatment group and visited the Museum of Discovery during a school field trip and either attended the Arkansas Animals or Awesome Science program, 0 otherwise. *Grade* indicates student *i*'s grade (3rd-8th) and is a vector of dummy variables. *Minority* is a dummy variable of 0 if student *i*'s race is white or 1 if student *i*'s race/ethnicity is Hispanic, black, Asian, American Indian, or other. *Totalvisits* is the total number of visits to science centers by student *i* prior to the study. *FRL* is the percent of students on Free or Reduced Lunch at student *i's* school. *Program* is a dummy variable where 1 is if the program was *Arkansas Animals* or 0 if the program was *Awesome Science*. Lastly, ε_i, is a stochastic error clustered by the school group that student *i* came with to visit the science center.

Question 3: Affective and cognitive impacts for several student populations. To determine the impacts on different groups of students' attitudes and knowledge after visiting a science center and attending an educational program compared to students of the same subpopulation who had not yet visited the science center, I estimated using the following two models for student *i*:

(Model 1)

$$Affective_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 Grade_{ij} + \beta_3 Minority_{ij} + \beta_4 Totalvisits_{j} + \beta_5 FRL_{ij} + \varepsilon_i$$

(Model 2)

$$Cognitive_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 Grade_{ij} + \beta_3 Minority_{ij} + \beta_4 Totalvisits_{ij} + \beta_5 FRL_{ij} + \beta_6 Program_{i} + \varepsilon_i$$

where *Affective* is student *i* in subpopulation *j's* standardized score on the Overall Affective Instrument. Cognitive is student *i* in subpopulation *j*' standardized score on either the Awesome Science Assessment or the Arkansas Animals Assessment. I ran separate estimations for each sup-population, *j*. The subpopulations estimated were white students, minority students, girls, boys, students who were a first time visitor to a science center, and students who had previously visited q science center prior to the study. These regressions compared the subpopulation of interest in the control group with the subpopulation of interest in the treatment group.

If the subpopulation of interest is also a predictor variable such as *minority*, I did not include the predictor variable in that particular regression. The modified models were for white students and minority students which did not have the vector dummy variable for minority, and first time and multiple time visitors which did not have the variable *Totalvisits*. *T* is a binary variable equal to 1 if a student is in the treatment group, in the population of interest, and visited the Museum of Discovery during a school field, 0 if a student is in the control group and the population of interest. *Grade* indicates student *i*'s grade (3rd-8th). *Minority* is a dummy variable of 0 if student *i*'s race is white or 1 if student *i*'s race/ethnicity is Hispanic, black, Asian, American Indian, or other. *Totalvisits* is the total number of visits to science centers by student *i* prior to the study. *FRL* is the percent of students on Free or Reduced Lunch at student *i's* school. *Program* is a dummy variable where 1 is for students who completed the *Arkansas Animals* Assessment or 0 if otherwise. Lastly, ε, is a stochastic error clustered by the school group that student *i* came with to visit the science center.

Preview of Chapter Four

In Chapter 4, I estimate the treatment effect for all five scales using ordinary least squares. I also estimate the treatment effect for the subpopulations of interest: minority students, white students, girls, boys, first-time visitors, and multiple-time visitors. I follow the models outlined above for each regression.

Chapter 4: Affective and Cognitive Impacts from Visiting a Science Discovery Center which Included a Model of Direct Instruction

The study reported here used ordinary least squares to determine if students who visited the Museum of Discovery compared to students who had not yet visited the Museum of Discovery on a school field trip showed a difference in content knowledge or attitudes based on the field trip experience. I evaluated these outcomes for the overall group and for minority students, white students, girls, boys, students who are visiting a science center for the first time, and students who have previously visited a science center. As stated in the chapters above, the three research questions are:

1.) After visiting a children's discovery science center during a school field-trip, how do <u>student's attitudes</u> regarding science centers and studying science differ from students who had not yet made such a visit? (see Table 4.1 below)

2.) After visiting a children's discovery science center during a school field-trip, how do <u>students' science knowledge</u> differ from students who had not yet made such a visit? (see Table 4.2 below)

3.) What impacts in attitudes or knowledge does the science center visit offer on certain sub-groups of students? Such populations include: white students, minority students, girls, boys, students who were a first time visitor to a science center, and students who have visited a science center/science museum previously. (see Table 4.3, 4.4, and 4.5 below)

Question 1: Affective Impacts

Table 4.1 provides the effect size and standard error for students who visited the MOD during a school field trip on the Overall Affective Instrument, Interest in Visiting Science

Centers/Museums Subscale, and Interest in Studying Science Subscale compared to students who had yet to visit the MOD on a school field trip.

Table 4.1

Impact of Visiting a Science Center on the Affective Domain

	Ordinary Least Squares Estimations		
	Overall Affective Instrument	Interest in Studying Science Subscale	Interest in Science Center Subscale
Treatment	0.19*** (0.07)	0.10 (0.07)	0.26*** (0.07)
Number of Visits to Science Museums	0.05*** (0.01)	0.04*** (0.01)	0.05*** (0.01)
School FRL %	0.002 (0.001)	0.003** (0.001)	0.00 (0.00)
4th grade	-0.10 (0.08)	-0.06 (0.09)	-0.14 (0.11)
5th grade	-0.49*** (0.07)	-0.39*** (0.09)	-0.49*** (0.10)
6th grade	-0.63*** (0.09)	-0.55*** (0.12)	-0.57*** (0.10)
7th grade	-0.65 (0.38)	-0.61* (0.35)	-0.54 (0.35)
8th grade	-0.70 (0.44)	-0.51 (0.33)	-0.78 (0.50)
Minority Students	-0.08 (0.05)	-0.04 (0.06)	-0.12** (0.05)
N	1830	1830	1830

Each column represents a single regression. Effect sizes are in standard deviation. Grade levels were compared to 3rd grade students. Standard errors in parenthesis and are clustered to the group level (G=32) and $p<0.01$=***, $p<0.05$=**, and $p<0.1$=*, two-tailed.

Students who visited the MOD on average scored 19% of a standard deviation higher on the Overall Affective Inventory ($p<0.05$) than students who had not yet visited the MOD during a school field trip. Also students who visited during a school field trip were 26% of a standard deviation more interested in visiting science centers in the future than students who had yet to visit during a school field trip. Also as students advanced to a higher grade level regardless of if the student was in the treatment or control groups, students became less and less interested in studying science and visiting science centers.

Question 2: Cognitive Impacts

Table 4.2 compares the effect size on the Combined Program Assessment, the Awesome Science Program Assessment, and the Arkansas Animals Program Assessment for students who visited the Museum of Discovery on a school field tip and attended either of these formal programs compared to students who had yet to visit.

Table 4.2

Impact of Visiting a Science Center on the Cognitive Domain

	Ordinary Least Squares Estimations		
	Overall Knowledge Assessment	Awesome Science Assessment	Arkansas Animals Assessment
Treatment	0.21 (0.15)	0.22 (0.14)	0.34** 0.12)
Number of Visits to Science Museums	0.03** (0.01)	0.02 (0.02)	0.05*** (0.01)
School FRL %	0.00 (0.00)	-0.00 (0.01)	-0.00 (0.00)
Minority Student	-0.35*** (0.08)	-0.47*** (0.11)	-0.20 (0.13)
4th grade	0.36** (0.15)	0.28 (0.18)	
5th grade	0.45*** (0.10)	0.48*** (0.11)	
6th grade	0.68*** (0.13)	0.81*** (0.06)	
7th grade	1.01*** (0.29)	1.02*** (0.30)	
8th grade	0.41** (0.15)	0.41*** (0.15)	
Arkansas Animals Program	0.08 (0.13)		
N	1350	915	435

Each column represents a single regression. Effect sizes are in standard deviation. Standard errors are in parenthesis and clustered to the group level (G=23 combined, G=17 Awesome Program, G=6 Animals Program) and p<0.01=***, p<0.05=**, and p<0.1=*, two-tailed.

Only students who visited the MOD during a field trip and attended the Arkansas Animals Program scored significantly higher (34% of a standard deviation) than students who had yet to visit the MOD during a field trip ($p<0.05$). Similar to the affective domain, for each additional visit to a science center, students scored 3% of a standard deviation higher on the knowledge assessment, regardless of having visited the MOD and attending the program or not.

Also, as expected, students scored higher on the knowledge assessment for each advance in grade level. Finally, minority students scored significantly lower on the knowledge assessments (40% of a standard deviation) regardless of if the students visited the MOD and attended either program or not ($p<0.01$).

Question 3: Affective and Cognitive Impacts for several student populations.

Table 4.3, 4.4, and 4.5 compares the effect size of the affective survey and the cognitive assessments for several subpopulations of students who visited the Museum of Discovery on a school field tip compared to that same subpopulation of students who had yet to visit. Table 4.3 provides affective and cognitive outcome estimates for minority students who visited the MOD on a school field trip compared to minority students who had yet to visit and for white students who visited the MOD on a school field trip compared to white students who had yet to visit. Table 4.4 provides affective and cognitive effect estimates for girls who visited the MOD on a school field trip compared to girls who had yet to visit and for boys who visited the MOD on a school field trip with boys who had yet to visit. Table 4.5 provides affective and cognitive effect estimates for students who have never visited a science center until this visit compared to students who had never visited a science center and had yet to visit (first time novices) and for students who had previously visited a science center before this visit compared to students who had also previously visited a science center but had not yet visited as part of this study.

Effects of visiting a science center by race/ethnicity. Visiting a science center has statistically significant ($p<0.05$) and small to moderate impacts on minority students but did not have statistically significant effects on white students, holding all else equal (see Table 4.3 below). After visiting a science center during a school field trip, minority students scored 21% of a standard deviation higher on the affective survey (p<0.01) and 33% of a standard deviation higher on the knowledge assessment than minority students who had not visited the MOD.

Table 4.3

Overall Impacts of Visiting a Science Center by Race

	Ordinary Least Squares Estimations			
	Minority Students		White Students	
	Overall Affective Scale	Overall Knowledge Scale	Overall Affective Scale	Overall Knowledge Scale
Treatment Effect	0.21*** (0.06)	0.33** (0.14)	0.13 (0.09)	0.14 (0.13)
Total Visits	0.03** (0.01)	0.02 (0.18)	0.06*** (0.02)	0.04*** (0.01)
Grade Level	-0.25*** (0.06)	0.17 (0.09)	-0.21*** (0.04)	0.19*** (0.05)
School FRL %	0.002** (0.001)	-0.00 (0.00)	0.004* (0.002)	-0.00 (0.00)
Ark. Animals Program		0.37** (0.15)		0.02 (0.10)
N	602	402	1228	948

Each column represents a single regression. Effect sizes are in standard deviation. Standard errors are in parenthesis and clustered to the group (G) level. For minority students, G=29, affective, and G=21, knowledge. For white students, G=31, affective and G=23, knowledge and p<0.01=***, p<0.05=**, and p<0.1=*, two-tailed.

Effects of visiting a science center by gender. Visiting the MOD during a school field trip did not have any significant impacts on girl students affectively or cognitively (see Table 4.4 below). Boy students who visited the MOD during a school field trip scored 27% of a standard deviation higher on the affective survey than boy students who had yet to visit the MOD (p<0.01) and 40% of a standard deviation higher on the Combined Cognitive Assessment after visiting the MOD and attending either the Arkansas Animals program or Awesome Science program than boy students who had yet to visit (p<0.05).

Table 4.4

Overall Impacts of Visiting a Science Center by Gender

	Ordinary Least Squares Estimations			
	Female Students		Male Students	
	Overall Affective Scale	Overall Knowledge Scale	Overall Affective Scale	Overall Knowledge Scale
Treatment Effect	0.06 (0.07)	0.00 (0.11)	0.27*** (0.09)	0.40** (0.15)
Total Visits	0.05*** (0.01)	0.05*** (0.02)	0.04** (0.02)	0.02 (0.02)
Grade Level	-0.22*** (0.06)	0.17*** (0.09)	-0.20*** (0.05)	0.19*** (0.05)
School FRL %	0.00 (0.00)	0.00 (0.00)	0.005*** (0.002)	-0.00 (0.00)
Minority	-0.04 (0.07)	-0.31*** (0.11)	-0.08 (0.06)	-0.35*** (0.12)
Ark. Animals Program		0.06 (0.08)		0.21 (0.14)
N	964	696	865	653

Each column represents a single regression. Effect sizes are in standard deviation. Standard errors in parenthesis and are clustered to the group level (G). For the affective survey, G=31, and for the knowledge assessment, G=23. p<0.01=***, p<0.05=**, and p<0.1=*, two-tailed.

Effects of visiting a science center for new visitors and returning visitors. Students visiting a science center for the first time as part of the treatment group have a more favorable attitude about science and science centers than students in the control group who have not previously visited a science center, about 20% of a standard deviation ($p<0.05$). Also, first time visitors in the treatment group scored 37% of a standard deviation higher on the knowledge assessments than students who have never visited a science center from the control group ($p<0.05$). However, students in the treatment group who previously visited a science center did not score significantly higher at the 90% confidence level on the overall knowledge assessment than students in the control group who had previously visited a science center but had not yet visited for this study. See Table 4.5 below.

Table 4.5

Overall Impacts of Visiting a Science Center by 1st Time and Multiple Time Visitors

	Ordinary Least Squares Estimation			
	New to Science Centers		Previously Visited a Science Center	
	Overall Affective Scale	Overall Knowledge Scale	Overall Affective Scale	Overall Knowledge Scale
Treatment Effect	0.22**	0.37**	0.20***	0.16
	(0.10)	(0.13)	(0.07)	(0.13)
Grade Level	-0.23***	0.12**	-0.19***	0.22***
	(0.06)	(0.05	(0.03)	(0.05)
Minority	-0.17***	-0.40**	-0.00	-0.31***
	(0.06)	(0.14)	(0.07)	(0.07)
School FRL %	0.00*	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Ark. Animals Program		0.11		0.16
		(0.09)		(0.12)
N	549	430	430	920

Effect sizes are in standard deviation. Standard errors in parenthesis and are clustered to the group level (G). For affective, G=31. For knowledge, G=23. p<0.01=***, p<0.05=**, and p<0.1=*, two-tailed.

Preview of Final Chapter

In chapter five, I will summarize the study design and administration of the surveys and then discuss key findings. I will also compare these findings to the findings of the three RCT studies about the impacts students receive when visiting a science center during a school field trip. Lastly, I will discuss several practical implications and provide suggestions for future research.

Chapter 5: Conclusions, Discussion, and Direction for Future Research

The purpose of this study was to determine the affective impacts that students receive from visiting a science center during a school field trip and the cognitive impacts of attending a short educational program during the visit compared to similar students who had not yet visited the science center during a school field trip. The research questions are:

1) After visiting a children's discovery science center during a school field-trip, how do students' attitudes regarding science centers and studying science differ from students who had not yet made such a visit?
2) After visiting a children's discovery science center during a school field-trip, how do students' science knowledge differ from students who had not yet made such a visit?
3) What impacts in attitudes or knowledge does the science center visit offer on certain sub-groups of students? Such populations include: white students, minority students, girls, boys, students who were a first time visitor to a science center, and students who have visited a science center/science museum previously.

I randomly sorted 31 school groups with 1,830 students into either a treatment group where they visited the Little Rock, Arkansas Museum of Discovery (MOD) during a school field trip or the control group where they had not yet visited the Museum of Discovery during a school field trip but would visit after participating in the study. Several weeks before beginning the study, the MOD had just finished a $9 million renovation transforming the venue from a children's science museum to a more hands-on Exploratorium-style science center. Therefore, many of the study participants who had previously visited the MOD prior to the study would now have a brand-new experience, making this an ideal time period to conduct such a study.

The MOD reservation staff provided a list of the school groups who would be visiting the MOD during the spring and fall months of 2012. Each lead teacher was contacted and asked if the school would be interested in participating in the study. Although 50% of 80 school groups agreed to participate, only 39% (31 school groups) returned the survey materials. Both the control and treatment groups completed the same survey at their school. The survey asked questions about interest in studying science, interest in visiting science centers/museums, and questions from either the Arkansas Animals or Awesome Science formal-style educational programs teachers self-selected for their students to attend. The programs lasted approximately 30 minutes and students were free to explore the science center for the remaining part of their field trip. Survey responses were coded into an Excel spreadsheet and analyzed using ordinary least squares regression in STATA. This chapter discusses the findings for each of the three research questions, overall general conclusions from this study and other similar studies, and areas for future research.

Key Findings

Research question one. This question targeted understanding of issues in the affective domain by focusing on student attitudes about interest in science. The survey (see Appendix C) had two sections addressing these interests. The first section asked about a student's interest in studying science and the second section asked about a student's interest in visiting science centers. The two sections were combined for an overall affective impact as well as analyzed separately. Students who visit a science center on a school field trip have a more overall positive attitude about science (19% of a standard deviation (SD), $p<0.01$), particularly towards a greater interest in visiting science centers again (26% SD, $p<0.01$) than students who have yet to visit the MOD on a school field trip (See Table 4.1). Students who visited the MOD did have a slightly higher interest in studying science (10% of a standard deviation) but this finding is not statistically significant at the 90% confidence level.

The 19% effect size for the overall affective items survey is small by Cohen's (1988) standards. Cohen (1988) categorized effect sizes as small (around 20% of a standard deviation), medium (around 50% of a standard deviation), and large (around 80% of a standard deviation), which he calculated by taking the difference of the two groups' means and dividing by a standard deviation for the data, known as Cohen's *d*. One possible problem with the effect sizes for the affective domain is that this study compares attitudes and interests of students who had just returned from the field trip with attitudes and interests of students who had not yet visited the MOD but were about to visit and knew of this upcoming visit. Knowing about the trip may have increased excitement about the visit, which could reduce the effect sizes. Nonetheless, these findings show that students are at least slightly more excited about visiting science centers and have a greater attitude about science in general after visiting a science center than students who had yet to visit a science center but knew of the upcoming visit.

Research question two. For the cognitive domain, the study focuses specifically on learning gains from participating in a more formal science center lab-based classroom experience. Science centers offer these formal programs as an additional option for school groups, generally for an extra charge. The programs last about 30 minutes. Students who visited the science center and attended either the Arkansas Animals program or the Awesome Science program were asked six questions about their respective programs. Their responses were compared to student responses in the control group who had not yet visited the science center.

Students who attended the Arkansas Animals program outperformed students who had not visited the MOD and attended the program (34% SD, $p<0.05$) (see Table 4.2). The Arkansas

Animals assessment had a small to medium effect size, where on average the students who attended the program while visiting the science center scored about 10% higher on the assessment, holding all else equal, than students who had not yet visited the science center and attended the program. Questions, for example, asked students if they could choose an invertebrate, a reptile, an amphibian out of a list of four possible choices for each question. Only six school groups with a total of 437 students completed the Arkansas Animals Assessment.

Students who attended the Awesome Science program were not statistically different at the 90% confidence level than students who had not yet visited the MOD. Seventeen school groups with 915 students completed the Awesome Science Assessment. For example, questions asked about Bernoulli's Principle, sublimation, why helium floats, and the chemical name of dry ice. The percent of correct answers from both programs was also standardized and then merged to create one assessment, Overall Knowledge Assessment. Students' scores on the combined assessment after visiting the MOD and attending one of these two programs were not statistically different at the 90% confidence level than students' scores who had yet to visit the MOD. Based on the combined program numbers, students who visit a science center and attended a formal educational program do not have a better understanding about the program topic than students who had yet to visit the MOD and attend the program, holding all else equal. However, since the amount of understanding varies by program, more research is needed to determine why one program had a greater impact than the other program.

Research question three. Not only did this study focus on the affective and cognitive impacts for all students but also for various sup-populations: white students, minority students, girls, boys, students who were first time visitors to a science center, and students who had previously visited a science center before the study. All of the outcomes measured where positive, although not necessarily statistically significant at the 90% confidence level. Table 4.3, 4.4, and 4.5 in chapter 4 provides effect sizes and significant levels for each sub population.

Race and ethnicity. Minority students had a more overall positive attitude about science (21% SD, $p<0.01$) and higher achievement levels on the Overall Knowledge Assessment (33% SD, $p<0.05$) than minority students who had yet to visit the MOD, holding all else equal. Minority students include black students (16.5% of the study population), Hispanic students (6.1% of the study population), Asian students (2.1% of the study population), American Indian (3.3% of the study population), and other students (4.5% of the study population). White students who visited the MOD during a school field trip were not statistically different from white students who had yet to visit the MOD on a school field trip at the 90% confidence level (see Table 4.3).

Gender. The visit to the MOD had a larger impact on male students than female students. Male students who visited the MOD had a more overall positive attitude about science (27% SD, $p<0.01$), and higher achievement levels on the Overall Knowledge Assessment (40% SD, $p<0.05$) than male students who had not yet visited the MOD as part of a school field trip, holding all else equal. Female students who visited the MOD were not statistically different than female students who had yet to visit the MOD on a school field trip on either the Overall Attitude Scale or the Overall Knowledge Assessment at the 90% confidence level (see Table 4.4)

First time visitors and students who previously visited a science center. The MOD reopened after several months of an intensive renovation just a few weeks before the start of this study which created a new experience even for students who had previously been to the MOD. However, one subpopulation of interest is those students who had never visited any science center at all, either as part of a school field trip or outside of school. These first-time visitors had a more overall positive attitude about science (22% SD, $p<0.05$) and scored higher on the Overall Knowledge Assessment (37% SD, $p<0.05$) than other students who had yet to visit any science center. Also for students who had visited a science center already and were in the treatment group, they scored 20% of a standard deviation higher than students who had previously visited a science center and were in the control group. The comparable effect sizes on the Overall Affective Instrument between first time visitors to a science center and those of previous visitors to a science center suggest that the experience of visiting the newly renovated MOD exhibits was the same for both groups, while only new visitors showed a statistically significant positive difference on the knowledge assessment at the 95% confidence level.

General Conclusions

During the 1980s, the students who visit an ISEI brought with them "millions of television images and a flare for the dramatic" (Ambach, 1986, p. 36) while students today bring hours of virtual reality, Google, and a profound connection with their smartphones. Even though the learning environment and technology has advanced, limited quality research exists that suggests students learn from visiting ISEIs, specifically when those visits are part of a short, one-day school field trip that are often structured, guided, and controlling. This study's findings match the findings of two of the three RCT studies, Holmes (2011) and Itzek-Greulich (2015). The average student does not gain knowledge by attending an educational program at an ISEI compared to what they are already learning in the classroom. However, these formal off-site programs are cognitively beneficial for males, minorities, and first-time visitors.

When teachers take students to science centers, students are often closely monitored and unable to freely roam, have limited choice on what they want to study in detail, and are often unable to choose their own learning groups (Eshach, 2007; Griffin 2004). This could even be more pronounced when a researcher is involved. Of the seventeen studies included in the systematic review, all 17 studies included a formal-learning style activity. These formal-learning activities included a guided tour (Ballouard et al., 2012; Basten et al., 2014; Futer, 2005; Holmes, 2011; Sentürk & Özdemir, 2014; Stavrova & Urhahne, 2010; Sweet, 2014), work-station tasks (Holmes, 2011; Itzek-Greulich et al., 2015; Jarvis & Pell, 2002; Jarvis & Pell, 2005; Kamarainen et al., 2013; Prokop et al., 2007; Puhek et al., 2012; Stavrova & Urhahne, 2010; Strum & Bogner, 2010; Wilde & Urhahne, 2008), and/or lectures with demonstrations (Freedman, 2010; Sentürk & Özdemir, 2014).

Unfortunately only two RCT studies from the systematic review reported affective benefits, and of those, Holmes (2011) had students in the control groups complete the survey after arriving at the site location. At this point, both groups of students were equally excited about science, so any affective benefits compared to staying in the classroom is unclear. The other study, Prokop et al. (2007), did not calculate a treatment effect between the control group and treatment group, so again it is unclear what the effect size is and if that effect size is significant. Regardless of methodology or of the degree of structure in an ISEI, studies reported that students on field trips who visit these ISEIs (and depending on the specific ISEI) are more accepting of unpopular organisms such as snakes (Ballouard et al., 2012); are more excited about learning science in a tropical forest (Futer, 2005); have lowered anxiety and created a stronger social context (Jarvis & Pell, 2005), are better able to understand focal topics and complete science related skills (Kamarainen et al., 2013); a greater interest in biology as a school subject, the natural environment, and future biology work (Prokop et al, 2007); a positive change in overall attitude (Sentürk & Özdemir, 2014); and a greater desire to do science experiments at home, become a scientist, and are less frightened about going on a trip to space (Jarvis & Pell, 2005). However, these findings do not tell us if there is a difference in attitude and motivation between students who visited these venues and students who stayed in the classroom. Again, any classroom lab, project, movie, virtual field trip, or other activity could have had a similar impact without the students actually needing to take a field trip.

Future Research

Benefits secondary students receive from visiting a science center. The majority of those who have studied museum learning have called for more research. As Andre et al. (2017) note, students in elementary schools learn differently than secondary students. We know very little about elementary student benefits to visiting ISEIs but even less about secondary students. Even after eighty years of research on the benefits of school field trips, most questions about the affective and cognitive benefits are not answered, particular for secondary students. So, how do ISEIs cater most effectively to both groups of students? Do they even cater to both groups of students? What types of field trips are secondary teachers choosing if any? What are the benefits of field trips for secondary students compared to elementary students?

After reviewing the registration information that the Museum of Discovery provided, the overwhelming majority of reservations are from primary and elementary school groups. Only four secondary groups were listed out of 84 total school groups. Secondary students may not take the same number of academic field trips as those students in the elementary or middle school grades or the teachers may choose to visit other venues than science centers. Nonetheless, part of the $9 million renovation was to make the MOD more appealing to adults and older children (Tidwell, 2011).

When Andre et al. (2017) conducted a systematic review on the research about school field trips to all types of museums; they excluded articles that focused on students above the age of 12. They believe secondary student experiences are qualitatively different from those of younger children. I did not intentionally exclude studies in the systematic review but of the 17 studies, none of the studies evaluated the affective domain of students past 6th grade. For studies investigating cognitive outcomes, only three studies (KrombaB & Harms, 2008; Puhek et al., 2012; Stavrova & Urhahne, 2010) evaluated field trip outcomes on students between grades 7 and 9. Only one of these studies (Puhek et al., 2012) used a control group (although the groups were not randomly created). All three studies reported positive achievement gains between the pre- and post-assessments.

With the developments and technological improvements in virtual reality, many secondary schools may choose virtual experiences over taking students to a real field trip site. Puhek et al. (2012) and Sweet (2014) compared learning outcomes between students who visited an ISEI and students who stayed in the classroom and visited the ISEI through a virtual learning experience. Both studies found positive impacts by the virtual experience and the real experience, however, students who actually went on the field trips had larger gains than those who experienced a virtual reality. Neither study compared students' attitudes between the two groups nor did they actually estimate to see if the difference in outcomes between the two groups were significant. Also, Sweet (2014) studied 3rd graders and Puhek et al. (2012) studied 8th graders. To better understand the benefits of school field trips, we need more information about secondary student field trip practices and experiences.

Finding the right ISEI to spark female science interest. A second topic not adequately addressed in the literature is what type of ISEI has the greatest impact for creating interest in studying science by girls. Science educators have known for some time of a disparity between female and male students regarding science interest, motivation, and future careers as scientists which have led to the publications of an abundant amount of research on females and science (Carlone, Webb, Archer, & Taylor, 2015). More recent studies (Archer, DeWitt, & Willis, 2014; Broadway & Leafgren, 2012; Hughes, 2001) have focused on the characteristics and types of boys who have long-term science aspirations. Other studies have confirmed the global perception of a scientist as a white male with crazy, untamed hair wearing a lab coat and glasses surrounded by beakers and other dangerous chemicals (Barman, 1999; Koran & Bar, 2009; Song & Kim, 1999). Ultimately, the term ‘scientist’ conjures images of Albert Einstein.

Osborne, Simon, and Collins (2003) reviewed the literature from the last 20 years on gender and attitudes towards science. One of their key findings is the amount of evidence showing a decline in students choosing to pursue scientific careers, and that number is lower for females. For example, in England, the male to female ratio in secondary advanced physics classes is to 3.4 to 1 compared to advanced biology which is 1 to 1.6 female. Also, the number of students taking advanced physics has declined from 45,000 to 30,000 between 1990 and 2000, while the number of students in advanced biology has stayed consistently around 50,000 during this time period. So, although some science subjects such as physics and chemistry attract more boys than girls, the decline in interest in the physical sciences is not a just a ‘female’ concern. Based on research findings, females are not all that interested in studying these subjects in the first place, so a decline in students enrolling in advanced physical science classes suggests even males are losing interest in future careers in the physical sciences.

From this study here, boy students who visited the MOD had larger and more significant outcomes than girl students across both the affective and cognitive domains. On the two outcomes evaluated (overall attitude in science and greater understanding about a program’s topic), boy students had moderate, positive, and significant outcomes on both, while girl students did not have any significant outcomes at the 90% confidence level. Instead of changing science centers to cater to girls, researchers need to determine if there are any specific types of ISEIs that empower girls to want to be scientists so schools can provide field trips experiences to those types of ISEIs.

First time visitors (aka novelty effect). Holmes (2011) conducted a randomized control treatment study using pre/post analysis of 228 students who visited the IDEA Place at the Louisiana Tech University Children's Science Museum. Students who were in the control group completed the assessment the moment they arrived at the science center while the treatment group completed the assessment after visiting the science center. Holmes did not find any statistically significant differences between the control group who had yet to visit the science center and the treatment groups who experienced either just the field trip or a field trip with a lesson. Holmes partially contributes this insignificant finding to the novelty effect. When students are unfamiliar with a setting they typically learn less than students who are familiar with the setting and are unable to focus on specific tasks assigned (Balling & Falk, 1980).

Several questions on both the spring survey and fall survey in this study asked students how many times they have visited a science center on a field trip and then how many times they have visited a science center with family or outside of school. All students in the control group who said they had never visited a science center are considered first time visitors. All the students in the treatment group who reported they had only visited a science center once (since they just went on the school field trip to a science center) are also considered first time visitors. When comparing first time visitors who just experienced the MOD with students who were about to visit the MOD for the first time, the results suggest these students received some cognitive and affective benefits from visiting the MOD. Compared to students in the control group who have never visited a science center, those students in the treatment group who just visited a science center, the MOD, for the first time and as part of a school field trip had more positive attitudes about science (22% SD, $p<0.05$) and demonstrated more knowledge after attending an educational program (37% SD, $p<0.05$). Before this study, the MOD transformed from a children's natural history museum to an Exploratorium-style science center. Both first-time visitors in the treatment group and students in the treatment group who had previously visited a science center had a similar effect size which suggests that the renovations did create a new experience even for return visitors. If a novelty effect does exist than the effect size would be larger after the next visit for all students. Researchers in the 1970s and early 80s suggest students become so overwhelmed with the experience that very little learning occurs (Balling & Falk, 1980; Falk, Martin, & Balling, 1978; Martin, Falk, and Balling, 1981).

Multiple ISEIs in a single study. Lastly, future research must include studies that evaluate multiple ISEIs using the same instruments. Although each ISEI is different and many offer specific educational programs catered to their specific visitor population, some benefits should carry over from one ISEI to another: interest in science, motivation, levels of engagement, and understanding of the nature of science. Also, are there any differences in visiting different types of ISEIs? What types of benefits, particularly affective, do students receive from visiting a zoo verses visiting a science center or by conducting an experiment at a local pond? Clearly each ISEI type would offer a different set of affective and cognitive outcomes but what are those specific sets of outcomes? For example, visiting the MOD did not have a significant impact on student attitudes about studying science in the future at the 90% confidence level. Does visiting a zoo, conducting experiments at a nature park, or visiting a different science center have a similar, little-to-no effect, on increasing student interest in studying science?

Conclusion

Continuing research on science centers and other ISEIs is important but the research is meaningless if these findings are based on weak study designs. This RCT study suggests affective benefits exist for students who visit a science center outside of the benefits they are receiving in the classroom. Increasing student knowledge from a field trip visit depends in part on the program the teacher chooses for the students to attend. In this study, students had a greater knowledge increase after attending the Arkansas Animals program compared to students who had not yet visited the MOD. Students who attended the Awesome Science program had similar outcomes as other students who stayed at school. For minority students, schools may provide the only opportunities for visiting places such as science centers, and this study found that minority students have moderate cognitive and affective benefits by such field trips, more so than the average student. Also, visiting science centers seems to have a greater impact on boys than on girls. Lastly, if a novelty effect exists, it was not demonstrated in this study. First time visitors in the treatment group actually scored better on the knowledge assessment than students in the control group who have never visited a science center before. Also, both first time visitors and repeat visitors from the treatment groups scored similarly on the Overall Affective Instrument suggesting that the MOD renovations did in fact create a new experience for returned visitors as well. However, to draw any general conclusions about the short-term impacts of visiting ISEIs, more rigorous studies are necessary.

References Cited

Ambach, G. (1986). Museums as places of learning. *Museum News*, *65*(2), 35-41.

American Alliance of Museums. (2014, April 10). *Testimony of Ford W. Bell, DVM, President of the American Alliance of Museums to the House Appropriations Subcommittee on Interior, Environment & Related Agencies NEH, NEA, the Smithsonian Institute and Historic Preservation Programs.* Retrieved from http://www.aam-us.org/docs/default-source/advocacy/fy15-bell-aam-interior-testimony.pdf?sfvrsn=2

Anderson, D., & Lucas, K.B. (1997). The effectiveness of orientating students to the physical features of a science museum prior to visitation. *Research in Science Education*, *27*(4), 485-95.

Andre, L. Durkson, T., & Volman, M.L. (2017). Museums as avenues of learning for children: A decade of research. *Learning Environmental Research, 20,* 47-76.

Archer, L., DeWitt, J., & Willis, B. (2014). Adolescent boys' science aspirations: Masculinity, capital, and power. *Journal of Research in Science Teaching, 51,* 1-30.

Association of Science Technology Centers (1996). *An invisible infrastructure: Effective summary.* Retrieved from http://www.astc.org/resource/education/infrastructure_sum.htm

Association of Science-Technology Centers (2014). *2013 science center and museum statistics.* Retrieved from http://www.astc.org/wp-content/uploads/2014/10/2013-Science-Center-Statistics.pdf

Balling, J., & Falk, J. (1980). A perspective on field trips: Environmental effects on learning. *Curator, 23,* 229-240.

Ballouard, J., Provost, G., Barre, D., & Bonnet, X. (2012). Influence of a field trip on the attitude of school children toward unpopular organisms: An experience with snakes. *Journal of Herpetology, 46,* 423-428.

Bamberger, Y., & Tal, T. (2008). Multiple outcomes of class visits to natural history museums: The students' view. *Journal of Science Education and Technology, 17,* 274-284.

Barman, C. (1999). Students' views about scientists and school science: Engaging K-8 teachers in a national study. *Journal of Science Teacher Education, 10*(1), 43-54.

Basten, M., Meyer-Ahrens, I., Fries, S., & Wilde, M. (2014). The effects of autonomy-supportive vs. controlling guidance on learners' motivational and cognitive achievement in a structured field trip. *Science Education, 38,* 2114-2132.

Behrendt, M., & Franklin, T. (2014). A review of research on school field trips and their value in education. *International Journal of Environmental & Science Education, 9,* 235-245.

Birney, B. A. (1988). Criteria for successful museum and zoo visits: Children offer guidance. *Curator,* 31, 292-316.

Bitgood, S. (1989). School field trips: An overview. *Visitor Behavior, 4*(2), 3-6.

Bitgood, S., Serrell, B., & Thompson, D. (1994). The impact of informal education on visitors to museums. In V. Crane et al. (Eds.), *Informal science learning: What the research says about television, science museums, and community-based projects*. Massachusetts: Research Communications Ltd.

Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.

Blosser, P.E., & Helgeson, S.L. (Eds.). (1986). *Investigations in science education* (Vol. 12). Columbus, OH: University of Ohio.

Broadway, F. S., & Leafgren, S. L. (2012). Unmasking: On violence, masculinity, and superheroes in science education. *Cultural Studies of Science Education, 7*, 719-733.

Burchenal, M. K., & Grohe, M. (2008). Reimagining school programs. *Visual Arts Research*, *34*, 66-75.

Bustillos, D., & Thornock, B. (2013). Marking Twain. *Journal of Medical Humanities*, *34*(4), 455-458.

Cameron, A.C., & Miller, D.L. (2015). *A practitioner's guide to cluster-robust inference*. Davis, CA: University of California. Retrieved from http://cameron.econ.ucdavis.edu/research/Cameron_Miller_JHR_2015_February.pdf

Carlone, H.B., Webb, A.W., Archer, L., & Taylor, M. (2015). What kind of boy does science? A critical perspective on the science trajectories of four scientifically talented boys. *Science Education*, *99,* 438-464.

Cleaves, A. (2005). The formation of science choices in secondary school. *International Journal of Science Education*, *27*(4), 471-486.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Columbus Dispatch (2012). School field trips in decline amid standardized testing. Retrieved from http://www.dispatch.com/content/stories/local/2012/10/13/school-field-trips-in-decline-amid-standardized-testing.html

Cooper H.M. (1984). The integrative research review: A systematic approach. *Applied Social Research Methods Series* (Vol. 2). Beverly Hills, CA: Sage.

Corpus, J. H., McClintic-Gilbert, M. S., & Hayenga, A. O. (2009). Within-year changes in children's intrinsic and extrinsic motivational orientations: Contextual predictors and academic outcomes. *Contemporary Educational Psychology, 34*, 154-166.

Crane, V. (1994). *Informal science learning: What the research says about television, science museums, and community-based projects*. Dedhma, MA: Research Communications.

Csikszenmihayli, M., & Hermanson, K. (1995). Intrinsic motivation in museums: What makes visitors want to learn? *Museum News*, *74*(3), 35-37 and 59-62.

Davidson, S.K., Passmore, C., & Anderson, D. (2010). Learning on zoo field trips: The interaction of the agendas and practices of students, teachers, and zoo educators. *Science Education, 94*(1), 122-141.

Dewitt, J., & Storksdieck, M. (2008). A short review of school field trips: Key findings from the past and implications for the future. *Visitor Studies, 11*, 181-197.

Dierking, L. (1991). Learning theory and learning styles: An overview. *Journal of Museum Education,16(1), 4-6.*

Dierking, L. D., Burtnyk, K., Buchner, K. S., & Falk, J. H. (2002). Visitor learning in zoos and aquariums: A literature review. Retrieved from http://www.aza.org/ConEd/VisitorLearning/Documents/VisitorLearning LiteratureReview.pdf

Dillon, J. (2007). School site to museum floor: How informal science institutions work with schools. *International Journal of Science Education, 29*, 1489-1507.

Dohn, N. (2013). Upper secondary students' situational interest: A case study of the role of a zoo visit in biology class. *International Journal of Science Education, 35*(16), 2732-2751.

Dorsen, J., Carlson, B., & Goodyear, L. (2006). Connecting informal STEM experiences to career choices: Identifying the Pathway. *ITEST Learning Resource Center*. Retrieved from http://stelar.edc.org/sites/stelar.edc.org/files/itestliteraturereview06.pdf

Effective Public Health Practice Project (2010). *Quality assessment tool for quantitative studies.* Retrieved from http://www.ephpp.ca/PDF/Quality%20Assessment%20Tool_2010_2.pdf

Eshach H (2007) Bridging in-school and out-of-school learning: Formal, non-formal and informal education. *Journal of Science Education and Technology, 16*, 171-190

Falk, J. (2013). Understanding museum visitor's motivations and learning. *Motivation and Learning Styles*. Seminar ARKEN Museum of Modern Art: 106-130. Retrieved from https://slks.dk/fileadmin/user_upload/dokumenter/KS/institutioner/museer/Indsatsomraader/Brugerundersoegelse/Artikler/John_Falk_Understanding_museum_visitors__motivations_and_learning.pdf

Falk, J.H., Martin W.W., & Balling, J.D. (1978). The novel field trip phenomenon: Adjustment to novel settings interferes with task learning. *Journal of Research in Science Teaching, 15*(2), 127-134.

Falk, J. H., & Storksdieck, M. (2005). Using the contextual model of learning to understand visitor learning from a science center exhibition. *Science Education, 89*, 1-35.

Freedman, M.R. (2010). A "healthy pizza kitchen" nutrition education program at a children's health museum. *Journal of Nutrition Education and Behavior, 42*, 353-354.

Futer, M. (2005). *Evaluating the effectiveness of environmental education essential elements in school field trip programming*. (Master thesis). Retrieved from Library and Archives Canada.

Garnett, R. (2001). *The impact of science centers/musuems on their surrounding communities.* Kingston, Australia: Questacon.

Gebotys, R. (2011). Handout on validity. *Wilfrid Laurier University*. Retrieved from http://legacy.wlu.ca/docsnpubs_detail.php?grp_id=247&doc_id=45759

Genaux-Hauser, A. (2010). *Literature Review. Science Experiments in the Rochester, NY Area*. Retrieved from https://sites.google.com/site/scienceinrochesterny/home/literature-review

Glick, M.P., & Samarapungavan, A. (2008). Wolves are beautiful and proud: Science learning from a school field trip. *Journal of Museum Education, 33*, 199-207.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549-576.

Greene, J.P., Kisida, B., & Bowen, D.H. (2014a) The benefits of culturally enriching field trips. *Education Digest*, *78*(8), 4-13

Greene, J.P., Kisida, B., & Bowen, D.H. (2014b). The educational value of field trips. *Education Next, 14*(1), 78-86

Griffin, J. (2004). Research on students and museums: Looking more closely at the students in school groups. *Science Education 88*, S59-S70.

Griffin, J. & Symingotn, D. (1997). Moving from task-oriented to learning-orientated strategies on school excursions to museums. *Science Education, 8,:* 763-779.

Gutwill, J. P., & Allen, S. (2012). Deepening students' scientific inquiry skills during a science museum field trip. *The Journal of the Learning Sciences, 20,* 130-181.

Harris, A.D., McGregor, J.C., Perencevich, E.N., Furuno, J.P., Zhu, J., Peterson, D.E., & Finkelstein, J. (2006). The use of interpretation of quasi-experimental studies in medical informatics. *Journal of American Medical Informatics Association, 13*(1), 16-23.

Henry, C.K. (1992). Retention and recall of images: Evaluating museum experiences of School children. *Visual Arts Research, 18*(2), 82-92.

Hodkinson, P., Colley, H., & Malcolm, J. (2003). The interrelationships between informal and formal learning. *Journal of Workplace Learning, 15*, 313-318.

Holmes, J.A. (2011). Informal learning: Student achievement and motivation in science through museum-based learning. *Learning Environment Research*, *14*, 263-277.

Honeycutt Spears, V. (2014). Stricter field trip policy gets Fayette school board approval. *Lexington Herald-Leader*. Retrieved from http://www.kentucky.com/news/local/education/article44468646.html

Hughes, G. (2001). Exploring the availability of student scientist identities within curriculum discourse: An anti-essentialist approach to gender-inclusive science. *Gender and Education, 13,* 275 – 290.

IDEA (2011). *Difference between a science museum and a science center.* Retrieved from http://www.idea.org/blog/2011/03/31/difference-between-a-science-museum-and-a-science-center/

Itzek-Greulich, H., Flunger, B., Vollmer, C., Nagengast, B., Rehm, M., & Trautwein, U. (2015). Effects of a science center outreach lab on school students' achievement: Are student lab visits needed when they teach what students can learn at school? *Learning and Instruction, 38,* 43-52.

Jarvis, T., & Pell, A. (2002). Effect of the challenger experience on elementary children's attitudes to science. *Journal of Research in Science Teaching, 39*, 979-1000.

Jarvis, T., & Pell, A. (2005). Factors influencing elementary school children's attitudes toward science before, during, and after a visit to the UK National Space Center. *Journal of Research in Science Teaching, 42*, 53-83.

Jones, L.S. (1997). Opening doors with informal science: Exposure and access for our underserved students. *Science Education*, *81*(6), 663-677.

Kamarainen, A.M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M.S., & Dede, C. (2013). EcoMOBILE: Integrating augmented reality and probeware with education field trips. *Computers & Education, 68*, 545-556.

Kirby, S. (2012). The state of women of color in the United States. *Center for American Progress*. Retrieved from https://www.americanprogress.org/issues/race/reports/2012/07/17/11923/the-state-of-women-of-color-in-the-united-states/

Kisiel, J.F. (2005). Understanding elementary teacher motivations for science field trips. *Science Education, 89,* 936-955.

Kisiel, J. F. (2010). Exploring a school-aquarium collaboration: An intersection of communities of practice. *Science Education, 94,* 95-121.

Koran, J.J., Koran, M.L., & Ellis, J. (1989). Evaluating the effectiveness of field experiences: 1939-1989. *Visitor Behavior, 4,* 7-10.

Koren, P., & Bar, V. (2009). Pupils' images of 'the scientist' among two communities in Israel: A comparative study. *International Journal of Science Education, 31,* 2485-2509.

Kowalczyk, N., & Truluck, C. (2013). Literature reviews and systematic reviews: What is the difference? *Radiologic Terminology, 85,* 219-222.

Kremer, K.B. & Mullins, G.W. (1992). Children's gender behavior at science museum exhibits. *The Curator: The Museum Journal,* 35(1), 39-48.

Krepel, W. J., & DuVall, C. R. (1981). *Field trips: A guideline for planning and conducting educational experiences.* Washington, DC: National Science Teachers Association.

KrombaB, A., & Harms, U. (2008). Acquiring knowledge about biodiversity in a museum: Are worksheets effective? *Journal of Biological Education, 42,* 157-163.

Kubota, C. A., & Olstad, R.G. (1991). Effects of novelty-reducing preparation on exploratory behavior and cognitive learning in a science museum setting. *Journal of Research in Science Teaching, 28*(3): 225-234.

Leachman, M., Masterson, K., & Figueroa, K. (2017). A punishing decade for school funding. *Center on Budget and Policy Priorities*. Retrieved from https://www.cbpp.org/research/state-budget-and-tax/a-punishing-decade-for-school-funding

Leeming, F.C., Dwyer, W.O., Porter, B., & Cobern, M. (1993). Outcome research in environmental education: A critical review. *Journal of Environmental Education, 24,* 8-21.

Madill, H. M., Ciccocioppo, A.L., Stewin, L., Armour, M.A., & Montgomerie, T.C. (2004). The potential to develop a career in science: Young women's issues and their implications for careers guidance initiatives. *International Journal for the Advancement of Counseling, 26(*1), 1-19.

Maltese, A. V., & Tai, R. H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education, 32*(5), 669-685.

Martin, W.W., Falk, J.H., & Balling, J.D. (1981). Environmental effects on learning: The outdoor field trip. *Science Education*, *65*(3), 301-309.

Martinello, M.L., & Kromer, M.E. (1990). Developing and assessing lower-SES Hispanic children's inferential thinking through a museum-school program. *Journal of Elementary Science Education, 2*(2), 21-36.

McComas, W.F. (2006). Science teaching beyond the classroom: The role and nature of informal learning environments. *The Science Teacher, 73*(1), 26-30.

McComas, W.F., Cox-Peterson, A., Olson, J., & Narguizian, P. (n.d.) *Research supported statements related to learning in informal/free choice environments such as science , field sites, and zoos.* Fayetteville, AR; University of Arkansas. Retrieved from https://pase.uark.edu/_resources/images/research.pdf

Melber, L.M. (2008). Informal learning and field trips: Engaging students in standards-based experiences across the K-5 curriculum. Thousands Oak, CA: Corwin Press.

Meldrum, M.L. (2000). A brief history of the randomized controlled trial: From oranges and lemons to the gold standard. *Hematology/Oncology Clinics of North America*, *14*(4), 745-760.

Moorehead, C. (n.d.) Types of experimental designs handout. *Dartmouth.* Retrieved from www.dartmouth.edu/~oir/docs/Types_of_Experimental_Designs_Handout.doc

Morag, O., & Tal, T. (2012). Assessing learning in the outdoors with the field trip in natural environments (FiNE) framework. *International Journal of Science Education*, *34*, 745-777).

Murdoch, M., Simon, A.B., Polusny, M.A., Bangerter, A.K., Grill, J.P., Noorbaloochi, S., & Partin, M.R. (2014). Impact of different privacy conditions and incentives on survey response rate, participant representativeness, and disclosure of sensitive information: A randomized control trail. *BMC Medical Research Methodology*, *17,* 1.

Museum of Discovery (n.d). *About.* Retrieved from https://www.museumofdiscovery.org/about.

Museum of Discovery (2013). *Your guide to discovery outside the classroom: 2013-2014 Program Guide.* Retrieved from http://museumofdiscovery.org/sites/default/files/2013-2014%20Program%20Guide_1.pdf

Nabors, M.L. Edwards, L.C., & Murray, R.K. (2009). Making the case for field trips: What research tells us and what site coordinators have to say. *Education, 129*, 661-667.

National Research Council (2009). *Learning science in informal environments: People, places, and pursuits.* Washington, DC: The National Academies Press.

National Science Board (2016). *Science and engineering indicators 2016: A broad base of quantitative information on the U.S. and international science and engineering enterprise.* Retrieved from https://nsf.gov/statistics/2016/nsb20161/#/

National Science Board. (2007). Science, technology, engineering, and mathematics (STEM) education issues and legislative options. In R. Nata (Ed.), *Progress in education* (Vol. 14, pp. 161-189). Washington, DC: Author.

Nazier, G.L. (1993). Science and engineering professors: Why did they choose science as a career? *School Science and Mathematics, 93, 3*21–324.

New Jersey School Board Association. (2012, Dec 18). *Survey results: Schools keep field trips intact.* Retrieved from http://www.njsba.org/news/sbn/20121218/survey-results-schoolskeep-field-trips-intact.php

Oakley, A., Gough, D., Oliver, S., & Thomas, J. (2005). The politics of evidence and methodology: Lessons from the EPPI-Centre. *Evidence & Policy, 1*(1), 5-31.

OECD (2012). *PISA in focus: Are students more engaged when schools offer extracurricular activities*. Retrieved from https://www.oecd.org/edu/pisa%20in%20focus%20n18%20(eng)--v05.pdf

Ohlandt, K. (2017). Field trip funding bill would give relief to low-income schools. *Exceptional Delaware 2018*. Retrieved from https://exceptionaldelaware.wordpress.com/2017/12/15/field-trip-funding-bill-would-give-relief-to-low-income-schools/

Organization for Economic and Development (2012). Are students more engaged when schools offer extracurricular activities? *PISA in Focus.* Retrieved from https://eric.ed.gov/?id=ED533439

Orion, N., & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a national environment. *Journal of Research in Science Teaching, 31,* 1097-1119.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049-1079.

Pace, S., & Tesi, R. (2004). Adult's perception of field trips taken within grades K-12: Eight case studies in the New York metropolitan area. *Education, 125*(1), 30-40.

Phillips, M., Finkelstein, D., & Wever-Frerichs, S. (2007). School site to museum floor: How informal science institutions work with schools. *Journal of Science Education, 29*, 1489–1507.

Plummer, M. (2014). Long Beach museum looks to buck trend of declining student field trips. *Southern California Public Radio*. Retrieved from http://www.scpr.org/blogs/education/2014/10/10/17383/long-beach-museum-looks-to-buck-trend-of-declining/

Porta, M. (2014). *A Dictionary of Epidemiology* (4th ed.). Oxford: Oxford University Press.

Porter, T.M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton, NJ: Princeton University Press.

Prather, J. P. (1989). Review of the value of field trips in science instruction. *Journal of Elementary Science Education, 1*, 10-17.

Preisendörfer, P., & Wolter, F. (2014). Who is telling the truth? A validation study on determinants of response behavior in surveys. *Public Opinion Quarterly, 78*(1), 126-146.

Price, S., & Hein, G.E. (1991). More than a field trip: Science programmes for elementary school groups at museums. *International Journal of Science Education, 13*(5), 505-519.

Prokop, P., Tuncer, G., & Kvansičák, R. (2007). Short-term effects of field programme on students' knowledge and attitude toward biology: A Slovak experience. *Journal of Science Education and Technology, 16*, 247-255.

Puhek, M., Perse, M., & Sorgo, A. (2012). Comparison between a real field trip and a virtual field trip in a nature preserve: Knowledge gained in biology and ecology. *Journal of Baltic Science Education, 11,* 164-174.

Ramey-Gassert (1997). Learning science beyond the classroom. *The Elementary School Journal, 97,* 433-450.

Ramey-Gassert, L., Walberg., H. J., III, & Walberg, H. J. (1994). Re-examining connections: Museums as science learning environments. *Science Education,78,* 345-363.

Rennie, L.J., & Johnston, D. J. (2004). The nature of learning and its implications for research in learning from museums. *Science Education, 88,* S4-S16.

Rennie, L.J., & McClafferty, T.P. (1995). Using visits to interactive science and technology centers, museums, aquaria, and zoos to promote learning in science. *Journal of Science Teacher Education, 6*, 175–185.

Rhan, M. (2017). Factor analysis: A short introduction, part 2-rotations. *The Analysis Factor*. Retrieved from https://www.theanalysisfactor.com/rotations-factor-analysis/

Rossi, P.H., Lipsey, M.W., & Freeman, H.E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage Publications.

Rudmann, C.L. (1994). A review of the use and implementation of science field trips. *School Science and Mathematics, 94*, 138-141.

Scarce, R. (1997). Field trips as short term experiential education. *Teaching Sociology*, *25*, 219-226.

Semper, R. J. (1990). Science museums as environments for learning. *Physics Today, 43*, 2-8.

Sentürk, E., & Özdemir, Ö. F. (2014). The effect of science centres on students' attitudes towards science. *International Journal of Science Education, Part B: Communication and Public Engagement, 4,* 1-24.

Shapiro, L. (1990, May 28). Guns and dolls. *Newsweek*. Retrieved from http://www.geocities.ws/pippin_ph/guns.htm

Sibbald, B., & Roland, M. (1998). Understanding controlled trials: Why are randomized control trials important? *BMJ, 316,* 201.

Song, J., & Kim, K. S. (1999). How Korean students see scientists: The images of the scientists. *International Journal of Science Education, 21*, 957-977.

Sparks, S. D. (2011, April 6). Researchers play catch-up in gauging beyond-school effects. *Education Week*. Retrieved from https://www.edweek.org/ew/articles/2011/04/06/27is-research.h30.html

Stavrova, O., & Urhahne, D. (2010). Modification of a school programme in the Deutsches Museum to enhance students' attitudes and understanding. *International Journal of Science Education, 32*, 2291-2310.

Sturm, H., & Bogner, F.X. (2010). Learning at workstations in two different environments: A museum and a classroom. *Studies in Educational Evaluation, 36*, 14-19.

Swan, D.W. (2014). Children who visit museums have higher achievement in reading, math, and science. *Institute of Museum and Library Services. Retrieved* from https://www.imls.gov/news-events/upnext-blog/2014/04/children-who-visit-museums-have-higher-achievement-reading-math-and

Sweet, A.M. (2014). *The effectiveness of virtual and on-site dairy farm field trips to increase student knowledge in science, social studies, and health and wellness standards.* (Master's thesis). UMI Dissertation. 1573767.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning Early for Careers in Science. Science, 312(5777), 1143-1144.

Terrero, N. (2012, Jan. 11). School budget cuts lead to less field trips. *NBC Latino*. Retrieved from http://nbclatino.tumblr.com/post/15688960472/school-budget-cuts-lead-to-less-field-trips

Tidwell, B. (2012, Jan. 11). The Museum of Discovery grows up. *Arkansas Times*. Retrieved from https://www.arktimes.com/arkansas/the-museum-of-discovery-grows-up/Content?oid=2012461

Torgerson, C.J., & Torgerson, D.J. (2001). The need for randomized controlled trails in educational research. *British Journal of Educational Studies*, *49*(3), 316-328.

Trochim, W.M.K. (2006). Experimental design. *Research Methods Knowledge Base.* Retrieved from http://www.socialresearchmethods.net/kb/desexper.php

U.S. Bureau of Labor Statistics (2014). STEM 101: Intro to tomorrow's jobs. *Occupational Outlook Quarterly.* Author: D. Vilorio. Retrieved from https://www.bls.gov/careeroutlook/2014/spring/art01.pdf

U.S. Department of Education, National Center for Education Statistics. (2016). *Digest of Education Statistics, 2015* (NCES 2016-014). Retrieved from https://nces.ed.gov/programs/digest/d15/ch_1.asp

U.S. Department of Education. (2007). *Report of the Academic Competitiveness Council.* Washington, DC: Author.

Weinburgh, M. (1995). Gender differences in student attitudes towards science: A meta-analysis of the literature from 1970- 1991. *Journal of Research in Science Teaching, 32*, 387-398.

Weinstein, M., Whitesell, E.R., & Schwartz, A.E. (2014). Museums, zoos, and gardens: How formal-informal partnerships can impact urban students' performance in science. *Evaluation Review, 38*, 514-545.

Wellington, J. (1990). Formal and informal learning in science: The role of the interactive science centres. *Physics Education, 25,* 247-252.

Wilde, M., & Urhahne, D. (2008). Museum learning: A study of motivation and learning achievement. *Journal of Biological Education, 42*, 78-83.

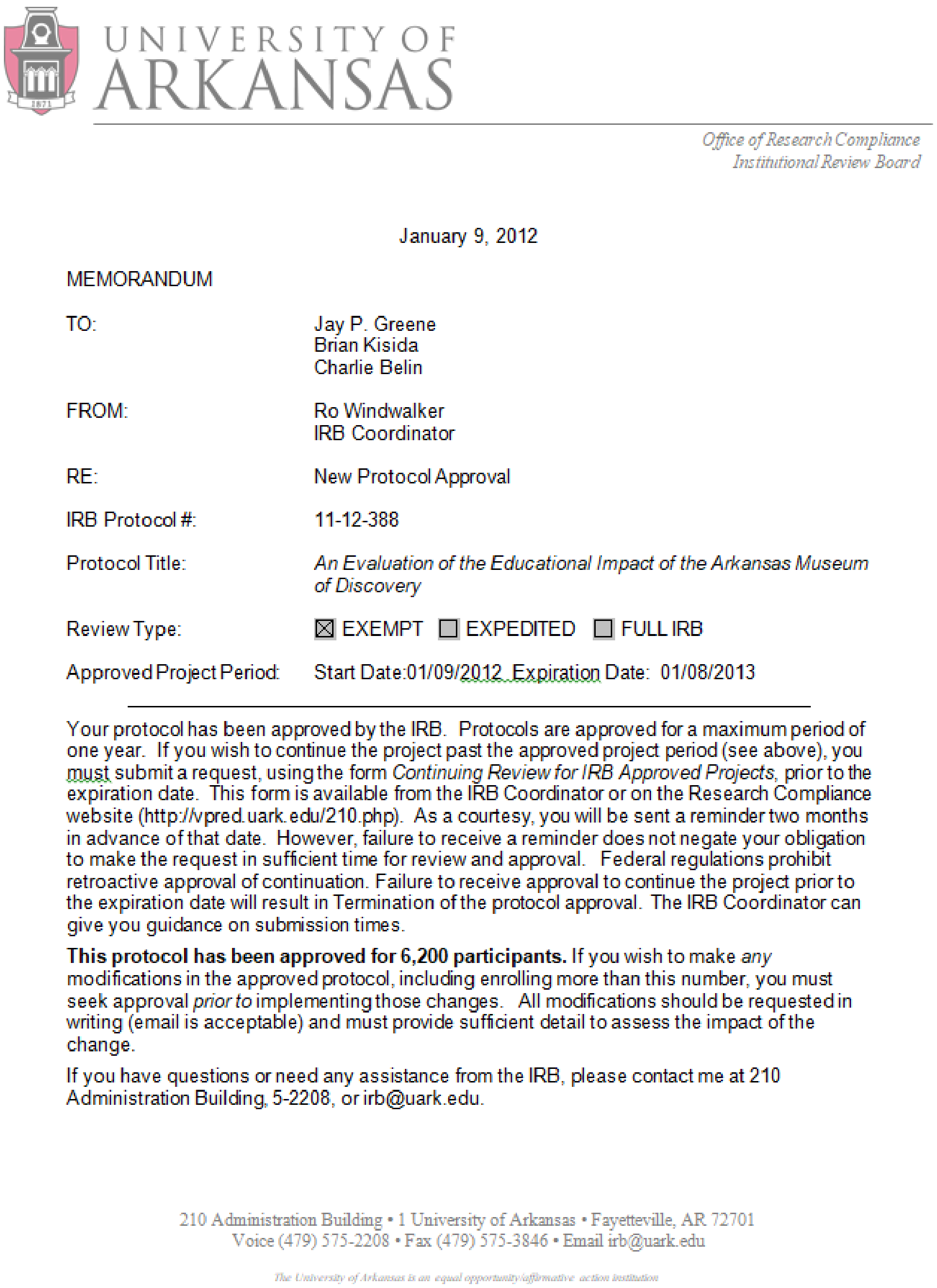
Williams, B., Onsman, A., & Brown, T. (2012). A rasch and factor analysis of a paramedic graduate attribute scale. *Evaluation and the Health Professions, 35*(2), 148-168.

Whitesell (2015). A day at the museum: The impact of field trips to informal science education institutions on middle school science achievement. *IESP Working Paper* #03-15.

Zoldosova, K,, & Prokop, P. (2006). Education in the field influences children's ideas and interest toward science. *Journal of Science Education and Technology, 15*, 304-313.

VI. Appendices

Appendix A: IRB Approval



UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

January 9, 2012

MEMORANDUM

TO: Jay P. Greene
Brian Kisida
Charlie Belin

FROM: Ro Windwalker
IRB Coordinator

RE: New Protocol Approval

IRB Protocol #: 11-12-388

Protocol Title: *An Evaluation of the Educational Impact of the Arkansas Museum of Discovery*

Review Type: ☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB

Approved Project Period: Start Date:01/09/2012 Expiration Date: 01/08/2013

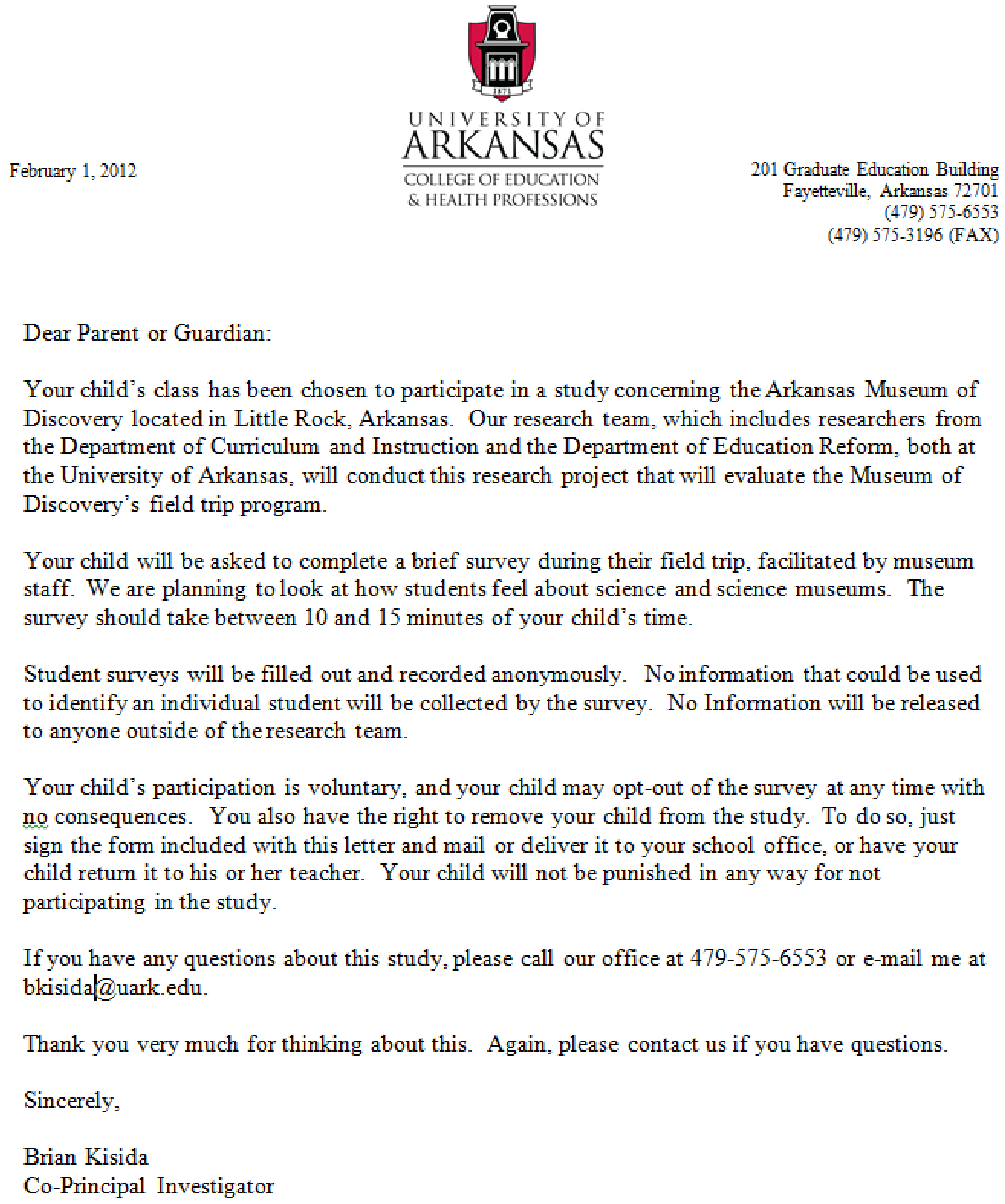
Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (http://vpred.uark.edu/210.php). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

This protocol has been approved for 6,200 participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 210 Administration Building, 5-2208, or irb@uark.edu.

210 Administration Building • 1 University of Arkansas • Fayetteville, AR 72701
Voice (479) 575-2208 • Fax (479) 575-3846 • Email irb@uark.edu

Appendix B: Letters to Parents and Teachers



UNIVERSITY OF
ARKANSAS
COLLEGE OF EDUCATION
& HEALTH PROFESSIONS

201 Graduate Education Building
Fayetteville, Arkansas 72701
(479) 575-6553
(479) 575-3196 (FAX)

February 1, 2012

Dear Parent or Guardian:

Your child's class has been chosen to participate in a study concerning the Arkansas Museum of Discovery located in Little Rock, Arkansas. Our research team, which includes researchers from the Department of Curriculum and Instruction and the Department of Education Reform, both at the University of Arkansas, will conduct this research project that will evaluate the Museum of Discovery's field trip program.

Your child will be asked to complete a brief survey during their field trip, facilitated by museum staff. We are planning to look at how students feel about science and science museums. The survey should take between 10 and 15 minutes of your child's time.

Student surveys will be filled out and recorded anonymously. No information that could be used to identify an individual student will be collected by the survey. No Information will be released to anyone outside of the research team.

Your child's participation is voluntary, and your child may opt-out of the survey at any time with no consequences. You also have the right to remove your child from the study. To do so, just sign the form included with this letter and mail or deliver it to your school office, or have your child return it to his or her teacher. Your child will not be punished in any way for not participating in the study.

If you have any questions about this study, please call our office at 479-575-6553 or e-mail me at bkisida@uark.edu.

Thank you very much for thinking about this. Again, please contact us if you have questions.

Sincerely,

Brian Kisida
Co-Principal Investigator

Arkansas Discovery Museum Study Removal Form

Complete this form and return it to your child's school to remove him or her from the study of the Discovery Museum.

I, _________________________, request that my child, ________________________, be removed

from the list of students participating in the Discovery Museum evaluation. I understand that my child will not be punished in any way for being removed from the study, and that any information that has been collected about him or her will be destroyed.

______________________________	__________________________
(Sign your name above)	(Today's date)

Return to:

Your child's school no later than XXXX X, 2012.

Letter to Teachers

During the 2012 school year, the Arkansas Museum of Discovery will be gathering data related to the impact a science museum field trip has upon student learning. The Department of Education Reform at the University of Arkansas will be helping to implement this study. As part of the study, some classrooms will be asked to complete a brief questionnaire as part of their field trip experience. The information gathered will provide valuable insight on the potential benefits of visiting a science museum to inform the practices of museums, educators, and policymakers, and provide students with an opportunity to critically reflect about their own field trips.

If you agree to participate, your classroom will fill out brief surveys at the Museum of Discovery during your assigned field trip, or a representative from the University of Arkansas may visit your classroom a few weeks before or after the field trip to administer the surveys.

We will supply you with an informational letter for parents that can be sent home with students as an accompaniment to field trip permission slips. The letter will provide information to parents and give them the opportunity to opt their child out of the study if they choose.

Information obtained from surveys will be confidential and any results will be reported anonymously. If you have additional questions about the study, please contact Brian Kisida at the Department of Education Reform by telephone at 479-575-6553 or by email at bkisida@uark.edu.

Thank you very much for thinking about this. Again, please contact us if you have questions.

Sincerely,

Brian Kisida
Co-Principal Investigator

Appendix C: Student Surveys

Instructions: Please complete this survey by supplying the requested information for each item. Please do not write your name on this survey. Do your best to answer EVERY QUESTION.

1) Are you a: ○ Boy? ○Girl?

2) What grade are you in? ○ 3rd ○ 4th ○ 5th ○ 6th ○ 7th ○ 8th ○ 9th ○10th ○ 11th ○ 12th

3) How would you identify yourself? ○ Hispanic/Latino ○White ○ American Indian

○Black or African American ○Asian ○Other: ______________

4) How many times have you ever visited the Arkansas Museum of Discovery on a school field trip?

○ None ○ 1 ○ 2 ○ 3 ○ 4 ○5 or more

5) How many times have you ever visited the Arkansas Museum of Discovery other than on a school field trip?

○ None ○ 1 ○ 2 ○ 3 ○ 4 ○5 or more

6) How many times have you ever visited any other science museum?

○ None ○ 1 ○ 2 ○ 3 ○ 4 ○5 or more

7) How interested are you in visiting science museums?

○ Not interested ○ A little interested ○ Interested ○ Very interested

8) How interested are you in learning about science?

○ Not interested ○ A little interested ○ Interested ○ Very interested

9) Would you like more science museums in your town? ○ Yes ○ No

10) I like school.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

11) I like science class.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

12) Science is an important part of my life.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

13) Trips to science museums are interesting.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

14) Trips to science museums are fun.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

15) I plan to visit science museums when I am an adult.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

16) I would be interested in joining a science club if my school offered one.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

17) I would like to study science in college.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

18) My favorite subject in school is:

○ Math ○ Reading/Writing ○ Science ○ History

19) When I grow up I want to be a(n):______________________________________

2012 Fall Student Survey

Instructions: Please complete this survey by supplying the requested information for each item. Please do not write your name on this survey. Do your best to answer EVERY QUESTION.

1) **Are you a**: ○ Boy? ○ Girl?
2) **What grade are you in?** ○ 3rd ○ 4th ○ 5th ○ 6th ○ 7th ○ 8th ○ 9th ○10th ○ 11th ○ 12th
3) **How would you identify yourself?**

○ Hispanic/Latino ○White ○ American Indian

○ Black or African American ○ Asian ○ Other: _______________

4) **How many times have you ever visited a science museum on a school field trip?**
   ○ None ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 or more
5) **How many times have you ever visited a science museum other than on a school field trip?**
   ○ None ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 or more
6) **How interested are you in visiting science museums?**
   ○ Not interested ○ A little interested ○ Interested ○ Very interested
7) **How interested are you in learning about science?**
   ○ Not interested ○ A little interested ○ Interested ○ Very interested
8) **Would you like more science museums in your town?** ○ Yes ○ No
9) **Have you been to the Arkansas Museum of Discovery on a school field trip?** ○ Yes ○ No
10) **I like school.**
    ○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree
11) **I like science class.**
    ○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree
12) **Science is an important part of my life.**
    ○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

13) Trips to science museums are interesting.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

14) I plan to visit science museums when I am an adult.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

15) I would be interested in joining a science club if my school offered one.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

16) I would like to study science in college.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

17) Trips to science museums are fun.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

18) I would like to learn more about science.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

19) Science is the best tool we have for understanding how the natural world works.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

20) Scientists often try to disprove their own ideas.

○ Strongly disagree ○ Somewhat disagree ○ Somewhat agree ○ Strongly agree

21) My favorite subject in school is:

○ Math ○ Reading/Writing ○ Science ○ History ○ Art/Music ○ PE

22) When I grow up I want to be a(n):______________________________________

Appendix D: Systematic Review Process in Detail

To find studies for the systematic review (see Figure D.1), I conducted a scoping search in multiple electronic databases for sources on the benefits of informal learning. Due to such a large number of responses (300,000-plus for multiple databases), I narrowed the research question to just field trips to Informal Science Education Institutions (ISEIs). I conducted scoping searches of just ISEI field trip studies and then outlined the inclusion criteria for sources to be considered for the review. I conducted the first search using four electronic databases and removed any duplicate studies from the four databases' results. After eliminating studies using information from the title and from any abstracts automatically generated that did not meet the inclusion criteria, I did a full-text review of all studies using the inclusion criteria. From this review of the full-texts, I compiled a list of more sources to potentially include. After realizing that many potential sources were not appearing in the four database searches, I conducted a second search of five databases using more general terms. I removed any duplicate studies from the second search that was already in the first search and eliminated studies from the second search based on the information provided in the title and from any automatically generated abstracts. I completed a full-text review of the studies in the second search not eliminated either as a duplicate or from the title/abstract review and reviewed all articles that were saved as additional potential sources to review. I continued this process for any new potential studies found from the full- text review and bibliographies of all studies that I had previously conducted a full-text review. I extracted data and quality assessed each study that was not eliminated during the full-text review and eliminated any study during data extraction that did not meet the methods' inclusion criteria. Lastly, I created study characteristic summary tables for the final studies and combined the positive and negative outcomes for studies included in the systematic review.

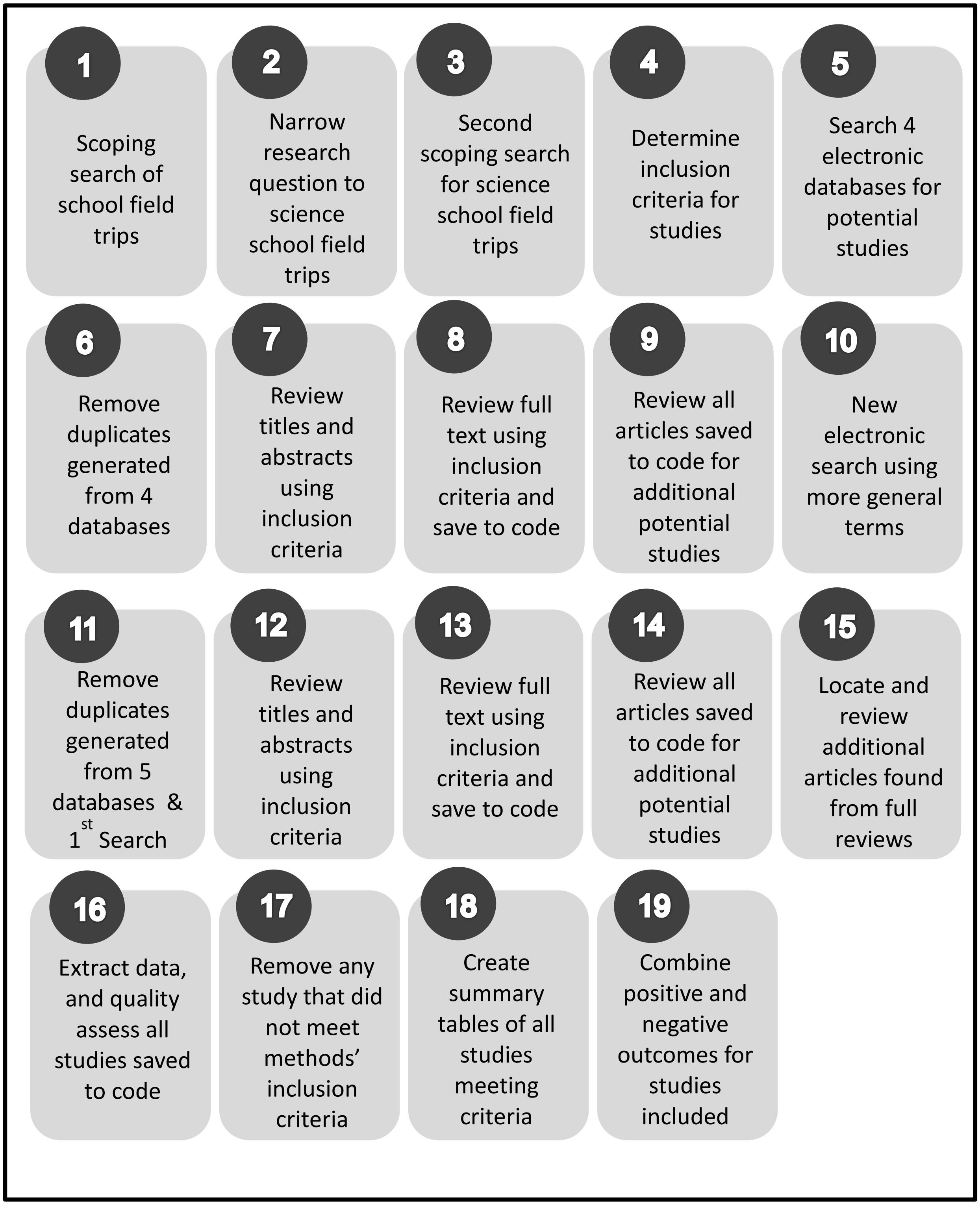


Figure D.1: Systematic Review Process

Identification and Organization of Studies

Identical studies. Occasionally, a researcher may publish multiple articles based on the data from a single study. I considered two papers to be from the same study if the papers used the same sample and data collection. Identical studies published separately will be cited together as one study. If the researchers analyze different subsets of data from different populations, the findings will be discussed separately. For example, Jarvis and Pell (2002) and Jarvis and Pell (2005) used the same venue for both studies with similar research questions but had a different population for each study.

Scoping searches. To have a general idea of the volume of literature on science field-trips to ISEIs, I conducted several scoping searches in JSTOR and ProQuest using key words such as *science, museum, student, school, field trip, out-of-school experience, informal learning,* and *education*. When I searched for *science* and *field trip* in the full text document, ProQuest found over 320,000 articles, while JSTOR found more than 780,000 articles.

Round one searches. With such a large number of articles produced from the scoping searches, I started adding more search terms, omitting studies with the terms *undergraduate* or *graduate* if mentioned in the abstract, and specifying which journals and sources to include. Although for each of the four search engines used to find studies, there were subtle differences in the search configuration. All first-round searches included the word *science*, *field trip, students*, and *schools*. Table D.1 provides information about the results of the first round of searches conducted on the 22nd day of February in 2015.

Table D.1

Search Terms for the First Search

Boolean Phrase	No Date Restriction	1/1/2000-2/22/2015
	Number of Results	
JSTOR: ((((ab:(science) AND ("field trip")) NOT ab:(graduate)) NOT ab:(undergraduate)) AND (schools)) AND (students))	56	44
EBSCO: science AND AB "field trip" AND schools AND students	98	86
ProQuest: all(science) AND ab("field trips") AND all(students) AND all(schools)	107	88
Web of Science: TOPIC: (science) AND TOPIC: ("field trip") AND TOPIC: (student) AND TOPIC: (schools)	36	34

Combined, the four search engines produced two hundred and fifty-two articles published between January 2000 and mid-February 2015. Twenty-six articles were duplicates in multiple search engines leaving 226 unique sources. For each article generated, I reviewed the title and the abstract based on the following inclusion criteria:

1. The study is about school field trips
2. which takes place in an Informal Science Education Institution
3. during a single day (one-stop trip)
4. by students in grades K-12th
5. with a minimum sample size of 30
6. that measure cognitive or affective outcomes
7. using either an RCT, CT or pre and post-test analysis
8. reporting on statistical significance
9. and published after January 1st, 2000.
10. in English (or can be translated into English with Google Translator)

Figure 2.2 shows the number of studies eliminated during each step based on the ten inclusion criteria. If I could not eliminate an article based only on the information provided by the abstract and/or title using the inclusion criteria criteria, the article was saved for further review. From the 226 articles, I reviewed the full text of 40 articles to determine if all inclusion criteria were met. Of the 40 articles, 8 articles met all the criteria and are included in the systematic review. Also, any stand-alone literature review was saved for review of potential sources. Every result from the search engines were stored in an Excel file and categorized based on primary topic. I also reviewed the literature review section and bibliography of the 40 articles for any other potential sources to include.

Round two searches. After reviewing the literature review section and bibliographies of 12 of the 40 articles, an additional 113 studies needed to be reviewed at the abstract level. With so many articles not appearing in the first round of searches, I decided to run another search of the four previous databases plus Science Direct using less restrictive search terms. For the second search, I searched only for *science*, *field trip*, and *education*. The search engine date ranges were from January 1, 2000 to May 2016, and all article citations and abstracts from this search were imported into EndNote software and organized into groups based on primary topic. The second-round search results from the five databases generated 1,487 articles. Of those 269 were duplicates either from within the five databases or from the first-round searches. This left 1,218 unique articles from the second round. After careful review of all titles and abstracts, 35 studies required a full-text review to determine if they met all the inclusion criteria. Of those studies only two met the search criteria and are included in the review. Table D.2 shows the results of the first and second searches by search engine.

Table D.2

Search Results by Search Engine

	JSTOR	EBSCO	Web of Science	ProQuest	Science Direct	Totals
1st Search "science" "field trip" "not graduate" not undergraduate" "schools" students" 2/22/2015	44	86	34	88	X	252
2nd Search "field trip" "science" "education" 5/9/2016	555	297	126	452	57	1487

Round three searches. Besides the two database searches, I also searched the bibliographies of all 75 sources that I conducted a full-text review. From the bibliographies, I made a list in Excel of any potential article to review at the abstract level. I reviewed 186 abstracts or skimmed the article if there was not an abstract. Of those 186 sources, I conducted 26 full-text reviews. I coded or started to code the methodologies of 15 of those articles, in which 7 met all inclusion criteria and are included in the review.

Merging the three searches. After merging the articles from the first search from February 22, 2015 with the second search from May 09, 2016, the total number of studies generated was 1,739. After removing the 26 duplicates from the first search, the 180 duplicates from the second search, and the 89 duplicates that were produced from both searches, plus adding in the 186 titles from the bibliography search, there were 1,630 unique studies. I conducted a total of 101 full-text reviews. I coded or started to code the results and methods of 38 studies. Of the 38 studies, only 17 met all the inclusion criteria and are included in this review. See Table 2.6 in chapter 2 for the combined totals of the three searches.

Types of field trip studies generated by search engines. Table D.3 shows the breakdown and number of the different field trip topics generated from queries of the electronic databases. Most articles generated from both searches were irrelevant, for example, an article about an elective in a PhD program for health care ethics education (Bustillos & Thornock, 2013). The second largest group of studies focused on student benefits in a K-12 setting, although many of these did not meet the inclusion criteria. For example, Dohn (2013) and Glick and Samarapungavan (2008) did not have a control group or a pre-test to evaluate if a change in learning or attitudes occurred after visiting the ISEI. See Table 3.3 for the other major topics generated.

Table D.3

Number of Different Field Trip Topics Generated by Search Engine

Category	# Articles
Irrelevant	952
Student Benefits	76
Post-Secondary	72
Professional Development	67
Part of a Curriculum Unit	65
Virtual Field Trips	61
Personal Reflection/News	42
Summer Camps/Multi-day/afterschool/site visits to schools	27
Museum Educators	19
Mobile/digital cameras/technology/worksheets	22
Educator Perceptions	17
Programs	13
Conference Field Trips	11

Screening studies using inclusion criteria. All articles of interest were obtained through the University of Arkansas library or an interlibrary loan. Using the Excel database from the first round and third round of searches and the Endnote database from the second round of searches, I applied the inclusion criteria to titles and abstracts (if available), full report, and again during data extraction and quality assurance. The exclusion reason was recorded in either the Excel database or the Endnote database. Any article excluded that may be important for background research and policy implications was saved into a separate folder. In cases with multiple reasons for exclusion, the highest inclusion criterion was labeled. For example, a study on college students who went on a single day field trip to an ISEI that did not have a pre-test or comparison group would be excluded for the fourth reason (not K-12 grade students) and not the sixth reason (no comparison group). See Figure D.2 for the numbers of studies eliminated by inclusion reason and by each search.

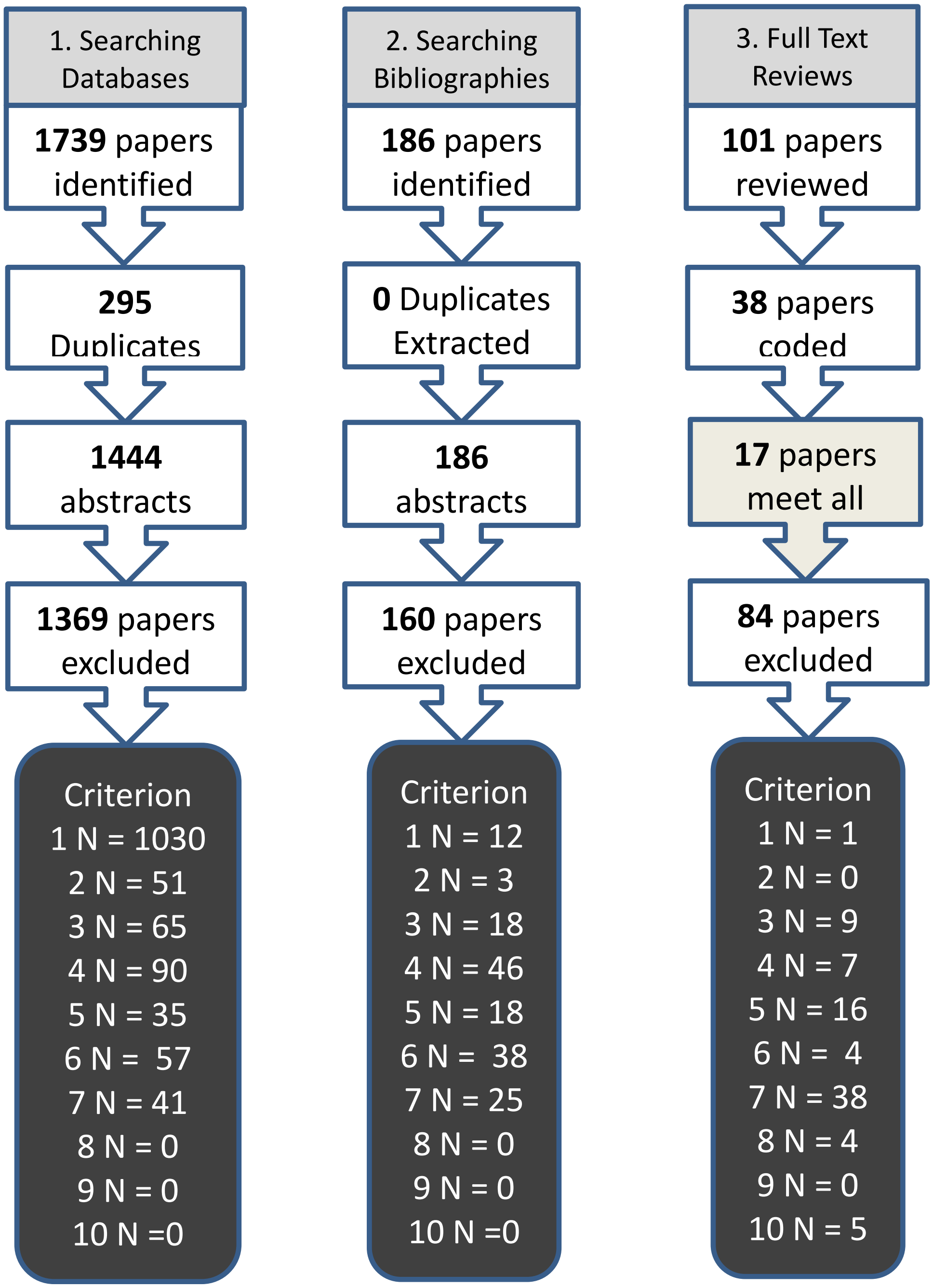


Figure D.2: Synthesis Map Used to Determine Final Studies

Coding studies after full-text review. After the full-text review, 38 articles that appeared to meet all the inclusion criteria were coded based on the following study characteristics: publication date, location of study, type of study, aim of study, science discipline, number of students, number of classes, number of schools, ages of participants, study design, instrument development, methods used to collect data, methods used to analyze data, conclusions, and overall quality. During the coding process, 11 articles were eliminated for not meeting the inclusion criteria. This left a final total of 17 articles to include in the review.

Data extraction and quality assurance. For each outcome listed, the pre-test or control group's average mean, the standard error or standard deviation from the pre-test/control group, the post-test or treatment group's standard mean, the standard error or standard deviation from the post-test/treatment group was recorded in the Excel database. To measure the quality of the study, I used the Effective Public Health Practice Project (2010) quality assessment tool for quantitative studies. I chose this tool because of the ease of use, the applicability of the components to education policy studies, and how the tool considers a mix of study methodologies from RCT to pre/post. Each study is evaluated on eight categories: (1) selection bias, (2) study design, (3) confounders, (4) blinding, (5) data collection methods, (6) withdrawals and drop-outs, (7) treatment integrity, and (8) and analyses. For each category, the rating is strong, moderate, or weak. Lastly, the category ratings are combined for an overall rating of strong, moderate-strong, moderate, weak-moderate, or weak.

Synthesis of evidence. The last step for this systematic review is to combine the studies and synthesize the findings to answer the systematic review question: What are the effects of visiting an Informal Science Education Institution as part of a school field trip on K-12 students' understanding in science or attitude towards science?

The studies were divided based on outcomes measured as either affective or cognitive and synthesized separately. Studies that had findings for both affective and cognitive outcomes would be included in both syntheses. Outcomes were positive (+), negative (-), or zero (0).

Limitations of the review. The review has two primary limitations. First, this review was conducted by a single researcher, so studies were not quality assessed by multiple people. I conducted all of the database and bibliography searches and determined which articles to include and exclude, which leaves room for error and a possibility some articles that met the criteria were missed. Second, the majority of studies on this topic did not incorporate a control-treatment method and not one researcher incorporated a large-scale RCT method. Therefore, for the pre-post only studies and studies that used a control group but the researcher did not make any statistical comparisons between the control and treatment group, I am unable to conclude if there is a difference in outcomes between students who visited an ISEI and students who did not.

Study Characteristics and Methodologies

Time between field trip experience and post-assessment. One major difference between this set of studies is the amount of time that passed between the field trip experience and the post-assessment. Six studies gave the post-assessment immediately following the field trip experience, and then three of those studies gave a delayed post-assessment anywhere from 1 week to 2 months after the field trip. Another three studies conducted the post-assessment within the first week, and one of those studies conducted a delayed post-assessment two months after the experience. All the post-assessments conducted between 1 and 5 months were considered delayed post-assessment where the researchers had previously conducted an earlier post-assessment. Basten et al. (2014) studied two different samples so their research is divided into study 1 and study 2, and for both studies they conducted a delayed post-assessment.

Only one study, Jarvis and Pell (2005), conducted three post-assessments at the 1 week, 2 month, and 4 month mark. Two studies did not provide any specific information about when the post-assessment was given to participants. Table D.4 provides information about the timeline between the field trip experience and the post assessment(s).

Table D.4

From Field Trip to Post-Assessment

Time between field trip and post-assessment	**Study**
Immediately following	Holmes, 2011
	Krombab and Harms, 2008
	Sentürk and Özdemir 2014
	Stavrova and Urhahne, 2010
	Strum and Bogner, 2010
	Sweet, 2014
	Wilde and Urhahne, 2008
Within 1 week	~Basten et al., 2014
	Prokop et al., 2007
	Futer, 2005
1 week	~Basten et al., 2014
	Jarvis and Pell, 2005
	*Sentürk and Özdemir 2014
	Itzek-Greulich et al., 2015
2 weeks	Ballouard et al., 2012
	Freedman, 2010
1 to 3 weeks	Jarvis and Pell, 2002
1 month to 2 months	*Holmes, 2011
	*Strum and Bogner, 2010
2 months to 3 months	*Basten et al., 2014
	*Jarvis and Pell, 2005
4 months to 5 months	*Basten et al., 2014
	*Jarvis and Pell, 2005
Unknown	Kamarainen, 2013
	Puhek, 2012

*The second post-assessment by a particular study, **the third post-assessment, ~a study with different sets of participants, one group attended a middle track school and the other group attended a college-ready, higher track school. Some variation in how the researchers conducted the study on each group.

Types of Methodologies. The inclusion criteria specifically focused on three types of methodologies used: pre-post assessment without a control, pre-post assessment with a control for comparison, and pre-post assessment with a control to determine if a difference exists between the treatment group and the control group. Ideally the control should be randomized. Studies that used descriptive data such as interviews or researcher observations were not included because these studies do not utilize a comparison group. The final 17 studies included a comparison group either through a different set of participants or through the pre-assessment. Eight studies did not include any kind of control (Basten et al., 2011; Freedman, 2010; Futer, 2005; Jarvis & Pell, 2002; Jarvis & Pell, 2005; Kamarainen et al., 2013; Krombaß & Harms, 2008; Stavrova & Urhahne, 2010). Five studies included a control but did not estimate a difference between the control group and the treatment group (Ballouard et al., 2012; Prokop et al., 2007; Puhek et al., 2012; Sweet, 2014; Wilde & Urhahne, 2008). Four studies included a control and calculated if the treatment group's change in attitude or change in knowledge was different from the control group's change in attitude or change in knowledge after the treatment students experienced the field trip (Holmes, 2011; Itzek-Greulich et al., 2015; Sentürk & Özdemir, 2014; Sturm & Bogner, 2010). Only three studies incorporated a randomized method for selecting treatment and control participants. For the three RCT studies, Holmes (2011) conducted the post-assessment the moment they arrived at the museum but before they had a chance to experience the field trip. Prokop et al. (2014) randomized the control and treatment groups at the school level, but they did not calculate any differences between the two groups. They only calculated differences within each group (pre-assessment to post-assessment). Itzek-Greulich et al. (2015) used a control group that did not receive the curriculum of the other three experimental groups (school-only, school and field trip, and field trip-only). For this systematic review and based on the analysis provided in Itzek-Greulich et al.'s publication, the school-only group was considered the control and the school and field trip group was the treatment. Although the researchers randomized classes and included the pre-score in the regression analysis, the school-only group scored much-higher on the pre-assessment than the school/field trip group.

Reliability and validity of instruments used. Generally, the authors created their own instruments to measure affective and cognitive changes; but in some instances, they selected and modified instruments created by other researchers from a previously published study. For the nine affective domain studies, all used Likert-style statements except one study (Ballouard et al., 2012) which used open and closed questions. Also, another study (Sentürk & Özdemir, 2014), did not provide details on the number of points in the Likert-scale. Three studies (Ballouard et al., 2012; Kamarainen et al., 2013; Strum & Bogner, 2010) did not provide any information about the reliability of the instrument. Three studies (Ballouard et al., 2012; Futer, 2005; Strum & Bogner, 2010) did not provide any information about the validity of the instrument, and three studies (Jarvis & Pell, 2002; Jarvis & Pell, 2005; Kamarainen et al., 2013) only claimed the instruments were previously validated in a different study. Of the studies that reported a reliability measure, all used Cronbach's alpha which ranged from 0.65 to 0.94, except one study (Holmes, 2011) which reported reliability using the Kuder-Richardson 21 formula. The studies ranged from seven statements (Kamarainen et al., 2013) to 47 statements (Ballouard et al., 2012), with a group average of 32 statements. Table D.5 provides information on the instruments used to measure the affective domain.

When measuring cognitive outcomes, all researchers created their own assessment based on the curriculum taught during each specific field trip. The shortest assessment consisted of 7 questions while the largest assessment had 41 questions. Ten studies included or used only multiple-choice formatted questions and three studies (Freedman, 2010; Jarvis & Pell, 2002; Kamarainen et al., 2013) did not provide any details of the type of questioning used. Seven studies did not provide any measure of reliability, and five studies measured reliability with Cronbach's alpha with the low at 0.163 and a high at 0.78. Seven studies did not provide any information about instrument validity, while five studies consulted with several experts such as teachers or museum educators. Table D.5 provides details on the instruments used for measuring cognitive outcomes.

Table D.5

Study Instruments' Validity and Reliability Measures

Study Instrument	**?s**	**Validity**	**Reliability**	**Study**
Closed/Open Questions	47	NA	NA	Ballouard et al., 2012
General feelings about snakes				
Willingness to protect snakes				
Possible influence of previous experiences with snakes				
Preferred activities				
4-point Likert Scale	10	NA	0.80<a<0.83	Futer, 2005
Overall environmental attitudes				
Level of internal Locus of control				
Sense of personal responsibility towards environment				
Multiple Choice	9		0.16<a<0.69	
Questions varied by program attended				
3-point Likert Scale	38	Previously validated	0.65<a<0.78	Jarvis and Pell, 2002
Science Enthusiasms Scale-engaging in science at school and home				
Science in a Social Context Scale-views on the uses of science to improve human life				
Knowledge test- Type of question not provided	8		0.63<a<0.67	
3-point Likert Scale	74	Previously validated		Jarvis and Pell, 2005
Science Enthusiasms Scale-engaging in science at school and home			0.72<a<0.78	
Science in a social context scale-views on the uses of science to improve human life			0.66<a<0.71	
Space Interest Scale-views about space exploration			a=0.71	
Planning and Teamwork-values planning with peers			a=0.73	
Working Confidence Scale-views of being a leader and responsibility for actions			a=0.77	
Anxiety Scale			a=0.71	
Open ended questions-recall	7	Cohen's K=.89	NA	Basten et al., 2016
Multiple choice-factual lnowledge	21			
Type of questioning not provided	19	NA	NA	Freedman, 2010
5-point Likert Scale	7	Previously validated	forthcoming	Kamarainen et al., 2013
Self-efficiency to ecosystem knowledge				
Skills				
Evaluation of environmental monitoring				
Knowledge test-Type of question NA		3 experts		
5-point Likert Scale Children's Academic Intrinsic Motivation Inventory	44	Reviewed by science educ. staff and teachers	KR21=0.31	Holmes 2011
Motivational orientation in science and other academic areas				
General orientation towards school learning				
Multiple Choice	30			
Knowledge test				

Table D.5 (cont.)

Study Instruments Validity and Reliability Measures

Study Instrument	**?s**	**Validity**	**Reliability**	**Study**
Affective: 5-point Likert Scale *Biology Attitude Questionnaire modified to measure chemistry*	12	Reviewed by three experts in the field	0.74<a<0.77	Prokop et al., 2007
Biology as a school subject				
Natural environment outside				
Future work in biology				
Cognitive-multiple choice	16		a=0.78	
Knowledge Assessment				
Likert Style Statements *Attitudes Towards Science Scale*	33	Reviewed by 22 experts. Piloted by 116 students	a=0.94	Sentürk & Özdemir, 2014
Self-concept in school science				
Science outside of school				
Practical work in school science				
Learning science in school				
Future participation in science				
Importance of science				
5 point Likert Scale *Intrinsic Motivation Inventory*	27	NA	NA	Sturm & Bogner, 2010
Interest and enjoyment				
Perceived choice				
Value and usefulness				
Perceived competence.				
Cognitive				
Multiple choice	17			
Open-ended questions	1			
Multiple choice	14	NA	NA	KrombaB & Harms, 2008
Open ended questions-analysis and evaluation	8	NA	a=0.71	Puhek et al., 2012
True/False	2	Reviewed by 2nd grade teachers	NA	Sweet, 2014
Matching	3			
Multiple choice	3			
Multiple choice and open ended questions	26	NA	a=0.65	Wilde & Urhahne, 2008
Multiple-choice	33	0.04>RMSEA>0.11	0.43>EAP>0.65	Itzek-Greulich et al., 2015
Rate familiar terms	8	RMSEA=0.19	EAP=0.67	

Quality-Assurance Results

Each study is evaluated on: (1) selection bias, (2) study design, (3) confounders, (4) blinding, (5) data collection methods, (6) withdrawals and drop-outs, (7) treatment integrity, and (8) and analyses. For each category, the rating is strong, moderate, or weak, and then the category ratings are combined for an overall rating of strong, moderate-strong, moderate, weak-moderate, or weak. Of the 17 studies, 11 studies are rated as having an overall moderate methodological approach, 3 studies are rated as having an overall weak methodological approach, 1 study is rated as having an overall moderate to strong methodological approach, 1 study is rated as having an overall strong methodological approach, and 1 study is rated as having a weak to moderate methodological approach. Table D.6 provides each study’s overall rating and ratings for the individual components. Over 75% of studies included in the systematic review had at a minimum a moderate methodological approach specifically regarding study design, treatment integrity and analysis.

Table D.6

Quality-Assurance Results

	Ballouard et al. (2012)	Basten et al. (2016)	Freedman (2010)	Futer (2005)	Holmes (2011)*	Itzek-Greulich et al. (2015)*	Jarvis & Pell (2002)	Jarvis & Pell (2005)	Kamarainen et al. (2013)	Krombab & Harms (2008)	Prokop et al. (2007)*	Puhek et al. (2012)	Sentürk & Özdemir (2014).	Stavrova & Urhahne (2010)	Sturm & Bogner (2010)	Sweet (2014)	Wilde & Urhahne (2008)
Selection Bias	M	W	W	M	W	S	M	M	W	W	M	M	M	M	M	W	W
Study Design	M	M	M	M	M	S	M	M	W	W	M	M	M	M	M	W	M
Confounders	W	S	M	W	M	S	M	W	W	W	S	M	M	M	S	W	M
Blinding	M	S	M	M	W	M	M	M	M	W	M	M	M	W	S	W	M
Data Collection	W	W	W	M	W	S	M	S	W	W	S	M	M	W	S	W	M
Withdrawals	W	W	M	W	W	S	W	M	W	W	S	M	M	W	M	W	W
Treatment Integrity	M	M	M	S	M	S	S	M	W	W	M	S	S	S	S	M	M
Analysis	M	M	M	M	M	S	M	M	S	M	S	S	M	M	M	W	M
Overall	**M**	**M**	**M**	**M**	**W** **M**	**S**	**M**	**M**	**W**	**W**	**M** **S**	**M**	**M**	**M**	**M**	**W**	**M**

*RCT design